

# WICC 2013

XV Workshop de Investigadores  
en Ciencias de la Computación

Paraná - Entre Ríos  
18 y 19 de abril de 2013  
18 19 18 19 18 19 18 19 18 19 18 19 18 19 18 19



Universidad Autónoma  
de Entre Ríos



RedUNCI

# WICC 2013

XV Workshop de Investigadores  
en Ciencias de la Computación



Paraná - Entre Ríos  
18 y 19 de abril de 2013

## XV Workshop de Investigadores en Ciencias de la Computación 2013

---

ISBN: 9789872817961



Facultad de Ciencia y Tecnología

Universidad Autónoma de Entre Ríos (UADER)

wicc2013@uader.edu.ar

### WICC 2013

El Workshop de Investigadores en Ciencias de la Computación es organizado, a partir de 1999, por la Red de Universidades Nacionales con Carreras de Informática (RedUNCI). El objetivo del Workshop es crear un foro para el intercambio de ideas entre Investigadores en Ciencias de la Computación, de modo de fomentar la vinculación y potenciar el desarrollo coordinado de actividades de Investigación y Desarrollo entre los mismos.

**Compilación:**

ASS Claudio Caluva  
Lic. Silvia Aranguren  
Lic. Rodolfo Muzachiodi





# INDICE

## Arquitectura, Redes y Sistemas Operativos

---

Un Mecanismo De IPC De Microkernel Embebido En El Kernel De Linux.	11
Análisis De Prestaciones De Tráfico Multicast En Redes Mixtas Ipv4 E Ipv6.	16
Internet Del Futuro Y Ciudades Inteligentes.	21
Conectividad WSN: Implementación De Un Middleware WSN-IP-WWW	28
Protocolos Para Mejorar La Performance De Las Aplicaciones Web	34
Redes Inalámbricas De Uso Comunitario	38
Aplicaciones De Internet De Las Cosas SIPIA6 - Red De Sensores Inalámbricos Con Ipv6	43
Desarrollo De Aplicaciones Nativas Para Ipv6	48
Administración De Qos En Ambientes De Redes De Servicios Convergente	53
Sistemas Embebidos En Red Con Requerimientos De Tiempo Real	58
Despliegue De Manets Para M-Learning En Zonas De Recursos Limitados	63
SARA Six Análisis , Implementación Y Evaluación De Servicios Colaborativos Competitivos Aplicados A Redes De Avanzada	66
Transaction Signature (Tsig). Una Alternativa De Seguridad Para Transferencias De Zonas Dns.	70
Generación De Una Biblioteca Para Control De Tareas En Tiempo Real En Un Sistema Operativo Didáctico	73

## Bases de Datos y Minería de Datos

---

Incorporar Actividades Virtuales En Educación Superior: Algoritmo De Segmentación De Docentes Según Sus Competencias.	78
Ingeniería De Software Para Clasificar Patrones Cognitivos Conductuales.	83
Redes Sociales Y Motores De Búsqueda.	87
Aplicación De Técnicas De Minería De Datos Al Análisis De Situación Y Comportamiento Académico De Alumnos De La Ugd.	94
Gestión De Datos Biometricos En Bases De Datos Objeto-Relacionales	97
Análisis De Fuentes De Información Para El Proceso De Diseño De Un Datawarehouse Sobre Pacientes Diabéticos	102
Análisis De Incidentes Informáticos Usando Modelos De Asociación Y Métodos Del Análisis De Datos Multivariante	107
Investigación Básica Y Aplicada En Bases De Datos Deductivas	102
Pertinencias De Planes De Estudio De Carreras De Informatica Con Normativas Establecidas Por Coneau	116
Relacionando Comentarios Textuales Y Valores Numéricos En Encuesta De Satisfacción De Usuarios	121
Minería De Datos En Redes Educativas	126
Generación De Un Procedimiento De Búsqueda De Outliers Sobre Campos Alfanuméricos En Logs De Auditoria	129
Patrones Estadísticos Relacionados Con El Perfil Del Alumno De La Facultad De Tecnología Y Ciencias Aplicadas	134
Instrumentos Para La Gestión De Proyectos De Explotación De Información	137
Elaboración Y Ensayo De Aplicación De Algoritmos De Ingeniería Del Conocimiento En La Gestión De Redes Sociales.	142
Bases De Datos De Objetos No Estructurados	147
Almacenamiento De Datos Xml En Oracle 11g	152
Visión Ampliada Del Management (Epm)	157
Generador Automático De Modelos De Datos Normalizados En Bases De Datos Relacionales	161
Bases De Datos No sql En Cloud Computing	166
Prácticas Y Aplicaciones De Ingeniería De Requisitos En Proyectos De Explotación De Información	172
Geometría Computacional Y Bases De Datos	176
Minería De Datos En Bio-Ciencias	181

Evaluación De La Calidad De La Información Extraída Por Wrappers, De Un Sitio Web	185
Minería De Datos Aplicada A La Conservación Ex Situ De Recursos Fitogenéticos De San Juan.	189
Indexando Texto En Memoria Secundaria	194
Minería De Datos Utilizando Sistemas Inteligentes	199
Sistema De Información Geográfica En La Web Para El Problema De Ruteo De Vehículos	204
Codifexa - Colecciones Digitales Para Exactas	209
Consolidación De Un Modelo Para Bases De Datos No Convencionales	215
Recuperación Eficiente De Datos En Bases De Datos Multimedia	220
Aplicando Estrategias Y Tecnologías De Inteligencia De Negocio En Sistemas De Gestión Académica	225

## Computación Gráfica, Imágenes y Visualización

---

Transformaciones Geométricas Para Facilitar La Identificación De Objetos En Imágenes Digitales	230
Ontologías Y Semántica En El Proceso De Visualización.	235
Tics Y Difusión Del Patrimonio Cultural.	240
Visualización En Ciencias Geológicas.	245
Reconocimiento De Acciones En Videos De Tenis Usando Flujo Óptico Y CRF	249
Simulación On De Personajes Conversacionales Virtuales Dentro De Un Entorno De Realidad Virtual	254
Tratamiento De Imágenes De Fibras Animales	259
Tracking De Múltiples Objetos Aplicado A Insectos	264
Líneas De Realidad Aumentada Aplicadas A La Geología Y Los Libros Aumentados	269
Visión Por Computador E Informática Gráfica. Realidad Virtual, Realidad Aumentada E Interfaces Avanzadas	274
Imágenes Satelitarias Para Detección Y Monitoreo Operacional De Los Cambios En El Bosque Nativo Del Norte De La Provincia De Entre Ríos.	280

## Innovación en Educación en Informática

---

Estrategias Didácticas En El Uso De Herramientas Software Para Favorecer La Comprensión De Los Alumnos En La Enseñanza Sobre Compiladores	285
El Pensamiento Crítico y los Entornos Virtuales de Aprendizaje - Perspectiva Sistémica	289
Motores de Juegos e Inteligencia Artificial para la Enseñanza	294
RemoteBot: una Aplicación que Combina Robots y Dispositivos Móviles	299

## Ingeniería de Software

---

Propuesta De Un Portal De Replicación De Experimentos - Análisis De La Personalidad En Los Equipos En El Desarrollo De Software.	305
Accesibilidad Web Como Medida De Calidad En El Marco Del Proyecto "Sistemas Y Tic: Técnicas Y Herramientas"	310
Caracterización De Los Riesgos Inherentes A La Ingeniería Reversa.Herramientas Para La Ingeniería Reversa.	314
Optimización De La Gestión Del Avalúo Fiscal En La Provincia De San Luis.	319
Generación De Sistemas De Recuperación De Información Para La Gestión Documental En El Área De Las Ciencias De La Computación.	324
Estudio Comparativo Y Análisis De Rendimiento De Los Lenguajes De Manipulación De Datos En Bases De Datos Orientadas A Objetos Y Bases De Datos Objeto-Relacionales.	329
Aproximacion Para Un Método De Elicitación Y Especificación De Requerimiento De Seguridad Para El	334

<b>Desarrollo De Software.</b>	
<b>Construcción De Un Prototipo De Repositorio Temático En El Campo De La Ingeniería De Software</b>	<b>339</b>
<b>Actualidad Del Estado De Una Nueva Metodología De Gestión Y Desarrollo De Proyectos De Software: Metodología Ágil Basada En Telecomunicaciones Mate</b>	<b>343</b>
<b>Análisis Metodológico Para La Utilización De Process Mining Como Tecnología De Optimización Y Respaldo De La Implementación De Procesos De Negocio Bajo El Marco De Bpm</b>	<b>348</b>
<b>Técnicas Y Herramientas Para Desarrollo De Sitios Web Accesibles</b>	<b>353</b>
<b>Trazabilidad De Versiones En Ingeniería De Requisitos</b>	<b>358</b>
<b>Evaluación De Metodologías Y Técnicas Para La Generación De Pruebas Aplicadas Un Caso Real A Partir De Sus Especificaciones Funcionales</b>	<b>363</b>
<b>Utilizando Contratos Jml Para Optimizar Diseños Orientado A Objetos Siguiendo Mda</b>	<b>367</b>
<b>Implementación De Una Red Social Académica Universitaria</b>	<b>372</b>
<b>Modelos Y Tecnologías En Gobierno Electrónico</b>	<b>378</b>
<b>El Modelado De Requerimientos En Las Metodologías Ágiles</b>	<b>383</b>
<b>Evaluación De La Calidad En Sitios Web Bancarios</b>	<b>388</b>
<b>Extendiendo Bpmn2 Para Soportar Workflows Científicos De Esteco</b>	<b>392</b>
<b>Modelo Para Aplicaciones Sensibles Al Contexto: Validación Y Evaluación</b>	<b>396</b>
<b>Optimización De La Calidad De Los Sistemas Móviles</b>	<b>401</b>
<b>Metodologías Ágiles Para El Desarrollo De Sistemas De Cómputo Distribuido</b>	<b>405</b>
<b>Reuso Orientado A Servicios</b>	<b>410</b>
<b>Interoperabilidad De Modelos De Simulación</b>	<b>415</b>
<b>Desarrollo De Modelos De Evaluación Usando Operadores De Una Lógica Continua</b>	<b>420</b>
<b>Los Sitios Web Gubernamentales Como Facilitadores Del Gobierno Electrónico - Análisis De La Evolución De Los Sitios Web Municipales Del Conurbano Bonaerense</b>	<b>425</b>
<b>Quco2: Una Herramienta Para Medir La Calidad De Las Aplicaciones Web</b>	<b>430</b>
<b>Generación Automática De Código Rmi A Partir Del Modelo De Análisis Utilizan Do Reglas Relations/Qvt</b>	<b>434</b>
<b>Modelado Y Desarrollo De Sistemas Robóticos Educativos</b>	<b>438</b>
<b>Ingeniería De Software En Crecimiento: Enfoque Conceptual Y Metodológico, Y Herramientas</b>	<b>443</b>
<b>Un Acercamiento En La Integración Entre Bpmn Y Soa</b>	<b>447</b>
<b>Metodología De Modelado De Aplicaciones Web Móviles Basada En Componentes</b>	<b>451</b>
<b>Reuso De Software Orientado A Dominios</b>	<b>456</b>
<b>Desarrollo De Un Modelo Basado En Servicios Digital Es Comunes Reutilizables Para Gobierno Electrónico</b>	<b>461</b>
<b>Herramientas De Modelado Y Simulación Para Sistemas De Gran Escala</b>	<b>466</b>
<b>Semántica Oculta En Modelos De Requisitos</b>	<b>471</b>
<b>Reuso De Conocimiento En Foros De Discusión Técnicos</b>	<b>476</b>
<b>Modelado Específico De Dominio (Dsm) Para Una Familia De Aplicaciones De Software De Simulación De Ecosistemas Naturales</b>	<b>481</b>
<b>Aproximación A Los Métodos Formales: Una Experiencia Con Un Método De Desarrollo Liviano</b>	<b>486</b>
<b>Validación De Especificaciones No Funcionales De Aplicaciones Web A Través De Técnicas De Testing De Usabilidad</b>	<b>490</b>
<b>Construcción Del Lel De Requisitos</b>	<b>495</b>
<b>Proyectos De Evaluación De Productos De Software Con Un Nuevo Framework De Calidad</b>	<b>500</b>
<b>Aplicación De Software Colaborativos O Groupware Empleando Metodología Scrum En Sistemas Y Organizaciones</b>	<b>505</b>
<b>Sistemas Ubiquos: Desarrollo Y Aplicaciones</b>	<b>510</b>
<b>Ingeniería De Software Aplicada A Un Sistema De Gestión De Calidad En Centros Educativos</b>	<b>517</b>
<b>Transformación De Modelos Conceptuales Mediante Uso De Xslt .</b>	<b>521</b>
<b>Conceptos De Dinamismo Aplicados A Servicios Y Work Flows En Bpms Basados En Cloud Computing</b>	<b>526</b>
<b>Validación De Requerimientos A Través De Modelos Conceptuales - Modelos Y Transformaciones</b>	<b>531</b>
<b>Redes Bayesianas Aplicadas A La Ingeniería De Software</b>	<b>538</b>
<b>Aplicación De Un Método Para El Estudio De La Calidad De Los Procesos De Negocio En La Comparación De Metodologías Ágiles De Desarrollo</b>	<b>543</b>
<b>Gestión Del Mantenimiento Aplicando Una Metodología Basada En Sistemas Y Tecnologías De Información - Fase Ii</b>	<b>548</b>
<b>Definición De Métricas En Un Metamodelo Usando Ocl Para Diseño Web</b>	<b>551</b>
<b>Aplicación De Técnicas De Comprensión De Programas Para La Obtención De Información De Modelos De Procesos Workflow</b>	<b>556</b>
<b>Comprensión De Especificaciones De Procesos De Negocios Escritas En Bpmn</b>	<b>562</b>
<b>Análisis Y Comparación De Plataformas Brms A Través De Una Prueba De Concepto</b>	<b>567</b>

Automatización De La Evaluación De La Usabilidad Del Software	572
Procesos De Ingeniería De Requerimientos Basados En Modelos De Madurez Orientados A La Certificación Iso 9001	576
Mejora De Procesos En El Desarrollo De Sistemas De Software Y En Procesos De Gestión. Experiencias En Pymes.	581
Ingeniería De Software En El Desarrollo De Aplicaciones Para Dispositivos Móviles.	586
Extracción De Información Estática De Programas Escritos Usando El Paradigma Orientado A Objetos	590
Métodos Y Procesos Para Sistemas Distribuidos. Experiencias	595
Análisis Comparativo De Modelos De Calidad Orientado Al Desarrollo De Software En Pymes	601
Plantilla Cupido: Automatización Y Avances	606
Refactoring : Secuencia De Aplicación	611
Análisis De Disciplinas De Las Ciencias De La Conducta Como Aporte A La Ingeniería De Requerimientos	616
Persistencia De Mediciones Como Apoyo A La Gestión De Proyectos De Software	621
Herramientas Software Para Una Optima Gestión Del Proceso De Desarrollo De Software , A Utilizarse En Las Empresas De Software De La Provincia De Catamarca	625
Definición Formal De Una Metodología Para La Generación De Sistemas De Software Orientados A Servicios	629
Criterios De Calidad En El Desarrollo De Aplicaciones Web	634

## Procesamiento Distribuido y Paralelo

---

Diseño de Algoritmos para Plataformas Paralelas.	638
Desarrollo y Sintonización Automática de Aplicaciones Paralelo/Distribuidas	643
Desarrollo de Aplicaciones Paralelo/Distribuidas orientadas a la Predicción de Incendios Forestales	648
Construyendo un Sistema de Cómputo Distribuido Multipropósito	653
Búsquedas por Rango sobre Plataformas GPU en Espacios Métricos	658
Aspectos energéticos de sistemas de computación	663
Elaboración de Estrategias Paralelas para Búsquedas por Similitud en Espacios Métricos	667
Simulación De Aplicaciones Paralelas Y Mecanismos De Tolerancia A Fallos	672
Escalabilidad y Paralelización mediante el uso de Hadoop Distributed File System	677
Algoritmos Paralelos y Distribuidos para Cómputo de Altas Prestaciones. Fundamentos , Aplicaciones y Evaluación de rendimiento	681
Arquitecturas Multiprocesador en HPC : Software de Base, Métricas y Aplicaciones	686
Tendencias en Arquitecturas y Algoritmos Paralelos	691
Cloud Computing en HPC	696
Utilizando GPU en la Recuperación de Información Multimedia	700

## Procesamiento de Señales y Sistemas de Tiempo Real

---

Red de nodos sensores y actuadores para control y supervisión de consumo de energía en dispositivos electrodomésticos.	705
Desarrollo e implementación de un sistema informático para el análisis del movimiento humano orientado al estudio de la marcha con procesamiento digital de imágenes	710
Metodologías de Diseño para Sistemas Embebidos	713
Desarrollo de Sistemas de Apoyo para la Toma de Decisiones en Procesos Productivos	718
Aplicación de interfases lectoras de bioseñales en el contexto de la domótica	723
Sistemas de Tiempo Real con Requerimientos Heterogéneos: Integración Hardware - Software	728
Análisis de Perturbaciones de Sistemas de Tiempo Real en Aplicaciones de Control	732
Reconocimiento Automático de Patrones , Análisis de Imágenes y Generación de Características	735
Tópicos de Diseño de Aplicaciones de Tiempo Real	740

# Seguridad Informática

---

Gestión Automática de incidentes e inventarios G.A.I.I.	744
Optimización de un esquema "Occupancy Problem" orientado a E-Voting	749
Aplicación De Sistemas Biométricos En La Administración Pública Local Para Protección De La Información	754
Reconocimiento flexible de formatos de archivos para el descubrimiento de evidencia digital	759
Detección de Anomalías en Oráculos Criptográficos tipo RSA por medio de análisis probabilísticas y estadísticos	764
Inicio de la Línea de Investigación "Ingeniería de Software y Defensa Cibernética"	769

# Tesis Doctorales

---

Identificación temprana de características transversales en el lenguaje de la aplicación capturado con el Léxico Extendido del Lenguaje.	774
Extracción De Conocimiento En Grandes Bases De Datos Utilizando Estrategias Adaptativas	784
Modelo De Proceso De Conceptualización De Requisitos	794
Engineering Accessible Web Applications. An Aspect - Oriented Approach	803

# Agentes y Sistemas Inteligentes

---

Cambio De Creencias: Operadores De Contracción Utilizando Cláusulas De Horn	813
Aplicación De Algoritmos Genéticos A La Corrección De Matrices Inconsistentes En El Proceso Analítico Jerárquico (Ahp)	818
Diseño De Un Sistema Para Alertar Riesgo En El Déficit De Micronutrientes	823
Framework Para Toma De Decisiones Multi - Experto Multi - Granular Con Información Lingüística	828
Desarrollo E Implementación De Prototipos De Robots Móviles Para La Navegación Autónoma Utilizando Técnicas De Inteligencia Artificial	833
Diseño E Implementación De Un Sistema Multisensorial En Un Robot Móvil Autónomo Controlado Mediante Un Sistema Experto Basado En Reglas Y Evaluación De La Navegación En Entornos Semiestructurados	837
Estrategias De Formación De Grupo En Videojuegos Rts Con Ambientes Dinámicos	841
Sistema De Predicción De Incendios Forestales Basado En El Índice Fwi Para La Provincia De Córdoba	845
Métodos, Técnicas Y Herramientas Para La Ingeniería De Software Orientada A Agentes	850
Agentes Inteligentes En Ambientes Dinámicos.	855
Agentes Inteligentes En Ambientes Dinámicos: Análisis De Opinión	860
Prototipo De Sistema De Control Fuzzy Aplicado Al Proceso De Curado De Tabaco Virginia	865
Mejoramiento De Los Servicios De Razonamiento Basados En Programación Lógica Rebatible	870
Metaheurísticas Avanzadas Y De Población Descentralizada Para Problemas De Ruteo De Vehículos.	875
Redefinición De Un Framework Que Formaliza La Argumentación Basada En Expertos	880
Un Enfoque Para El Debilitamiento Y Fortalecimiento Entre Argumentos	885
Integración De Ontologías Datalog +/-	890
Modelo De Servicio De Razonamiento Con Preferencias	895
Evaluación De La Calidad De La Información De Wikipedia En Español	900
Integración De Modelos De Razonamiento Práctico, Tecnologías De Acuerdo Y Aprendizaje	905
Uso De Técnicas Metaheurísticas Avanzadas Para Resolver Problemas De Optimización Combinatoria	910
Agentes Inteligentes En Ambientes Dinámicos: Simplificación De Texto	914
Agentes Bdi Con Función De Utilidad: Un Enfoque Hacia La Opinión Y Confianza En La Toma De Decisiones.	918
Desarrollo De Api De Consulta A Fuentes De Información En La Web Para Sistemas De Argumentación Rebatible	922
Estudio Comparativo De La Eficiencia Entre Controladores Difusos Del Tipo Mandani Y Sugeno "Un Caso	926

De Estudio En La Navegación Autónoma De Robot"	
Técnicas De Soft Computing Aplicadas A Biometría, Predicción Y Ruteo De Vehículos	930
Metaheurísticas Secuenciales Y Distribuidas: Adaptación De Parámetros Y Entornos De Ejecución	935
Argumentación Rebatible En Agentes Adaptativos De Escenarios Dinámicos	940
Desarrollo De Gestores De Noticias Incorporando Técnicas Basadas En La Noción De Confiabilidad	944

## Innovación en Sistemas de Software

---

Propuesta Preliminar De Sistema Experto Para Determinar El Método De Ensayo No Destructivo Aplicado A Una Central Hidroeléctrica.	949
Estudio De Soluciones Tecnológicas Para El Desarrollo De Un Modelo Factible De Participación Ciudadana.	954
Desarrollo De Sistema Experto Generador De Recomendaciones Para La Implementación Sistemas Informáticos De Gestión	959
Mashups En Ambientes Inteligentes	963
Desarrollo De Interface De Detección De Emociones Para Su Utilización En Redes Sociales Y Entornos Virtuales De Aprendizaje	973
Repositorio Ontológico Institucional	978
Conceptualización, Modelado E Implementación Experimental De Un Framework Sensible Al Contexto	982
Avances En El Proyecto Arquitecturas De Control Para Robots Autónomos Móviles Didácticos Basadas En Sistemas Embebidos	987
Sistemas Inteligentes En Arquitecturas De Motores Para Videojuegos	990
Automatización En La Captura De Datos Para El Modelado De Flujo Vehicular En Zonas Urbanas	994
La Gestión Del Conocimientos En Las Organizaciones	999
Integración De Computación Heterogénea Con Hadoop Para Cloud Computing	1004
Integración De Aplicaciones Web A Través De Sindicación De Contenidos	1009
Hacia El Fortalecimiento De La Sociedad En El Uso Y Aplicación De La Información Geospacial Y Las Tic	1014
Administración Y Control De Proyectos Extendiendo El Concepto De Indicadores	1018
Metodología Para El Proceso De Análisis De Adopción De Cloud Computing Para Pymes	1023
Métodos Y Tecnología Informática Aplicada Al Desarrollo De Sistemas De Gerenciamiento Energético En Apoyo A Iso 50001	1027
Sistemas De Información Gerencial En Los Procesos De Negocio De Las Pymes De Jujuy	1032
Arquitecturas Distribuidas Para Gobierno Electrónico	1037

## Tecnología Informática Aplicada en Educación

---

Experiencias Derivadas De La Representación, Modelización Y Construcción De Simuladores En Los Ciclos Lectivos 2011-2012	1041
Tecnología De Computo Ubicua Aplicada A La Educación.	1046
Nuevas Formas De Construcción Del Conocimiento: Redes Sociales.	1051
Indicadores Para La Evaluación Del Trabajo Colaborativo En Línea En El Área De Electrotecnia	1055
Creación De Libros Abiertos En América Latina: Meto Dología, Plataforma Y Estrategias	1060
Educación/Ingeniería De Software	1064
Determinación De Factores Resilientes En Los Alumnos Universitarios	1068
La Autorregulación De Los Aprendizajes Y El Aprendizaje Colaborativo Mediado Por Tic En El Nivel Superior	1073
Aprendizaje Basado En Problemas(Abp) En Los Alumnos De Carreras Técnicas	1077
Aportes Para La Enseñanza De Circuitos RI En Corriente Continua: Análisis De Un Applet	1082
Actualidad De La Capacitación En Inteligencia De Negocios Brindada Por Organizaciones	1087



Latinoamericanas	
Personalización En Entornos De U - Learning	1092
Aprendizaje Organizacional & Aprendizaje Móvil	1096
Redes Sociales: Condiciones Institucionales Y Competencias Tecnológicas Para Su Integración A La Educación En Ámbitos Universitarios	1099
Comunidad De Aprendizaje En Un Contexto Virtual: Red De Docentes De La Fi Unlz	1103
Condiciones Lingüísticas De La Producción Del Conocimiento Con Herramientas De Comunicación Asincrónica: El Foro En Espacios Virtuales De Aprendizaje	1106
M-Learning En La Universidad. Marco De Análisis. Aplicaciones Móviles Como Objetos De Aprendizaje.	1112
Espacios Virtuales Para Trabajo Colaborativo	1116
Laboratorio Virtual Y Remoto Para La Enseñanza De Diseño Y Administración De Redes De Computadoras	1121
Análisis Del Aprendizaje Usando El Nivel De La Incertidumbre En La Adquisición De Conocimiento	1126
Integrando Redes Sociales Y Agentes De Software En Entornos Educativos	1131
Tópicos Avanzados En La Programación De Computadoras	1136
Software Educativo Para La Resolución Numérica Y Gráfica De Integrales Y De Ecuaciones Diferenciales Ordinarias	1140
Gestión De Contenidos En La Era De Las Tic.	1145
Algoritmos. Un Análisis Comparado De Su Enseñanza Teórica Y Práctica En Universidades Del Noa Y Extranjeras.	1149
Televisión Digital Como Plataforma Educativa	1153
Resultados Del Estudio De Requerimientos De La Didáctica Para Una Aplicación De Mundos Virtuales En Educación	1157
Aplicación De Requerimientos Didácticos Y Desarrollo De Software De Mundos Virtuales Con Licencia Open Source	1160
Descifrando El Sustrato Atencional En El Entorno Telemático: ¿Atención Digitalizada?	1165
El Desafío De La Nuevas Tecnologías De La Información Y Las Comunicaciones En Los Contextos Educativos	1169
Innovación En Los Ambientes Virtuales De Aprendizaje: Las Tic Y Su Relación Con Los Sistemas Ubicuos	1176
Unnoba Blogs: Blogs De Asignaturas Y Contenidos Abiertos.	1181
Entornos Virtuales 3d Como Herramienta Para Promover El Trabajo Colaborativo En Educación.	1185
Aplicaciones De Las Tic En Educación Y Ciencias	1191
Bigdata En La Educación	1196
Objetos De Aprendizaje En Ambientes Centrados En El Alumno	1199
Taller Revisión De Temas De Matemática, Modalidad E-Learning Y Presencial.	1203
Identificación De Las Estrategias De Uso De Recursos Y Actividades En Las Aulas Virtuales En La Facultad De Ciencias Económicas De La Unpsjb	1208
Metodología De Desarrollo De Herramientas Informáticas Didácticas Para El Aprendizaje Del Cálculo Diferencial E Integra	1214
Un Abordaje De La Enseñanza De La Materia Inteligencia Artificial Desde Los Ambientes De Aprendizaje Colaborativos En La Carrera De Isi Utn Frc	1220
Tecnologías De La Información Y La Comunicación Aplicadas A La Formación De Profesionales En Turismo Rural	1225
Tics Aplicadas A Problemas De Gobierno Electrónico Y De E - Learning	1230
Aplicaciones De La Robótica En La Integración De Contenidos En Carreras Técnicas	1235
Tic, Gc Y E-Educación: Generación De Sistemas Informáticos.	1240
Tecnologías De La Información Y La Comunicación En Ámbitos Educativos : Análisis De Experiencias Y Resultados De Su Aplicación	1245
Las Tecnologías De La Información Y La Comunicación En El Ámbito De La Educación Especial.	1251
Plataforma De Hardware De Bajo Costo Para Robótica Educativa Resultados De La Primera Etapa	1256
Difusión De La Actividad Académica A Través De La Integración De Entornos Virtuales De Aprendizaje Con Redes Sociales, Sistemas Académicos Y Repositorios Digitales	1259
Metodologías De Enseñanza Interactiva Para Entornos Virtuales	1264

# Un mecanismo de IPC de microkernel embebido en el kernel de Linux

**Pablo Pessolani**

Departamento de Ingeniería en Sistemas de  
Información  
Facultad Regional Santa Fe  
Universidad Tecnológica Nacional  
Santa Fe, Argentina  
ppessolani@frsf.utn.edu.ar

**Silvio Gonnet**

CIDISI- Instituto de Desarrollo y Diseño  
(INGAR), CONICET  
Universidad Tecnológica Nacional  
Santa Fe - Argentina  
sgonnet@santafe-conicet.gov.ar

**Toni Cortes\***

Barcelona Supercomputing Center y  
Departamento de Arquitectura de Computadores  
Universitat Politècnica de Catalunya  
Barcelona - España  
toni.cortes@bsc.es

**Fernando G. Tinetti#**

III-LIDI  
Facultad de Informática  
Universidad Nacional de La Plata  
La Plata, Argentina  
Comisión de Inv. Científicas, Prov. Bs. As.  
fernando@info.unlp.edu.ar

## RESUMEN

Existe una marcada tendencia en la industria de comercializar procesadores con múltiples núcleos, conocidos como multi-cores. Se prevé que en el mediano plazo la cantidad de núcleos aumente significativamente hasta miles de núcleos por procesador [1] denominados many-cores. Esta tendencia requiere de sistemas operativos (OS) que puedan aprovechar éstas tecnologías y que adapten su funcionamiento de tal forma de que no se transformen en el cuello de botella de las aplicaciones que ejecutan sobre ellos como consecuencia de la contención de los recursos que comparten [2].

Los OS basados en microkernel, los OS multi-kernel, algunos exokernels, y ciertas tecnologías de virtualización ofrecen ventajas significativas en su diseño para su adaptación a sistemas many-cores. Estas arquitecturas requieren de un mecanismo de transferencia de mensajes para comunicar entidades tales como procesos, hilos, kernels o máquinas virtuales.

En este artículo se presentan el trabajo de investigación y desarrollo de un mecanismo de IPC basado en Minix 3 (denominado M3-IPC) embebido dentro del kernel de Linux. M3-IPC permite incorporar servicios de un OS basado en microkernel (Minix) dentro de un OS monolítico (Linux) conformando un sistema híbrido donde pueden convivir aplicaciones y servicios de ambas arquitecturas y adaptarlos de manera no abrupta a los sistemas many-cores.

**Palabras claves:** IPC, many-cores, microkernel.

## CONTEXTO

Este artículo forma parte de los trabajos de investigación y desarrollo del proyecto presentado en WICC 2012 [3], denominado: “*Sistema de Virtualización con Recursos Distribuidos*” (en inglés DRVS). Este proyecto de I/D involucra a investigadores de varios laboratorios y centros de investigación (ver afiliaciones de los autores), en un área que por su amplitud requiere de múltiples enfoques. Más específicamente, se deriva de las tareas de investigación en el contexto de los programas de postgrado de la Facultad de Informática de la UNLP.

## INTRODUCCION

La tendencia en la industria muestra que los computadores de uso general dispondrán de cientos e incluso miles de núcleos procesadores. En su gran mayoría, el software desarrollado se basa en un modelo de memoria compartida consistente. Para mantener la consistencia de memoria entre los distintos núcleos, los sistemas multi-core utilizan protocolos de invalidación de cache o de actualización de cache. A medida que aumenta el número de núcleos, estos protocolos presentan una mayor sobrecarga y latencia lo que los hace poco escalables. Por esta razón, tal como se argumenta en [4], el hardware futuro se parecerá más a un sistema distribuido unido por una red interna dentro del mismo chip que a un sistema de memoria compartida [5].

Los OS basados en microkernel como Minix[6] o L4 [7], el exokernel Corey[8], los OS multi-kernel como Barrelfish [4], Nix [9] o FOS [10] y algunas tecnologías de virtualización [11] ofrecen ventajas significativas en su diseño para su adaptación a sistemas many-cores. Estas arquitecturas utilizan la transferencia de mensajes como mecanismo para comunicar entidades tales como procesos, hilos, kernels, hipervisores o máquinas virtuales. Los

\* Este trabajo está siendo subvencionado parcialmente por parte del Ministerio de Ciencia y Tecnología de España por la ayuda TIN2007-60625 y del Gobierno Catalán por la ayuda 2009-SGR-980

# Investigador CIC Provincia de Buenos Aires

sistemas de microkernel como FOS[10] asignan un núcleo o core para cada servidor o device driver por lo que se evita el cambio de contexto del emisor al receptor durante la transferencia de mensajes. Los sistemas multikernel, asignan un kernel a cada núcleo comunicándose entre sí mediante IPC.

Se hace evidente la importancia y criticidad que representan para estos sistemas los mecanismos de IPC que utilizan.

## LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

La línea de I/D de este proyecto refiere a desarrollar un modelo de DRVS [3]. Un DRVS es un Sistema de Virtualización con Recursos Distribuidos que comparte características con los Sistemas de Imagen Unica (SSI), con la virtualización basada en OS y con los OS de microkernel multiservidores. Las aplicaciones y procesos servidores se ejecutan en un entorno o compartimento que se denomina Máquina Virtual (VM). A diferencia de otras tecnologías de virtualización estas VMs no están confinadas a un solo nodo, sino que pueden abarcar un subconjunto de nodos. Para el DRVS un nodo es un computador de un cluster o un núcleo de un sistema many-cores.

Dado que los procesos de una VM se comunicarán mediante transferencia de mensajes, son fundamentales la robustez, sencillez y rendimiento del mecanismo de IPC. Por sus características, el DRVS requiere comunicar procesos co-residentes en el mismo nodo (locales) y procesos residentes en diferentes nodos (remotos).

El proyecto del DRVS establece como requerimiento para el mecanismo de IPC su compatibilidad con las APIs de Minix 3, de allí surge la denominación de M3-IPC.

La decisión de crear un nuevo mecanismo de IPC dentro del kernel de Linux y no utilizar los mecanismos de IPC existentes se debió básicamente a que estos últimos no ofrecen la totalidad de las características requeridas por el proyecto de I/D:

- *Messages Queues, Pipes, FIFOs, Unix Sockets*: No disponen de la posibilidad de comunicarse con procesos remotos de forma transparente y uniforme.
- *RPC, UDP/TCP Sockets*: Tienen gran latencia en la inicialización de los canales de comunicación y bajo rendimiento de las transferencias entre procesos co-residentes (ver [Micro-benchmarks](#)).

Para el diseño de los mecanismos de IPC del DRVS se establecieron los siguientes principios:

- *Simplicidad*: Debe proveerse un número pequeño de APIs básicas, de rápida comprensión por parte de los programadores y que permita una sencilla traza de mensajes para la depuración de las aplicaciones.
- *Transparencia*: Las APIs de IPC deben ser transparentes a la ubicación de los procesos, es decir, no debe haber distinción entre procesos locales y remotos a fin de facilitar la programación.
- *Aislamiento*: Debe soportar VMs como compartimentos de procesos de tal forma que solo se admitirá la

comunicación entre procesos pertenecientes a la misma VM.

- *Escalabilidad*: Debe aprovechar eficientemente los recursos de un cluster o de un sistema many-cores.
- *Rendimiento*: El rendimiento en computadores de pocos núcleos debe ser equivalente a los mecanismos de IPC más veloces de los que se dispone en Linux.

Un DRVS es un contenedor de VMs que abarca a todos los nodos (computadores de un cluster o núcleos de un sistema many-cores). Una VM es un contenedor de procesos que abarca a un subconjunto de nodos, por lo que los procesos de una VM pueden estar ejecutando en diferentes nodos. Un nodo puede ser compartido por diferentes VMs, y por lo tanto pueden ejecutarse en él procesos de diferentes VMs, pero los mismos no tienen posibilidad de comunicarse entre sí (aislamiento).

En la primera etapa (que es la que se presenta en éste artículo) se procedió al diseño, desarrollo, implementación y pruebas de rendimiento de un mecanismo de IPC para comunicar procesos locales, es decir que se encuentran compartiendo un nodo. Para esta etapa, se estableció como objetivo desarrollar el software de IPC, programas de prueba y micro-benchmarks sobre Linux para plataforma x86.

Si bien el objetivo primario es satisfacer los requerimientos del proyecto, también se pretende brindar una nueva herramienta de comunicación para Linux que facilite nuevos desarrollos, como por ejemplo un microkernel como módulo instalable de un OS co-residente con el propio Linux de forma equivalente a como lo hacen algunos sub-kernels de tiempo real tales como RTLinux [12] o RTAI [13].

La configuración de los parámetros de operación del DRVS (número de VMs, número de nodos, etc.) puede hacerse en forma dinámica especificándolos como argumentos del comando de inicialización u obtenerse a partir de registros de un directorio LDAP.

Para poder utilizar M3-IPC el DRVS debe estar inicializado. Como el DRVS estará conformado por N nodos, debe asignársele un *nodeid* a cada nodo durante la inicialización. Esto puede hacerse directamente al iniciar Linux incluyendo el *nodeid* en la línea de parámetros de booteo, en un script de arranque, directamente ingresando el comando en forma manual o como parte de un programa invocando a *mx\_drvs\_init()*.

Luego de inicializado el DRVS, se deben inicializar las VMs. Cada VM tiene asignado un *vmid*, una serie de parámetros de configuración, los nodos que abarca y un nombre que facilita su identificación a los administradores del DRVS. Los parámetros de configuración de las VMs pueden establecerse por línea de comandos, como parte de un programa utilizando *mx\_vm\_init()*, o a partir de registros de un directorio LDAP.

Los procesos Linux ordinarios no pueden utilizar M3-IPC, previamente deben registrarse a una VM mediante *mx\_bind()*. Un proceso solo puede pertenecer a una VM. A cada proceso registrado se le asigna un identificador denominado *endpoint* que es único dentro de la VM y que será utilizado en las primitivas de IPC como origen/destino de los mensajes y demás APIs. Luego de registrarse se establece una relación biunívoca entre el PID de Linux y el

*endpoint* de una VM (PID $\leftrightarrow$ (vmid, endpoint)) y el proceso queda habilitado para utilizar las APIs de M3-IPC refiriéndose a los otros procesos mediante sus *endpoints* (sin mencionar el *vmid*, ya que es idéntico al propio).

Los mensajes en M3-IPC, al igual que en Minix, tienen diferentes formatos y tamaño fijo de 36 bytes para una plataforma x86 de 32 bits. La copia de bloques de datos entre procesos admite tamaños de bloques de hasta 4 MB.

La de-registración de un proceso de una VM se hace en forma explícita mediante *mnx\_unbind()*. Si el proceso finaliza en forma normal con *exit()* o porque ha recibido una señal que lo finaliza sin realizar *mnx\_unbind()*, el sistema lo hace en forma implícita. Si el proceso finalizado esperaba enviar mensajes a otro proceso, u otros procesos esperaban recibir mensajes de éste, los códigos de retornos de las APIs le informarán que el *endpoint* ha finalizado.

Para finalizar una VM se utiliza *mnx\_vm\_end()*, que envía un SIGPIPE a todos los procesos registrados en esa VM, y luego retorna sus recursos (memoria) a Linux.

Para finalizar el DRVS se utiliza *mnx\_drvs\_end()*, que finaliza todas las VMs, y por consecuencia todos los procesos pertenecientes a las VMs.

Las siguientes son las APIs para la transferencia de mensajes y copias de bloques de datos:

- *mnx\_send()*: Envía un mensaje de tamaño fijo desde el espacio de direcciones del proceso emisor al espacio de direcciones del proceso receptor. Si el proceso receptor no está esperando el mensaje, el emisor se bloquea.
- *mnx\_receive()*: Recibe un mensaje en el espacio de direcciones del proceso emisor. Si no hay pendiente de entrega un mensaje requerido, el receptor se bloquea.
- *mnx\_sendrec()*: Envía un mensaje a un proceso y espera por un mensaje de respuesta. El proceso permanece bloqueado hasta recibir la respuesta.
- *mnx\_notify()*: Envía un mensaje de notificación (sin contenido, similar a una señal o signal) a un proceso. Si el proceso receptor no está esperando dicho mensaje, queda registrada la notificación y el proceso receptor la recibirá cuando ejecute el próximo *mnx\_receive()*.
- *mnx\_vcopy()*: Permite copiar bloques de datos entre los espacios de direcciones de dos procesos.

El mecanismo de IPC resultante, su rendimiento y la sencillez de sus APIs permiten construir nuevos servicios equivalentes a los brindados en los OSs de microkernel en un OS monolítico como es Linux.

## Trabajos Relacionados

Existen proyectos similares a M3-IPC. Estos son:

- *Synchronous Interprocess Messaging Project for LINUX (SIMPL) [14]*: Es un proyecto open source para Linux que permite disponer de las APIs estilo QNX de transferencia síncrona de mensajes. SIMPL es una librería que utiliza FIFOs para las transferencias locales y TCP para las transferencias remotas. El rendimiento es bastante pobre dado que utiliza *pipes* para la sincronización de procesos y *shared memory (shmem)* para la copia de mensajes y datos.

- *Send/Receive/Reply (SSR)[15]*: Es una implementación de las APIs de QNX4 en Linux. A diferencia de SIMPL, SSR es un módulo instalable de kernel. La comunicación entre los procesos de usuario y el módulo de kernel se realiza utilizando operaciones IOCTL de un archivo de dispositivo.

## RESULTADOS Y OBJETIVOS

En esta sección se resumen las [Características de M3-IPC](#) detallando algunos aspectos referentes al diseño, y a fin de contrastar resultados con los objetivos de rendimiento establecidos para esta etapa del proyecto se realizaron una serie [Micro-benchmarks](#). Estos tests permiten comparar el desempeño de M3-IPC frente a otros mecanismos equivalentes disponibles en Linux, y frente al IPC de Minix en dos versiones que utilizan diferentes tecnologías de administración de memoria.

### Características de M3-IPC

Uno de los objetivos de M3-IPC es brindar un mecanismo de transferencia de mensajes entre procesos sencillo y fácil de comprender construido en base a componentes de software existentes en Linux.

Las características del M3-IPC resultante de esta etapa del proyecto de I/D son las siguientes:

- Es un mecanismo de IPC sincrónico sin utilización de buffers en el kernel y compatible con APIs de Minix 3.
- Utiliza el mismo vector interrupciones que Minix, que es diferente al utilizado por Linux para las llamadas al sistema. Esto permite que ambas APIs convivan en el mismo kernel y que aplicaciones desarrolladas para Minix seguir utilizando el mismo vector. Para habilitar este nuevo vector de interrupción se debió modificar el kernel de Linux, lo que seguramente dificultará el mantenimiento de M3-IPC en kernels futuros.
- Permite a los programadores configurar VMs para aislar el IPC entre sus aplicaciones.
- A fin de aprovechar el paralelismo en sistemas SMP y many-cores:
  - Las APIs soportan hilos (*threads*) y utilizan *mutexes* del kernel de Linux para realizar la exclusión mutua.
  - Se maximizó la granularidad de las regiones críticas a nivel de procesos utilizando *spinlocks* y *mutexes*. Esto permite el paralelismo de transferencias de mensajes entre múltiples pares de procesos diferentes.
  - La utilización de VMs permite el paralelismo de transferencias de mensajes dentro de VMs diferentes.
- Las estructuras de datos de los procesos registrados en una VM (*struct proc*), se encuentran alineadas con las líneas de cache L1 de CPU a fin de reducir el tiempo de acceso.
- Utiliza *Task Queues* y el mecanismo de *Event waiting* de Linux para la sincronización de procesos e hilos.
- Utiliza *Reference Counters* del kernel de Linux para controlar el número de procesos registrados en una VM.
- La copia de bloques de datos entre procesos y la transferencia de mensajes se realiza desde el espacio de



direcciones del proceso origen al espacio del proceso destino sin copia intermedia en un buffer del kernel.

- La copia de bloques de datos distingue entre la copia de bytes y la copia de páginas (más eficiente).
- La información del sistema M3-IPC, el DRVS, sus VMs, nodos y procesos se encuentra integrada al sistema de archivos de Linux */proc*.
- Desarrollado inicialmente para plataforma Linux x86 en lenguaje C como Open Source.

### Micro-benchmarks

Uno de los principios de diseño establece que el rendimiento esperado en computadores de pocos núcleos debe ser equivalente a los mecanismos de IPC mas veloces de los que se dispone en Linux.

A fin de evaluar el rendimiento de M3-IPC se prepararon micro-benchmarks sobre Linux 2.6.32 x86 de 32 bits para M3-IPC y para otros mecanismos de IPC tales como:

- *Message Queues*: Se realizaron benchmarks propios.
- *FIFOs, pipes, Unix Sockets, TCP Sockets*: Se utilizó *ipc-bench* [16] (modificado).
- *SRR, SIMPL*: Se utilizaron los benchmarks que forman parte de la distribución.

En todos los casos, los programas de pruebas fueron realizados o modificados de tal forma que las transferencias de mensajes y copias de bloques de datos se asemejen a la semántica M3-IPC.

Por otro lado se realizaron micro-benchmarks para Minix 3.1.2 que utiliza administración de memoria segmentada y Minix 3.2.0 que utiliza administración de memoria por paginación. En ambas versiones debieron alterarse ligeramente los microkernels y el administrador de procesos (PM) para que admita la realización de tests equivalentes a los de M3-IPC. Es importante aclarar que Minix 3.1.2 soporta sistemas monoprocesadores y Minix 3.2.0 se encuentra en etapa experimental de soporte SMP.

El equipo utilizado para realizar los micro-benchmarks es un dual-core Intel Core 2 E7400 @ 2.80GHz con cache de 3072 kB y línea de cache de 64 bytes.

### Transferencias de Mensajes

Las pruebas de rendimiento de transferencia de mensajes fueron denominadas:

- *IPC1*: Un proceso cliente envía un mensaje con *mnx\_send()* a un proceso servidor que espera con un *mnx\_receive()*. El servidor responde con un *mnx\_send()* al cliente que espera con un *mnx\_receive()*.
- *IPC2*: Un proceso cliente envía un mensaje con *mnx\_sendrec()* a un proceso servidor y espera por la respuesta. El servidor espera la petición del cliente con un *mnx\_receive()* y le responde con un *mnx\_send()*.
- *IPC3*: Un proceso cliente envía un mensaje con *mnx\_notify()* a un proceso servidor y continúa su ejecución hasta hacer un *mnx\_receive()* esperando por la respuesta. El servidor que espera con un *mnx\_receive()* responde con un *mnx\_notify()* al cliente.

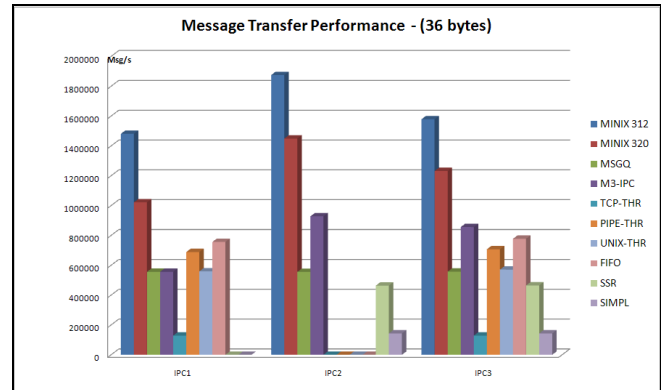


Figura 1. Rendimiento Comparativo de Transferencia de Mensajes.

En la [Figura 1](#) se presentan los resultados comparativos para los tres tests de transferencia de mensajes. A fin de realizar los tests equitativos entre Linux y Minix, se anuló en Linux el segundo núcleo de la plataforma dual-core. Se puede apreciar el rendimiento marcadamente superior de las dos versiones de Minix frente a todas los mecanismos de IPC de Linux. Esto básicamente se debe a la sencillez del microkernel de Minix y la complejidad del kernel de Linux (mayor camino crítico). Aún así, M3-IPC muestra el máximo desempeño en dos de los tres tests frente al resto de los mecanismos de IPC de Linux. Cuando se habilitaron los dos núcleos y múltiples pares cliente/servidor, el rendimiento fue de 1.460.000 [Msg/s] para IPC2.

Algunos tests no fueron realizados para ciertos mecanismos de IPC (valores cero en [Figura 1](#)) porque no era posible reproducir una semántica equivalente. Para el caso de IPC3 donde se utiliza *mnx\_notify()*, se asumió que la transferencia de 1 byte podría considerarse equivalente.

### Copia de Bloques de Datos

Previo a la realización de las pruebas de copia de bloques de datos entre procesos se realizó el test *lmbench*[17] que brinda información acerca del rendimiento de copia de datos dentro del espacio de direcciones del mismo proceso (*bw\_mem*). El rendimiento de *bm\_mem* fue de 16,7 [GB/s] para copia de bloques de 4 kB, estableciendo el límite máximo de rendimiento de cualquier mecanismo de copia de bloques de datos entre diferentes procesos.

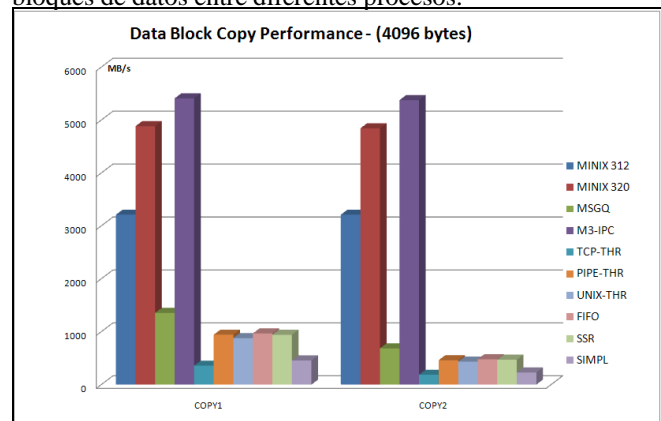


Figura 2. Rendimiento de Copia de Bloques de Datos.

Los test de copias de bloques de datos fueron denominados:

- **COPY1:** Copia un bloque de datos desde el espacio direcciones del proceso que invoca `mnx_vcopy()` hacia el espacio de direcciones de otro proceso.
- **COPY2:** Copia un bloque de datos desde el espacio direcciones de un proceso origen, hacia el espacio de direcciones de otro proceso destino. Un tercer proceso `requester` es el que invoca a `mnx_vcopy()`.

Como muestra la [Figura 2](#), el rendimiento de M3-IPC es máximo, incluso superando a ambas versiones de Minix. Tanto Minix como M3-IPC solo realizan una copia de datos entre el espacio de direcciones del proceso origen al espacio de direcciones del proceso destino. Los otros mecanismos de Linux realizan una doble copia (Origen→Kernel→Destino).

El rendimiento superior de M3-IPC se debe a que utiliza una optimización que ofrece el kernel de Linux que permite copiar páginas completas y que no requiere de cambios de contextos entre los procesos. Esta copia se lleva a cabo en una función que utiliza instrucciones MMX, en cambio Minix utiliza la instrucción MOVSB menos eficiente.

Se deben considerar algunos aspectos referentes a la copia de bloques de datos que pueden alterar su rendimiento:

- **Tamaños de las Cache:** El tamaño de cache L1 para el hardware utilizado es de 32 kB con un tiempo acceso reportado por `lmbench` de 1.07[ns], el tiempo de acceso a cache L2 es de 5.41[ns] y a cache L3 es de 11[ns]. El test `bw-mem` reportó que hay una reducción significativa (~50%) en el rendimiento cuando el tamaño de conjunto de trabajo (2 veces el tamaño de bloque) es mayor al tamaño de cache L1.
- **Alineación de los bloques de datos con respecto a las páginas:** Si los bloques de datos a copiar están alineados con el tamaño de página, se utiliza la copia de páginas completas que es un mecanismo más eficiente debido a utiliza instrucciones MMX como se indicó anteriormente.
- **Alineación de los bloques de datos origen y destino entre sí:** Si la dirección de inicio del bloque de datos origen se encuentra desplazada `src_off` bytes de la alineación de página, y la dirección de inicio del bloque de datos destino se encuentra desplazada `dst_off` bytes, si `src_off ≠ dst_off`, se duplica el número de ciclos de copia que hay que realizar y se impide realiza copias de páginas completas debiéndose realizar copias de bytes.

## Conclusiones

Como resultado de la primera etapa del proyecto de I/D se obtuvo un mecanismo de IPC apto para la transferencia de mensajes entre procesos de un mismo nodo, con un rendimiento similar, incluso mejorado, comparado con otros mecanismos de IPC de Linux.

M3-IPC ha sido desarrollado siguiendo las recomendaciones para su operación eficiente en sistemas multi-cores y many-cores, pero dado que utiliza las facilidades que ofrece el kernel de Linux (mutexes, spinlocks, reference counters, etc.) su escalabilidad está limitada al rendimiento que éstos presentan operando sobre ese tipo de sistemas.

## FORMACIÓN DE RECURSOS HUMANOS

El trabajo central de la línea de I/D presentada se plantea en principio como una tesis de doctorado, aunque por la amplitud del tema, seguramente dejará abiertas muchas cuestiones que podrán ser tratadas en tesinas de grado, tesis de magister/doctorado o en trabajos de otros investigadores.

En particular, el mecanismo de IPC admite, por ejemplo, la investigación los siguientes tópicos:

- **Seguridad:** M3-IPC no incluye aspectos relativos a seguridad. Podría integrarse con las *Capabilities* de Linux.
- **Integración con Linux Containers(LXC)[18]:** Las VMs de M3-IPC podrían integrarse al mecanismo de contenedores o compartimentos LXC disponible en Linux.

## REFERENCIAS

1. Shekhar Borkar. "Thousand core chips: a technology perspective". In DAC '07 ACM, pages 746–749, 2007.
2. Siddha, S.; Pallipadi; A Mallick. "Process scheduling challenges in the era of multi-core processors". Intel Technology Journal, 2007.
3. Pablo Pessolani, Toni Cortes, Silvio Gonnet, Fernando G. Tinetti. "Sistema de Virtualización con Recursos Distribuidos". WICC 2012. Pág. 59-43. Argentina, 2012.
4. Andrew Baumann, Paul Barham, Pierre-Evariste Dagand, Tim Harris, Rebecca Isaacs, Simon Peter, Timothy Roscoe, Adrian Schüpbach, and Akhilesh Singhanian. "The Multikernel: A new OS architecture for scalable multicore systems". In Proceedings of the 22nd ACM Symposium on OS Principles, 2009.
5. Sriram Vangal, et al. "An 80-Tile Sub-100-W TeraFLOPS Processor in 65-nm CMOS," 1. IEEE Journal of Solid-State Circuits, Vol. 43, No. 1, Jan 2008.
6. Tanenbaum A., Woodhull A., "Operating Systems Design and Implementation, Third Edition", Prentice-Hall, 2006.
7. Liedtke J.; "On  $\mu$ -kernel construction"; Symposium on Operating System Principles ACM, 1995.
8. Silas Boyd-Wickizer, Haibo Chen, Rong Chen, Yandong Mao, Frans Kaashoek, Robert Morris, Aleksey Pesterev, Lex Stein, Ming Wu, Yuehua Dai, Yang Zhang, and Zheng Zhang. "Corey: an operating system for many cores". In Proceedings of OSDI'08. USENIX Association, 43-57. 2008.
9. Francisco J. Ballesteros, Noah Evans, Charles Forsyth, Gorka Guardiola, Jim McKie, Ronald Minnich, Enrique Soriano-Salvador. "NIX: A Case for a Manycore System for Cloud Computing". Bell Labs Technical Journal 17(2): 41-54 (2012).
10. Matarneh, R., 2009. "Multi Microkernel Operating Systems for Multi-Core Processors". JCS, 5: 493-500.
11. Carl Gebhardt and Allan Tomlinson, "Challenges for Inter Virtual Machine Communication", Technical Report, 2010.
12. M. Barabanov and V. Yodaiken. "Real-Time Linux". Linux journal, February 1997.
13. E. Bianchi and L. Dozio. "Some experiences in fast hard-real time control in user space with RTAI-LXRT". In Workshop on Real Time Operating Systems and Applications and 2nd Real Time Linux Workshop, November 2000.
14. John Collins and Robert Findlay, "Programming the SIMPL Way", Lulu.com, ISBN 0557012708, 2008.
15. SRR- QNX API compatible message passing for Linux. <http://www.opcdatahub.com/Docs/books.html>
16. ipc-bench: A UNIX inter-process communication benchmark. <http://www.cl.cam.ac.uk/research/srg/netos/ipc-bench/>
17. Carl Staelin and Hewlett-packard Laboratories, "lmbench: Portable Tools for Performance Analysis", In USENIX Annual Technical Conference, 279-294, 1996.
18. LXC Linux Containers, <http://lxc.sourceforge.net/>



## Análisis de Prestaciones de Tráfico Multicast en Redes Mixtas IPv4 e IPv6

Higinio Facchini, Santiago Pérez, Alejandro Dantiacq, Gastón Cangemi  
GRID TICs (Grupo UTN de Investigación y Desarrollo en TICs)  
Laboratorio de Análisis de Tráfico y Seguridad  
Facultad Regional Mendoza, Universidad Tecnológica Nacional  
Rodríguez 273, Mendoza, Argentina  
0261-5244576  
@frm.utn.edu.ar

### Resumen

Las comunicaciones IP tradicionales permiten que un host (estación o puesto de trabajo) pueda enviar paquetes a otro host (transmisiones unicast o unidifusión) o a todos los hosts (transmisiones broadcast). Multicast o multidifusión IP ofrece una tercera alternativa de comunicaciones: permite que un host pueda enviar los paquetes a un grupo que está compuesto por un subconjunto de los hosts de la red.

Multicast IP es un tecnología para conservar el ancho de banda, específicamente diseñada para reducir el tráfico, transmitiendo un único flujo de información potencialmente a miles de destinatarios. De esta forma, se sustituyen las múltiples copias para todos los beneficiarios con la entrega de un único flujo de información. Por lo tanto, la multidifusión IP es capaz de reducir al mínimo la carga, tanto en los hosts origen y destino, y simultáneamente el tráfico total de la red.

Dentro de una red multicast, los routers son los responsables de replicar y distribuir el contenido de multidifusión a todos los hosts que están escuchando a un determinado grupo multicast. Los routers emplean protocolos multicast que construyen árboles de distribución para transmitir el contenido multicast, que aseguran la mayor eficiencia para el envío de datos a múltiples receptores. Cualquier alternativa multicast IP requiere que la fuente envíe más que una copia de los datos. El unicast tradicional a nivel de aplicación, por ejemplo, requiere que el

origen transmita una copia para cada receptor del grupo.

En el presente trabajo de investigación, se pretende realizar un análisis cualitativo y cuantitativo de tráfico multicast, tanto en los protocolos IPv4 como IPv6, sobre redes mixtas, midiendo la performance en una red de laboratorio, utilizando equipamiento real, combinado con generadores de tráfico real y sintético.

**Palabras claves:** tráfico multicast, protocolos multicast, routing multicast

### Contexto

La línea de investigación está inserta en el proyecto homólogo, en el ámbito del Laboratorio de Análisis de Tráfico y Seguridad (Latys) del Grupo UTN GRID TICs (GRUPO UTN DE INVESTIGACION Y DESARROLLO EN TICs), del Departamento Ingeniería en Electrónica, de Facultad Regional Mendoza, de la Universidad Tecnológica Nacional. Y además, está relacionado con el trabajo de tesis de maestría del Ing. Higinio Facchini, en la Facultad de Informática, de la Universidad Nacional de La Plata.

### Introducción

Hoy en día más y más aplicaciones de red requieren el reenvío de los paquetes de una o más fuentes a un grupo de receptores. Estas aplicaciones comprenden la mayoría de la transferencia de datos (por ejemplo, la entrega de actualizaciones de software de los

desarrolladores para los usuarios finales), la transmisión de medios (audio, vídeo, texto), el intercambio de datos (por ejemplo, una videoconferencia entre compartida distribuida participantes), la entrada de datos (por ejemplo, acciones, participaciones), la caché de Web y actualización de los videojuegos interactivos (por ejemplo, distribuidos entornos virtuales o juegos multijugador como Quake). Una abstracción muy útil en representación de cada una de estas aplicaciones es el concepto de multicasting.

Casi todas las comunicaciones en Internet (o en una red empresarial) hoy en día son unicast. En el nivel IP, cada paquete enviado se reenvía al host de destino identificado por la dirección IP de destino en la cabecera del paquete IP. Los enrutadores IP se especifican en las tablas de enrutamiento para reenviar paquetes basados en la dirección de destino. Además de unidifusión, también hay multidifusión. Para multicast, la dirección IP de destino se refiere a un grupo de hosts IP: la idea es que un paquete enviado a la dirección de grupo de multicast, debe llegar a todos los hosts del grupo.

El Protocolo de Internet versión 6 (IPv6) es la nueva generación del protocolo básico de Internet. La versión actual de IP (IP versión 4) tiene algunos inconvenientes que complican y, en algunos casos presentan una barrera para el desarrollo futuro. IPv6 debería eliminar estas barreras y proporcionar un ambiente rico en funciones para el futuro de la interconexión mundial. Con el avance masivo del tráfico de voz, video y de aplicaciones de tiempo real, se espera que las aplicaciones utilicen el tipo de tráfico multicast como un ahorro de ancho de banda en los enlaces troncales de la red, como así en todo punto de posible congestión de tráfico.

Además en las redes inalámbricas como 802.11, el rendimiento todavía es muy bajo, y decae fuertemente en la medida que se incrementa el número de usuarios, por lo que el ahorro de ancho de banda es fundamental.

## Ambientes multicast

Las soluciones multicast ofrecen ventajas relativas a la conservación del ancho de banda de la red. En el caso de una aplicación de gran ancho de banda, como vídeo MPEG, el multicast IP puede beneficiar aún en las situaciones que tengan sólo unos pocos receptores, dado que escasas secuencias de vídeo podrían consumir una gran parte del ancho de banda disponible en la red (Figura 1). Incluso en aplicaciones de bajo requerimiento de ancho de banda, multicast IP conserva los recursos cuando las transmisiones involucran miles de receptores.

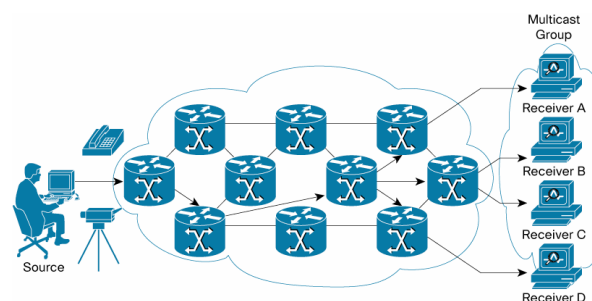


Figura 1. Ambiente multicast

Además, multicast IP es la única alternativa no broadcast para situaciones que requieren la transmisión de información simultánea a más de un receptor.

Para aplicaciones demandantes de bajo ancho de banda, una alternativa a multicast IP podría ser replicar los datos el origen. Esta solución, sin embargo, puede deteriorar las prestaciones de la aplicación, introducir latencias y retardos variables que afectarán a los usuarios y las aplicaciones, y requieren servidores costosos para administrar las réplicas y la distribución de los datos. Tales soluciones también involucran transmisiones múltiples del mismo contenido, y el consumo de una enorme cantidad de ancho de banda de la red. En las aplicaciones demandantes de mayores anchos de banda, multicast IP es la única opción viable.

## Aplicaciones multicast

Las aplicaciones multicast pueden clasificarse en aplicaciones de tiempo real y de no tiempo real. En las primeras se encuentran la

aplicaciones multimediales como televisión IP, video en vivo, videoconferencias, audio de Internet en vivo, y las aplicaciones de sólo datos como el flujo de noticias por Internet, juegos interactivos, entre otras. En las segundas se encuentran las aplicaciones multimediales como replicación de video, servidores web y transmisiones de contenido, y las aplicaciones sólo de datos como transmisión de información, servidor a servidor, servidor a estación, replicación de bases de datos y distribución de software. Otras aplicaciones que toman ventaja de multicast IP son:

- Comunicaciones corporativas,
- Transmisión de canales de música, y
- Educación a distancia o e-learning.

## Despliegue y tecnologías

Multicast IP esta soportado en:

- Redes IPv4
- Redes IPv6
- VPNs MPLS
- Redes móviles e inalámbricas

En la última década pasada, la adopción de multicast IP en aplicaciones empresariales y del sector público ha crecido enormemente (Figura 2), y los proveedores de servicios han respondido añadiendo más VPN multicast a sus carteras de servicios. Hoy en día, cualquier proveedor de servicios debe desplegar multicast IP para permanecer competitivo.

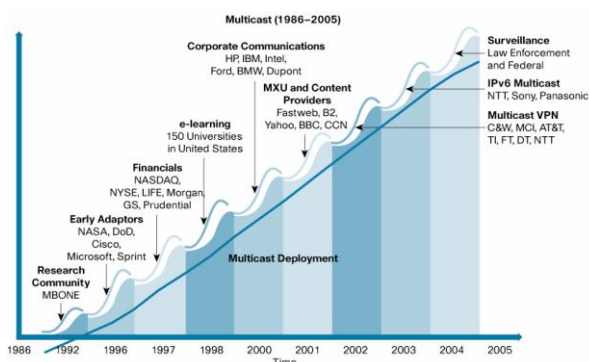


Figura 2. Crecimiento multicast

El despliegue de los servicios de video ofrece más incentivos para el fortalecimiento de la plataforma multicast IP de un proveedor de

servicios, ya que los vuelve más eficientes, en función de los costos medios de soportar tráfico de datos, de voz y de video.

## Grupos y alta disponibilidad

Las redes que utilizan multicast IP entregan contenido desde un origen a múltiples usuarios (hosts o receptores) que están interesados en el flujo de datos. Un canal multicast se refiere a la combinación de un contenido y la dirección IP de origen, con la dirección del grupo multicast IP donde el contenido se está difundiendo.

A diferencia de las direcciones unicast y broadcast, los grupos multicast no tienen ningún límite físico o geográfico, y los interesados en unirse a receptores pueden estar situados en cualquier punto de una red o de Internet, en la medida en que el camino multicast esté establecido.

Para recibir una determinada secuencia de datos multicast, los hosts deben unirse a un grupo multicast enviando un mensaje del protocolo de administración de grupo de Internet (IGMP) a su router multicast local. Casi todas las redes y las aplicaciones utilizan, ya sea, la versión 2 o 3. IGMPv2/3 permite a las personas adherirse a los receptores o salir de un grupo.

La dirección del grupo multicast se ubica en la clase D del espacio de direcciones IP. El proveedor de contenidos, el propietario y los proveedores de los servicios seleccionan la dirección multicast en base a la política local de direcciones multicast (dependiendo si las aplicaciones multicast son locales o de alcance global).

Para asegurar la confiabilidad y alta disponibilidad para las aplicaciones multicast críticas, los administradores de red, deberían:

- Eliminar cualquier simple punto de falla,
- Diseñar redes que puedan dinámicamente responder a los problemas,
- Construir redes multicast escalables,
- Emplear técnicas de alta disponibilidad,
- Brindar seguridad a nivel del servicio,
- Soportar control de acceso y de admisión,
- Aplicar políticas propias de las redes multicast.

## Desarrollo y Trabajos de Experimentación

Los trabajos experimentales se realizarán sobre una red piloto de Laboratorio. La topología tendrá una composición mixta de redes cableadas e inalámbricas, routers, switches, access points, y equipos intermedios para dar soporte a los distintos tráfico. Además, equipos finales tanto cableados como inalámbricos que permitan simular una red empresarial de cierta envergadura.

El núcleo de la red tendrá como funcionalidad principal el ruteo de todos los tráfico, sobre el cual se realizarán las configuraciones de los distintos caminos del tráfico multicast, mediante el protocolo PIM en sus diferentes versiones, mientras que la conectividad primaria IP se podrá realizar sobre cualquier protocolo de ruteo unicast. Se consideran distintas opciones de configuración:

- Todo direccionamiento bajo IPv4,
- Todo direccionamiento bajo IPv6, y
- Direccionamiento IPv4 en el núcleo de la red e IPv6 en los bordes y viceversa.

En los bordes de esta red estarán los posibles usuarios que se podrán unir a los grupos multicast generados. Para los mismo, se trabajará con protocolo IGMP (para IPv4) y con MLD (para IPv6).

La generación del tráfico multicast se realizará de dos maneras diferentes:

- Con un generador sintético como es el software IP Traffic, que permite gestionar distintas sesiones de multicast, y
- Con un emisor real, como una cámara de video IP que realice streaming multicast.

Una vez realizadas las configuraciones básicas se generará tráfico multicast y se realizaran mediciones de rendimiento, de cantidad de paquetes, de retardo y variación de retardo (jitter), de errores, de consumo de ancho de banda, etc., para obtener los resultados generales y particulares, y las conclusiones de prestaciones buscadas.

## Resultados y Objetivos

En este trabajo se pretende hacer un análisis del tipo de tráfico multicast, en un escenario

de laboratorio con equipos reales sobre redes cableadas e inalámbricas, con los siguientes objetivos iniciales:

- Analizar el direccionamiento multicast tanto en el protocolo IPv4 como IPv6,
- Analizar el protocolo de ruteo multicast PIM en sus diferentes opciones tanto para IPv4 como IPv6,
- Analizar y comparar los protocolos de interacción de los miembros de los grupos multicast IGMP (para IPv4) y MLD (para IPv6),
- Comparar el rendimiento del tráfico multicast frente a unicast en la medida que la cantidad de miembros multicast aumenta, variando las condiciones de la red (redes cableadas, redes inalámbricas, IPv4, IPv6, y las mezclas de los mismos).

Obtenidos los datos primarios mencionados anteriormente, se buscarán los siguientes objetivos finales:

- Obtener conclusiones sobre la conveniencia de tráfico multicast sobre unicast, especialmente en aquellos tipos de tráfico de tiempo real, y
- Obtener conclusiones sobre el rendimiento de tráfico multicast en cuanto a la utilización de protocolos IPv6 vs IPv4.

## Líneas de investigación y desarrollo

El proyecto estará direccionado hacia los siguientes temas:

- Direccionamiento IPv4 e IPv6,
- Tráfico multicast,
- Direccionamiento multicast IPv6,
- Protocolos de ruteo multicast PIM,
- Protocolos IGMP y MLD de multicast, e
- Interacción del tráfico multicast en redes mixtas.

## Formación de Recursos Humanos

El equipo de trabajo está integrado por docentes investigadores, y becarios graduados y alumnos del Grupo GRID TICs (Grupo UTN de Investigación y Desarrollo en TICs) de la Universidad Tecnológica Nacional, Facultad Regional Mendoza. Entre los



integrantes, hay en curso una tesis de magister y una de grado, relacionadas con la línea general de investigación.

## Plan de Trabajo

El esquema tentativo del plan de trabajo considera las siguientes actividades:

- a. Buscar y obtener la información primaria sobre los protocolos a estudiar,
- b. Estudiar los distintos protocolos,
- c. Desplegar la red física propuesta,
- d. Configurar dicha red,
- e. Verificar la generación de tráfico multicast y evaluar los primeros resultados,
- f. Verificar la generación del tráfico multicast y analizar los resultados, y
- g. Efectuar las conclusiones.

Se aclara que los dos primeros puntos están muy avanzados, mientras que para los puntos “c” y “d” ya se han realizado pruebas de laboratorio en una red a escala.

## Recursos y equipamiento

Las actividades se llevarán a cabo en el ámbito de las instalaciones del grupo de investigación GRID TICS, de la UTN Regional Mendoza. Dicho Grupo cuenta con sus propias áreas de trabajo, 1 oficina técnico-administrativa, 2 Laboratorios con 11 computadoras cada uno, con material y con el siguiente equipamiento:

- 4 (cuatro) Routers CISCO 2811,
- 6 (seis) Routers CISCO 1721,
- 3 (tres) Switchs CISCO 2950,
- 2 (dos) Switchs CISCO 2960,
- 2 (dos) Switchs CISCO 3560,
- 1 (uno) ASA CISCO 5505,
- 2 routers Mikrotik,
- 4 Access Point Cisco y 2 Mikrotik,
- Placas inalámbricas de red,
- 2 (dos) cámaras de video IP con soporte de string multicast IPv4/IPv6,
- 22 (veintidos) CPs con Sistema Operativo Linux, Windows XP y Windows 7,
- Software IP Traffic de ZTI – Generador de tráfico IPv4/IPv6 unicast/multicast/broadcast y Medidor de performance (throughput,

cantidad de paquetes, jitter, número de errores, tanto enviados como recibidos, etc.),

- Software Analizador de tráfico Wireshark,
- Hardware Air Pcap para captura de tráfico wireless,
- Conexión a Internet por IPv4 e IPv6, y
- Servidor HP Proliant con Linux base y Máquinas Virtuales.

## Referencias

- [1] RFCs de direccionamiento multicast
- [2] Data and Computer Communications – William Stalling - 9º edición
- [3] [www.isoc.org/inet2000/cdproceedings/1c/1c\\_3.htm](http://www.isoc.org/inet2000/cdproceedings/1c/1c_3.htm)
- [4] [www.isoc.org/inet2000/cdproceedings/1c/1c\\_2.htm](http://www.isoc.org/inet2000/cdproceedings/1c/1c_2.htm)
- [5] [www.isoc.org/inet2000/cdproceedings/1c/1c\\_1.htm](http://www.isoc.org/inet2000/cdproceedings/1c/1c_1.htm)
- [6] [www.isoc.org/pubs/int/cisco-1-6.html](http://www.isoc.org/pubs/int/cisco-1-6.html)
- [7] [www.zte.com.cn/endata/magazine/ztecommunications/2006year/no3/articles/200610/20061010\\_162407.html](http://www.zte.com.cn/endata/magazine/ztecommunications/2006year/no3/articles/200610/20061010_162407.html)
- [8] [www.juniper.net/techpubs/en\\_US/junos12.2/topics/concept/ipv6-multicast-flow-overview.html](http://www.juniper.net/techpubs/en_US/junos12.2/topics/concept/ipv6-multicast-flow-overview.html)
- [9] [www.h3c.com/portal/Products\\_\\_\\_Solutions/Technology/IP\\_Multicast/](http://www.h3c.com/portal/Products___Solutions/Technology/IP_Multicast/)
- [10] [www.elisanet.fi/teemuki/translator/thesis.pdf](http://www.elisanet.fi/teemuki/translator/thesis.pdf)
- [11] <http://web.dit.upm.es/~jmseyas/linux/mcast.como/Multicast-Como.html>
- [12] <http://www.wireshark.org>

# Internet del Futuro y Ciudades Inteligentes

Eduardo O. Sosa<sup>(1,a)</sup>; Diego A. Godoy<sup>(2)</sup>; Raúl Neis<sup>(2)</sup>; Gabriel Motta<sup>(2)</sup>; Raúl Luft<sup>(2)</sup>; Darío Sosa<sup>(2)</sup>; Hernán Bareiro<sup>(2)</sup>; Paola Quiñones<sup>(2)</sup>

<sup>(1)</sup> Secretaría de Investigación y Posgrado (SECIP). Facultad de Ciencias Exactas, Químicas y Naturales, UNaM

<sup>(2)</sup> Centro de Investigación en Tecnologías de la Información y Comunicaciones (CITIC). Universidad Gastón Dachary

es@fceqyn.unam.edu.ar; diegodoy@citic.ugd.edu.ar, {raulneis, mottitag, raulluft, suedrosa, hernanbareiro, paola4280}@gmail.com

## Resumen

En este documento se presentan los avances de una serie de trabajos en proceso en el marco de proyectos de investigación desarrollados tanto en la Universidad Nacional de Misiones como en la Universidad Gastón Dachary relacionados con la Internet del Futuro y Ambientes Inteligentes. Los trabajos presentados son: Simulación de Redes de Sensores Remotos mediante interfaz basada en la Web, Contribución a la Gestión de Residuos Domiciliarios como una Aplicación en Ciudades Inteligentes y una Plataforma para la publicación de datos de Redes de Sensores Inalámbricos, orientada a la visión de la Internet de las Cosas.

**Palabras clave:** Internet del Futuro, Ciudades Inteligentes, Redes de Sensores Inalámbricos.

## Contexto

El trabajo presentado aquí se realiza en el marco de diversos proyectos de investigación en la Secretaría de Investigación y posgrado de la Facultad de Ciencias Exactas, Químicas y Naturales de la Universidad Nacional de Misiones sobre el tema “Internet del Futuro” e “Internet de las Cosas”, a saber: Proyecto 16Q457 “Hacia la Programación de Sensores Inalámbricos en la Forma web 2.0”, proyecto 16Q474 “Simulaciones de Sistemas Modernos de Comunicación” y el proyecto (en evaluación) “Ambientes Inteligentes. Una Mirada a Internet del Futuro”. Asimismo del proyecto de investigación denominado “Diseño de arquitecturas de soporte a la Internet del Futuro y Ambientes Inteligentes para Ciudades Inteligentes” que se ha formalizado por la resolución (19/A/12) de la Universidad Gastón



Dachary mediante el llamado al 5to Concurso de Proyectos de Investigación en 2012.

## Introducción

El continuo avance de la ciencia y la tecnología a lo largo del tiempo ha conllevado a un sinnúmero de innovaciones. Basta observar los avances que se han experimentado en la astronomía, basados en la mejora y evolución de las tecnologías de desarrollo de telescopios; como también los progresos a los largo de las últimas décadas en las ciencias médicas con todo lo referente al mundo microscópico. Cada una de estas innovaciones ha tenido que ver, indudablemente, con el desarrollo explosivo de las Tecnologías de la Información y Comunicaciones. Hoy es posible realizar medidas y cálculos a partir de las imágenes obtenidas por unidades satelitales, utilizando avanzados algoritmos computacionales para el cálculo.

Mark Weiser, conocido como el padre de la computación ubicua, a principios de los años 1990 ha apostado a la mencionada teoría, la que debería evolucionar hasta interactuar con los diferentes usuarios de las tecnologías[1],[2]. Estos principios enunciados originalmente por Weiser han sido validados por diferentes investigadores y estudiosos, convirtiéndose en un escenario cierto en el futuro cercano.

Los entornos inteligentes -hogares, oficinas, escuelas, y por supuesto ciudades; etc.- representan entornos avanzados TIC's en las cuales interaccionan objetos de uso cotidiano en continua evolución para usuarios no experimentados. Los entornos inteligentes han surgido rápidamente como un paradigma nuevo y emocionante que tiende a incluir diferentes

campos de investigación de la computación ubicua, o pervasiva, y en red.

La domótica, o smart-home, comprende varios enfoques, pero siempre tiende a una mejor calidad de vida de los habitantes. Los diversos enfoques van desde la mejora en la vida diaria, a permitir una vida más independiente para las personas mayores y los discapacitados. La tarea de los objetos inteligentes, implantados en artículos de uso diario, es detectar el entorno inmediato usando varios tipos de sensores, para luego procesar esta información. Esta función asigna un tipo de inteligencia artificial a los objetos comunes y bien conocidos y permite el procesamiento integral de información y la interconexión de casi cualquier tipo de objeto cotidiano[3].

La computación ubicua encuadra a un gran número de tecnologías y aplicaciones; desde dispositivos móviles utilizados diariamente, pasando por artefactos "inteligentes" para propósitos especiales (hornos, heladeras, etc.) y hasta a los juegos domiciliarios en red. Tiene que ver, y está íntimamente relacionada, con la evolución de los dispositivos electrónicos. Esta revolución se materializa a través de la miniaturización de los dispositivos y la mayor sociabilización de los contenidos disponibles. De acuerdo a la ley de Moore, el número de transistores, y por ende la capacidad de almacenamiento y cálculo de un circuito; se duplica cada uno o dos años[4]. Por ello, en la actualidad, es un desafío establecer y orientar los estudios hacia las regiones del conocimiento donde las TICs, convergen apuntalando a las ciencias.

La existencia de las numerosas y variadas simulaciones en los diversos campos de las ciencias son consecuencia del crecimiento exponencial de las propiedades y capacidades de los semiconductores que fueran

desarrollados por la industria en los últimos años. Así, hoy es posible realizar simulaciones numéricas para estudiar y analizar fenómenos utilizando métodos computacionales, los que se volverían prohibitivos si se pretendiera realizarlos por métodos empíricos.

La convergencia de sistemas de computación ubicuos embebidos en objetos de todos los días, los cuales se comunican inalámbricamente, con interfaces de nueva generación, sensores de diversos tipos, agentes inteligentes, sistemas de personalización, etc., tienden a conformar una inteligencia ambiental que nos rodea potenciando nuestras capacidades cognitivas. En tales ambientes el objetivo es proporcionar servicios informáticos y de comunicación de una manera mucho más simple, cómoda, agradable y transparente, donde los usuarios puede acceder de forma remota y controlar la información y aplicaciones en su entorno utilizando los diversos servicios derivados de la cooperación integrada de objetos heterogéneos intercomunicados entre sí.

Los aparatos que compondrán este ambiente aprenderán de las necesidades de las personas y luego las anticiparán. Crearán un ambiente inteligente, a nuestra disposición. La Inteligencia Ambiental está centrada en la persona; basados en ordenadores emocionales, nano-tecnología, vida artificial, sistemas virtuales, ubicuidad, biometría, sistemas de personalización, agentes inteligentes, Papel Electrónico, Bioclimas activos, cientos de microcomputadoras embebidas en la ropa, en los muebles, sensores inalámbricos, etc.

La posibilidad de acceder a una gran cantidad de información y proceder al procesamiento de la misma implica la existencia de una gran cantidad de elementos distribuidos en un determinado entorno físico que constituyen redes entre ellos. Este proyecto

pretende proponer algunas herramientas que serán aplicables a los entornos inteligentes del futuro.

De la misma manera, la globalización tecnológica ha transformado el orden del conocimiento y la innovación a escala global en un período de tiempo increíblemente corto, provocando con ello nuevos desafíos y aparición de oportunidades. La única manera que las ciudades, como elemento organizado, puedan competir en este nuevo entorno, e siendo cada vez más innovadoras y capaces de responder de manera efectiva a las preferencias y necesidades de sus ciudadanos. El concepto de Ciudades Inteligentes está muy extendido en el contexto europeo y en algunos países de Asia, donde estas tecnologías se despliegan en el mundo real no solo con el fin de analizar su viabilidad, sino también para considerar su impacto socio-económico. Las ciudades conforman así, hace un buen tiempo, un tecnopaisaje urbano en nuestra sociedad. Con ello están dando paso a un nuevo modelo productivo basado en el conocimiento, y en la utilización del mismo como aporte de soluciones técnicas que se ocupan de una mejor calidad de vida y de la sustentabilidad. Sin embargo se observa escasa actividad incipiente en las ciudades latinoamericanas.

Cuando el objetivo de una acción es obtener información válida de algún entorno con el cual se interactúa, la más de las veces es prácticamente imposible establecer vínculos fijos permanentes (cableados normales). Para estos escenarios se ha establecido una nueva tecnología, basada en un nuevo paradigma para los sistemas de cómputo, que han venido a denominarse redes de sensores inalámbricos (WSN).

Merece ser destacado aquí, por similitud operativa, la tecnología de identificación por

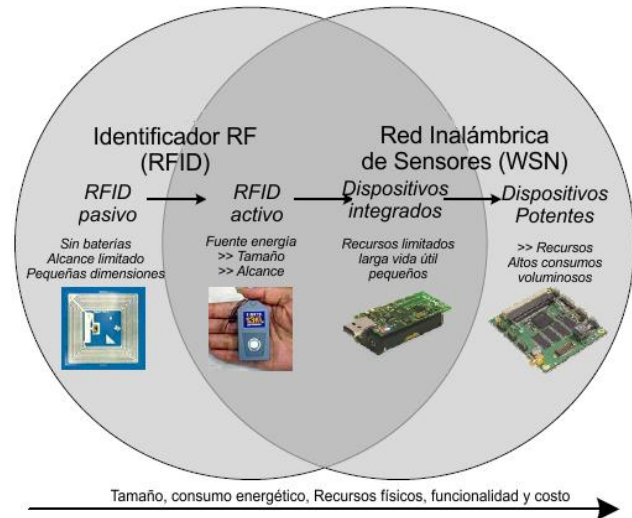
radio frecuencia (RFID), que consiste en un sistema de almacenamiento y recuperación de datos remotos, utilizados por transponders<sup>1</sup> RFID. Con éste sistema, se transmite radialmente la identidad determinada de un objeto por medio de un código y/o número de serie. Los elementos necesarios para ello son las etiquetas y los lectores. Los últimos, son los que detectan e identifican a determinado objeto físico, y que el mismo pueda ser controlado e inventariado en un sistema de cómputos al efecto.

Las etiquetas RFID se diferencian entre las pasivas y las activas. Las primeras carecen de una fuente de energía, pero son capaces de ser energizadas desde el dispositivo que realiza la consulta. Los RFID activos son dispositivos provistos de alguna fuente de energía, guardando una relación con pequeños nodos integrados de WSN[5]. Recientemente han sido desarrollado dispositivos RFID clase 4 y 5, los que son capaces de establecer entornos de seguimiento, ubicación, red y seguridad.

El deseo de mejorar el rendimiento de los servicios de las cadenas de provisión globalizadas ha impulsado el desarrollo de los sistemas RFID, pero su implementación ha sido problemática[5], en primer lugar por los costos de manufactura de las etiquetas y, fundamentalmente, por invasión de la privacidad de los usuarios adquirentes de productos etiquetados con RFID.

De estudios realizados en diferentes trabajos y publicaciones, se infiere el éxito de la computación ubicua; basado fundamentalmente en la adopción de las implementaciones de WSN y RFID. La ilustración 1 muestra la

superposición existente entre las tecnologías RFID y las WSN.



**Ilustración 1. Intersección de tecnologías RFID y WSN**

Una red de sensores inalámbricos consiste en una serie de dispositivos distribuidos desordenadamente en un área geográfica dada. Cada nodo posee capacidad de comunicación inalámbrica, de procesamiento de la información obtenida y fundamentalmente, cada uno de los nodos es capaz de promover el establecimiento de una red de datos eficiente.

Las WSN han trascendido el ámbito académico-científico, habiéndose vuelto una realidad en cuanto a la utilidad práctica de la tecnología. El alcance y diversidad de las aplicaciones de las redes de sensores, los requerimientos, diseños y plataformas son virtualmente ilimitados [6]

## Líneas de investigación y desarrollo

Se avanza en diversas líneas, caracterizadas por las iniciativas de los estudiantes a graduarse, con la orientación de docentes e investigadores de la Universidad de Misiones y

<sup>1</sup> Dispositivos electrónicos para comunicaciones inalámbricas; fusión de transmisor-responder

las Universidad Gastón Dachary. Detallando las líneas de investigación y desarrollo se deben mencionar:

1) *Simulación de Redes de Sensores Remotos mediante interfaz basada en la Web*; cuyo objetivo principal es desarrollar un prototipo de Interfaz de simulación de WSN basado en la Web. La simulación por computadora ha permitido a los científicos e ingenieros experimentar fácilmente con ambientes virtuales, alcanzando un nuevo nivel de detalle el análisis de las aplicaciones naturales y artificiales; que fuera desconocido en las primeras etapas del desarrollo científico. Además proporciona una gran ayuda en el diseño y análisis de aplicaciones complejas.

Es conocida la dificultad de modelar analíticamente a las WSN, dado que se tiende a realizar análisis simplificados. Toda simulación requiere de un modelo apropiado basado en fundamentos teóricos y sobre todo, de fácil implementación práctica [7], dado que los resultados de la simulación se extrapolan del escenario particular de análisis, con determinadas presunciones, que ciertas veces no encierran al comportamiento real de las WSN, comprometiendo seriamente con ello la credibilidad de las simulaciones.

En el trabajo se fijaron los siguientes objetivos específicos: a) Estudiar el estado del arte en la simulación de redes de sensores remotos basado en la web, b) Analizar las propiedades y eventos necesarias para reproducir el comportamiento de una Red de Sensores Remotos c) Diseñar la interfaz Web de obtención de parámetros, incorporación de los archivos particulares del proyecto y visualización de los resultados de simulación, d) Diseñar una solución del lado del servidor Web para procesamiento de los datos colectados y generación de los resultados de

simulación e) Desarrollar un prototipo de interfaz de simulación de WSN basado en la Web que a través de una interfaz pueda obtener parámetros de simulación, incorporar archivos particulares del proyecto, procesar los datos ingresados y generar los resultados de la simulación, f) Realizar pruebas para comprobar el funcionamiento correcto del sistema. La justificación para la realización de éste trabajo de investigación, es ofrecer una sistema que permita fácilmente tener un entorno de simulación disponible para múltiples usuarios. Esto permitirá abstraer a los usuarios de la complejidad del funcionamiento del sistema de simulación en sí, haciendo más amigable su uso. Integrando todo potencial de la herramienta de simulación Shawn[8] con todas las ventajas de los sistemas basados en la Web se permitirá simplificar el uso del simulador y disminuir los tiempos asociados a la instalación. Además dado que el procesamiento de la simulación se realizará en un servidor, esto permitirá que se creen luego nuevas formas de procesar los datos. Se podrá por ejemplo incrementar la capacidad del servidor realizando el procesamiento en un cluster de computadoras.

2) *Contribución a la Gestión de Residuos Domiciliarios como una Aplicación en Ciudades Inteligentes*, cuyo objetivo principal es diseñar un prototipo basado en redes de sensores inalámbricas para la detección de contenedores de residuos domiciliarios llenos, como una aplicación en ciudades inteligentes. En muchos campos, las diferentes redes de sensores inalámbricos[9] han desarrollado aplicaciones para resolver problemas de gestión con implementaciones inteligentes. Uno de los casos de éxito tiene que ver con la gestión de los residuos sólidos. En este trabajo se propone con el objetivo de



mejorar el manejo en el lugar y la optimización de transferencia en el proceso de gestión de residuos.

Como objetivos específicos se han fijado: a) analizar bibliografía sobre Ciudades Inteligentes, Internet del Futuro y Redes de Sensores Inalámbricas, b) Definir componentes de una red de sensores inalámbricas, detallando las características y funcionalidad de cada uno de sus elementos, c) Desarrollar un prototipo de software para dotar de funcionalidad a los sensores, d) Diseñar una interfaz para visualización de ruta óptima, e) Realizar pruebas en laboratorio, con hardware real, con el fin de validar la arquitectura propuesta.

Las áreas urbanas crecen con gran velocidad. Más de la mitad de la población mundial vive ya en ciudades. La previsión que maneja Naciones Unidas señala que el 70% de los seres humanos habitarán en centros urbanos en 2050. Este organismo advierte de que el aumento de la población de las ciudades puede convertirse en un auténtico problema, a no ser que se logre mantener la armonía entre los aspectos espacial, social y ambiental de las localidades, así como entre sus habitantes. En este nuevo escenario sociológico y demográfico, con claros efectos económicos, políticos y medioambientales, cobra fuerza el concepto de ciudad inteligente. Una ciudad inteligente es aquella que hace uso de los avances tecnológicos para mejorar la calidad de vida de sus habitantes. Son aquellas donde la inteligencia "se pone al servicio del ciudadano" a fin de gestionar de la mejor forma los recursos.

3) *Plataforma para la publicación de datos de Redes de Sensores Inalámbricos, orientada a la visión de la Internet de las Cosas*, cuyo objetivo principal es diseñar una plataforma para la captura, almacenamiento y publicación de datos de Redes de Sensores Inalámbricos,

orientada a la visión de Internet de las Cosas, persiguiendo a) Estudiar las tecnologías WSN y establecer el estado del arte de las mismas, b) Evaluar las alternativas existentes para la publicación de datos de WSN útiles para la visión de la IoT, c) Diseñar un prototipo de *plataforma* que permita la captura, almacenamiento y publicación de los datos obtenidos de la WSN, d) Probar el prototipo en un escenario donde se verifiquen las posibilidades de la plataforma.

El término Internet de las Cosas se refiere a una visión en la que la Internet se extiende conectando los objetos cotidianos a la red, permitiéndoles conectarse entre sí y cooperar unos con otros, en forma ubicua, en cualquier momento, basados en la afirmación que se hiciera Mark Weiser, "*las tecnologías más exitosas son las que desaparecen, se entretienen con la trama de la vida cotidiana hasta que no se distinguen de ella*" [1]. Cabe remarcar que la IoT es una visión de la Internet del Futuro (FI, Future Internet) y como tal, es un campo aún en desarrollo. No existe una implementación concreta que pueda llamarse IoT, sino que es un conjunto de tecnologías de internet que permitirán, en el futuro, concretar la visión de la interconexión de los objetos cotidianos a la red.

Se observa de los lineamientos expresados, la pretensión de desarrollar e implementar algunas aplicaciones prototipo para ver el impacto producido sobre ambientes válidos en el entorno de ejecución del proyecto, y sobre todo si es factible su implementación práctica a escala macro. Se pretenden abordar cuestiones relativas a la conexión entre las tecnologías y aplicaciones inteligentes para los distintos ambientes, realizando análisis y evaluación de la gestión del tráfico, soluciones para el estacionamiento diario de los ciudadanos,

manejo de residuos domiciliarios, control de la contaminación de agua y aire, la calidad del agua superficial, la eficiencia del alumbrado público, control del ruido y sonidos urbanos, transporte público, etc. Así se deberían identificar a las aplicaciones pertinentes para los distintos ambientes. Las aplicaciones y prototipos más prometedores serán seleccionados, y aplicados en el entorno real.

## Formación de Recursos Humanos

Relacionados a este proyecto existen 6 tesinas de grado en el tema de Internet de las cosas e Internet del Futuro, en trámite de evaluación en a UGD, como también trabajos de interacción con la industria del Té, de manera de mejorar la medición de parámetros en el proceso de fermentación en industrias de la Provincia de Misiones. Un trabajo de Especialización en Ingeniería del Software, UNLP, se encuentra en fase de evaluación de trabajo final. El grupo de trabajo está liderado por un doctor en ciencias informáticas, un ingeniero en sistemas y un ingeniero en Telecomunicaciones.

La Universidad Gastón Dachary, como soporte a la resolución 19/A/12, y apoyando a los proyectos de investigación en la UGD, ha asignado recursos suficientes para que varios alumnos en etapa de preparación de sus tesis de grado sean asimilados como becarios, los que son solventados con fondos propios.

## Referencias

- [1] Mark Weiser, "The computer for the 21st century," vol. 265, pp. 66-75, Septiembre 1991.
- [2] Satyanarayanan Mahadev, "Pervasive computing: Vision and challenges," *IEEE Personal Communications*, vol. 8, Agosto 2001.
- [3] S. Dengler, A. Awad, and F. Dressler, "Sensor/Actuator Networks in Smart Homes for Supporting Elderly and Handicapped People," in *21st International Conference on Advanced Information Networking and Applications Workshops*, vol. 2, 2007, pp. 863-868.
- [4] D. Estrin, and M. Srivastava D. Culler, "Overview of sensor networks," *IEEE Computer*, pp. 37(8):41-49, August 2004.
- [5] M., Kranenburg, R. Ward and G. Backhouse, "RFID: Frequency, standards, adoption and innovation," Department of Design, Goldsmiths College, University of London, London, 2006.
- [6] Th. Arampatzis, J. Lygeros, and S. Manesis, "A Survey of Applications of Wireless Sensors and Wireless Sensor Networks," in *Proceedings of the 2005 IEEE International Symposium on, Mediterrean Conference on Control and Automation.*, 2005, pp. 719-724.
- [7] E. Egea-López and et al., "Simulation Tools for Wireless Sensor Networks," in *Summer Simulation Multiconference - SPECTS 2005*, 2005, pp. 1-9.
- [8] S.P. Fekete, A. Kroller, S. Fischer, and D. Pfisterer, "Shawn: The fast, highly customizable sensor network simulator," , Braunschweig, 2007, pp. 299 - 299.
- [9] S. Longhi et al., "Solid Waste Management Architecture Using Wireless Sensor Network Technology," in *5th International Conference on New Technologies, Mobility and Security (NTMS)*, 2012, pp. 1-5.



# Conectividad WSN: Implementación de un Middleware WSN-IP-WWW

Benítez, Juan de Dios<sup>(1a)</sup>, Gloza, Gonzalo Martín<sup>(2b)</sup>, Eduardo Omar Sosa<sup>(3c)</sup>,  
Diego Godoy<sup>(4d)</sup>

<sup>(1,4)</sup> Centro de Investigación en Tecnologías de la Información y  
Comunicaciones (CITIC). Universidad Gastón Dachary

<sup>(3)</sup> Secretaría de Investigación y Posgrado (SECIP). Facultad de Ciencias  
Exactas, Químicas y Naturales, UNaM

(a,d) juan.benitez, diegodoy@citic.ugd.edu.ar;(b) ggloza@gmail.com;(c)es@fceqyn.unam.edu.ar

## Resumen

En el paradigma de las WSN encontramos que una de las principales características es la reducción del tamaño de los dispositivos y la disminución de los costos, lo que da lugar a que cada nodo deba tener únicamente el hardware estrictamente necesario para realizar de manera eficiente la tarea que se le ha encomendado. Esto conlleva a situaciones en que la conectividad de los nodos sensores con redes existentes sea limitada o nula, requiriendo así entonces un elemento capaz de realizar las interacciones necesarias de interconexión, para así lograr una homogenización de los datos y hacerlos accesibles por ejemplo a través de internet. En Este trabajo se presenta un middleware WSN-IP-WWW para la integración de redes de sensores inalámbricos con la Web.

**Palabras clave:** Nodos Sensores, Internet, WSN, middleware.

## Contexto

El trabajo presentado aquí se realiza en el marco del proyecto de investigación denominado “Diseño de arquitecturas de soporte a la Internet del Futuro y Ambientes Inteligentes” que se ha formalizado por la resolución (19/A/12) de la U.G.D. mediante el llamado al 5to Concurso de Proyectos de Investigación en 2012, en dicho proyecto se encuentran trabajando 3 tesinas de grado un becario de investigación graduado y un ingeniero colaborador ad honorem de la ciudad de Córdoba. Asimismo se relaciona directamente con diversos proyectos de investigación en la Secretaría de Investigación y posgrado de la Facultad de Ciencias Exactas,

Químicas y Naturales de la Universidad Nacional de Misiones sobre el tema “Internet del Futuro” e “Internet de las Cosas”, a saber: Proyecto 16Q457 “Hacia la Programación de Sensores Inalámbricos en la Forma web 2.0”, proyecto 16Q474 “Simulaciones de Sistemas Modernos de Comunicación” y proyecto (en evaluación) “Ambientes Inteligentes. Una Mirada a Internet del Futuro. Referente a este proyecto existen 2 tesinas de grado y 1 de posgrado en curso entre los participantes.

## Introducción

El desarrollo alcanzado por tecnologías de fabricación de circuitos integrados, tales como microcontroladores y dispositivos de lógica programable FPGA, han provocado el auge de las aplicaciones de sistemas embebidos. Las mismas se han convertido en la solución a gran número de problemas que anteriormente requerían de un sistema mucho más complejo, por ejemplo una PC.

Una aplicación de este tipo de sistemas es la interacción entre las WSN y las redes existentes, ya que la tendencia de los nodos de una WSN es la miniaturización de los nodos, lo que conlleva a la necesidad de limitar el hardware que puede llegar a contener cada nodo sensor, tanto por el tamaño como así también por la necesidad de bajos consumos de energía[1]. Así cada nodo contará únicamente con el hardware estrictamente necesario para poder realizar de manera eficiente la tarea que se le ha encomendado. Es así que la posibilidad de conectividad de cada nodo de la WSN juega un papel importante en esta miniaturización y en el consecuente uso de energía. De acuerdo a ello, podemos dividir la conectividad en a) entre nodos y b) con otras redes ya existentes como se puede ver en la Ilustración 1.

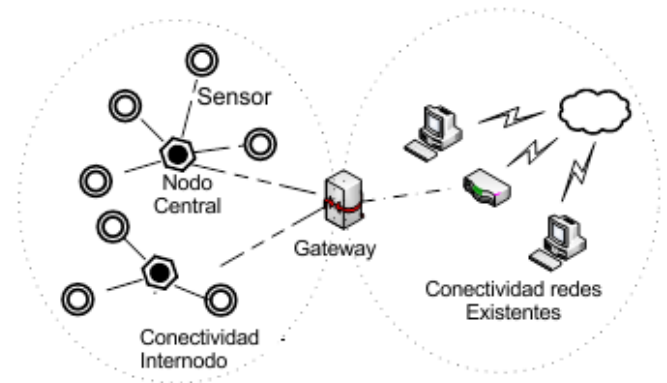


Ilustración 1. Conectividad en redes WSN

La necesidad de conectividad entre los nodos permite transferir información entre nodos vecinos, y además nos permite realizar un encaminamiento multi-salto de los paquetes de información, reduciendo así, la potencia de transmisión de cada nodo y con ello el consumo energético.

Sin embargo estos datos brindados por la WSN son poco útiles si no podemos tener un acceso de forma sencilla a ellos, para lo que es necesaria la conectividad de las WSN hacia las redes existentes, tanto locales como Metropolitanas. Considerando las capacidades limitadas de hardware de los nodos, no es común encontrar nodos sensores, capaces de realizar los dos tipos de conectividad, internodo y extra-nodo. Dentro de ellos encontramos a algunos con capacidad de manejo de paquetes IP[2].

6LoWPAN[3] brinda un estándar a las WSN con fragmentación y compresión de las cabeceras de los paquetes IPv6 para ser transmitidos por 802.15.4. Esto no les permite la conexión directa a las redes, lo que hace necesaria la utilización de una puerta de enlace capaz de direccionar los paquetes IPv6 de la

WSN hacia el exterior. Si bien Existen variadas aplicaciones que brindan la posibilidad de publicar y manejar los datos de una WSN, requieren en todos los casos una infraestructura para el envío de datos [4].

La arquitectura propuesta se puede ver en la Ilustración 2.

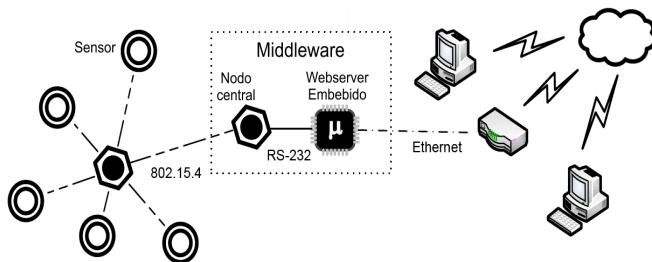


Ilustración 2. Arquitectura Middleware Propuesta

El middleware propiamente dicho está compuesto por un nodo sensor de la WSN, conformado por un nodo iSense [5] (Ilustración 3), con procesador Jennic JN5139, un sistema radial de 2,4 GHz compatible con normas IEEE 802.15.4, con ancho de banda estimado de 250kbit/s, sistema de encriptación por hardware AES, capacidad de cálculo de tiempo de vuelo (ToF) para cálculo de distancias entre nodos, 192kB de ROM, 96kB of RAM, como una variada posibilidad para la utilización de periféricos analógicos y digitales.

Por medio de un vínculo RS-232 se encuentra interconectada a una plataforma compuesta por un microprocesador ARM Cortex M4[6], donde el acceso desde/hacia redes TCP/IP es realizado mediante un servidor web embebido en dicha plataforma. La conectividad TCP/IP se logró implementando la librería uIP, la cual es una aplicación *open source* en lenguaje C para micro controladores de la pila de protocolos TCP/IP, cuya principal

particularidad es la minimización de la memoria utilizada. Ésta soporta los protocolos TCP, UDP, IP, ICMP y ARP.



Ilustración 3. Módulo principal iSense

La librería uIP provee una interfaz de programación de aplicación (API) que define como las aplicaciones interactúan con la pila TCP/IP. Dicha API se denomina “protosockets”, la cual es muy similar a la API de sockets Berkeley[7] pero sin la sobrecarga de los entornos multitareas. Ciertas modificaciones se realizaron para portar la stack uIP de manera adecuada a la arquitectura ARM, para evitar incompatibilidades en cuanto a los tipos de datos y temporizadores requeridos en la implementación.

Mientras que la librería uIP proporciona conectividad TCP/IP, las funcionalidades de la capa de enlace de datos se implementó mediante el controlador Ethernet ENC28J60 de Microchip, añadido como periférico al procesador Cortex M4 conectados mediante un bus SPI. Por otra parte la integración con la WSN se lleva a cabo a través de un nodo sensor de la red utilizado como concentrador de datos, donde allí por medio de un enlace 802.15.4 intercambia información desde y hacia la WSN, pudiendo determinar por medio de simple

direccionamiento MAC desde que sección de la red provienen los datos o hacia que sección enviar los mismos. La comunicación con el micro controlador se realiza por medio del protocolo RS-232, proveyendo un enlace bidireccional a ambos dispositivos para el intercambio de datos.

## Líneas de investigación y desarrollo

Para la implementación del servidor web se utilizó una placa de desarrollos Stellaris Launchpad de Texas Instrument® [8].

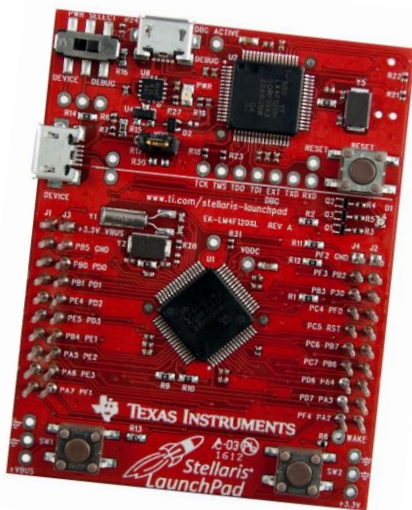


Ilustración 4. TI Stellaris Launchpad

La placa está conformada por un micro controlador ARM Cortex M4 - LM4F120H5QR- con punto flotante, capacidad de trabajo de hasta 80MHz, 256Kb de memoria Flash, 32Kb de SDRAM, 8 transceptores asíncronos universales (UART), 4 buses de intercomunicación en serie I<sup>2</sup>C, 4 interfaces seriales periféricas, 27 Timers, interfaz USB para ICDI (In Circuit Debug Interface), Micro B USB para debug, Micro B USB para

conexión con aplicaciones, 2 botones para aplicaciones de usuario, botón de reset y RGB LED (Ilustración 4).

La implementación consistió de tres etapas. La primera de ellas fue la implementación de un nodo WSN sumidero, el cuales el responsable de capturar todas las comunicaciones de la red WSN. La segunda parte del trabajo consistió en la implementación del servidor web sobre la plataforma Stellaris Launchpad, capaz de generar contenido dinámico a partir de la incorporación de llamadas a funciones en lenguaje C desde el código de las páginas web.

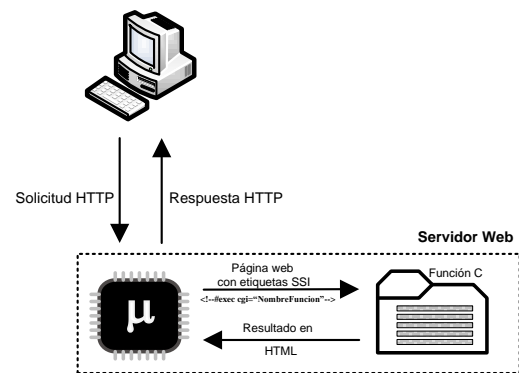


Ilustración 5. Servidor http

La inserción dinámica de contenido en una página solicitada del servidor web (ilustración 5) es posible mediante el conjunto de directivas SSI (Server Sides Includes), siendo capaz de procesar solicitudes de códigos html, css y shtml, siendo esta ultima la extensión asociada a páginas con contenido dinámico, como imágenes en formato jpg, gif y png.

La tercera etapa del trabajo consistió en lograrla interacción de los dos sistemas antes explicados. La comunicación entre las plataformas se llevó a cabo a través de una



conexión por puerto serie con el protocolo RS-232.

El nodo sensor sumidero recolecta la información proveniente de la WSN y por medio del direccionamiento MAC podrá determinar desde que sección de la WSN fue recibido el paquete de datos, una vez recibido el paquete, se extraerá la información y se lo identificará dentro de un paquete nuevo para ser enviado hacia el servidor web por medio del enlace RS-232.

## Resultados y Objetivos

En este trabajo se presenta un modelo de interoperabilidad middleware para la Integración de Redes IP y la Web con Redes de Sensores Inalámbricos con el fin de generar agentes con capacidades para percibir el mundo, tomar decisiones y actuar sobre el entorno.

## Formación de Recursos Humanos

El equipo de trabajo se encuentra formado por un Doctor en Ciencias Informáticas, Un Doctorado en Ingeniería Telemática, un auxiliar de investigación graduado, un auxiliar de investigación graduado residente en la ciudad de Córdoba Capital, y ocho auxiliares (Resolución rectoral 21/I/12) de investigación en período de realización de trabajos finales de grado. El número de tesinas de grado en curso con proyecto aprobado es tres y el número de trabajo de especialidad en curso con proyecto aprobado es uno. Los proyectos de grado se titulan “Plataforma para la publicación de datos de Redes de Sensores Inalámbricos, orientada a la visión de la Internet de las Cosas, Ambientes Inteligentes y Mashups”, “Diseño de un prototipo para monitoreo eficiente de iluminación basado en WSN utilizando

HTML5” y “Contribución a la Gestión de Residuos Domiciliarios como una Aplicación en Ciudades Inteligentes”. El Trabajo de Especialidad se titula “Plataformas para la creación de Mashups Sensibles al Contexto en Entornos de Inteligencia Ambiental”

## Trabajos citados

- [1] Eduardo Sosa, "Tesis Doctoral: "Contribuciones al Establecimiento de una Red Global de Sensores Inalambricos Interconectados", La Plata, 2011.
- [2] T. Teubler, M. Hail, and H. Hellbruck, "Transparent Integration of Non-IP WSN into IP Based Networks," in *IEEE 8th International Conference on Distributed Computing in Sensor Systems*, 2012, pp. 353-358.
- [3] Internet Engineering Task Force (IETF). (2013) IPv6 over Low power WPAN (6lowpan). [Online]. <http://goo.gl/eMyk1>
- [4] E. O. SOSA, S. FISCHER, and F. J. DIAZ, "Twitter, Soporte de una Red de Sensores Inalámbricos," in *XII REUNIÓN DE OTOÑO DE POTENCIA, ELECTRÓNICA Y COMPUTACIÓN, ROPEC, IEEE*, Colima, Mexico, 2010.
- [5] Coalesenses GmbH. (2013) Bridging the gap between virtuality and reality. [Online]. <http://www.coalesenses.com/>
- [6] ARM. (2013) The Architecture of the Digital World: Cortex-M4 Processor. [Online]. <http://tiny.cc/h9nftw>
- [7] University of Princeton. Berkeley sockets. [Online]. <http://goo.gl/Z4GwL>



- [8] Texas Instruments. Stellaris LM4F120 LaunchPad Evaluation Kit. [Online]. <http://www.ti.com/tool/ek-lm4f120xl>
- [9] coalesenses GmbH. Bridging the Gap Between Virtuality and Reality. [Online]. <http://www.coalesenses.com>

# Protocolos para Mejorar la Performance de las Aplicaciones Web

Marcelo F. Fernández<sup>1</sup>, Fernando Tinetti<sup>2</sup>, Gabrel H. Tolosa<sup>1</sup>

mail@marcelofernandez.info, fernando@info.unlp.edu.ar, tolosoft@unlu.edu.ar

<sup>1</sup> Departamento de Ciencias Básicas, Universidad Nacional de Luján.

<sup>2</sup> Facultad de Informática, Universidad Nacional de La Plata.

## Resumen

El protocolo HTTP ha sido revisado en un par de ocasiones, la última, ya hace más de una década. Sin embargo, las características de la Web actual, los requerimientos de sus usuarios, y los niveles de masividad hicieron que los recursos que se proveen mediante éste hayan alcanzado un punto donde se han encontrado algunas limitaciones inherentes al diseño original del protocolo. Es por ello que el IETF (*Internet Engineering Task Force*), dentro del *HTTPbis Working Group*, está analizando modificaciones, ajustes y/o mejoras de fondo en pos de alcanzar un futuro estándar HTTP 2.0.

Como una propuesta reciente se presentó el protocolo SPDY, cuyo objetivo primordial es mejorar el rendimiento del servicio Web. Hoy en día se ha convertido en la base sobre la cual dicho WG está trabajando.

Este proyecto de investigación plantea estudiar las deficiencias mencionadas de HTTP y las mejoras disponibles, entre los cuales se encuentra SPDY. Además, producir un estudio académico de algunas de las condiciones que optimicen su implementación, tanto desde el punto de vista técnico y general como además involucrando un análisis de condiciones geográficas. Por último, se desarrollarán piezas de software que faciliten dichos análisis y sirvan como herramientas de apoyo para una eventual migración a SPDY o a HTTP/2.0.

**Palabras clave:** HTTP, SPDY, protocolos web, www.

## Contexto

Esta presentación corresponde al proyecto de investigación “Estudio y propuestas de protocolos para mejorar la performance de servicios basados en Web”, aprobado por el

Departamento de Ciencias Básicas de la Universidad Nacional de Luján, para el cual el primer autor cuenta con una beca de investigación. categoría “perfeccionamiento” de la Secretaría de Ciencia y Técnica de la mencionada universidad.

## Introducción

La Internet de hoy es una Red global omnipresente, donde el crecimiento cuasi-exponencial de la demanda de servicios de información ha “exprimido” en buena medida las capacidades y posibilidades de todos los protocolos que han sido pilares fundacionales de la misma desde un principio.

El protocolo HTTP, que pertenece a este selecto grupo y es uno de los protagonistas principales en el uso de la red, cuenta ya con más de dos décadas de haberse desarrollado (la versión 0.9 en 1991) y más de una de haber visto su última revisión, la versión 1.1 en 1999 [12].

Pero, según varias investigaciones [1,4,20], adolece de debilidades bajo ciertas condiciones, además de algunos problemas de larga data [3, 6,18,25], que han sido potenciados por la riqueza (multimedia, interactividad, etc.) contenida y provista mediante las páginas web actuales.

Esto tiene diversas consecuencias, donde por un lado impacta directamente en el tamaño, cantidad y variedad de recursos que HTTP debe transportar.

Y por el otro, el funcionamiento mismo de HTTP v1.1, utilizando conexiones persistentes (*Keep-Alive*) más *Pipelining* (ambos para optimizar el flujo de datos a nivel de TCP), sumado a múltiples conexiones en paralelo para incrementar la velocidad en la descarga, no es suficiente para conseguir transferencias

eficientes de una página web más todos los recursos que ésta referencia, desaprovechando la capacidad de la red que se dispone y afectando tanto a usuarios como proveedores [5,15].

Si bien su característica de *stateless* permite alta flexibilidad y escalabilidad, el uso de *Cookies* para mantener el estado es percibido usualmente como error de diseño [17]. Además, obliga a que los clientes web utilicen técnicas de *polling* para obtener notificaciones del servidor, cuando sería más eficiente que éstos notifiquen al navegador cuando sea necesario. Asociado con el punto anterior, pero también con la complejidad de la web en la actualidad, la cantidad de encabezados por cada petición HTTP se ha incrementado considerablemente, aumentando la cantidad de tráfico a transportar.

Por otro lado, la falta del soporte de prioridades en la obtención de recursos también puede verse como una desventaja que originalmente no existía por la sencillez de las páginas y servicios web. Y, al ser un protocolo definido mediante directivas en formato de texto ASCII, el procesamiento computacional de sus órdenes y validación de su especificación es más complejo y propenso a errores que si fuera binario. Las redes móviles y satelitales bajo estas nuevas condiciones son casos extremos y particulares que deben ser analizados en detalle.

Finalmente, la falta de confidencialidad, en esta etapa de plena vigencia de redes *wireless*, es otra problemática subsanada por TLS (*Transport Layer Security*) [11], pero que no ha reemplazado a HTTP, sino que la ha complementado de manera opcional.

Por ende, a nivel mundial, se considera que es momento de revisar si el funcionamiento de la web a este nivel puede ser modificado, mejorado, renovado y/o optimizado de acuerdo a los tiempos que corren [7,8,10]; aprovechando la confiabilidad y velocidad de los enlaces sobre los equipos de conectividad de los que se dispone hoy en día, transportando el tipo y volúmenes de recursos Web actuales.

Aún así, esta revisión debe llevarse a cabo respetando las enormes virtudes de las especificaciones preexistentes, y sin descuidar asuntos sumamente importantes como la compatibilidad hacia atrás, complejidad de implementación, facilidades para la migración,

etc., a fin de maximizar las posibilidades de resistir las diferentes pruebas en el campo, para que finalmente se defina una versión HTTP/2.0.

Es por eso que desde que fue presentado el protocolo SPDY como un proyecto abierto [7], a fines de 2009, ha convocado interés de muchos investigadores y organizaciones fuera de Google, su principal impulsor, dado que persigue los siguientes objetivos de diseño:

- Reducir al mínimo la latencia necesaria para la descarga de recursos HTTP.
- Reducir al mínimo el *overhead* (y por consiguiente el ancho de banda) en la subida y descarga de recursos HTTP.
- Minimizar la complejidad de instalación e implementación.
- No requerir ningún cambio en los contenidos de los sitios web.

SPDY, en su versión actual (número 3) [2], presenta las siguientes características técnicas como más salientes:

- Multiplexación de peticiones HTTP concurrentes en una única sesión TCP.
- Encabezados HTTP comprimidos.
- Niveles de prioridades en las peticiones de recursos HTTP.
- *Server-Pushed Streams*: Recursos enviados desde el servidor al cliente sin existir una petición previa exclusiva y asociada de éste.
- Implementa una capa de compatibilidad con HTTP, preservando su semántica: *Cookies*, *headers*, *encoding*, etc., funcionan de la misma manera que HTTP.
- Fuertemente acoplado con HTTP; la integración de protocolos forma parte de la especificación. Sin embargo, SPDY puede ser utilizado para transportar otros protocolos.
- Podría considerarse ubicado sobre TCP y entre SSL/TLS y HTTP, en la capa 5 (Sesión) del modelo OSI.
- En la práctica funciona sobre TCP y TLS/SSL, aunque no está atado a éstos; sí requiere un transporte confiable de mínima.

Esto permitiría, en principio:

- Mayor Escalabilidad: poder manejar más cantidad de usuarios y atenderlos más rápido desde el punto de vista del proveedor de contenidos.
- Velocidad: que los usuarios obtengan una mejora sustancial percibible en la carga de páginas y recursos Web.
- Mayor Seguridad: ya que de facto se utiliza TLS para la implementación.
- Compatibilidad hacia atrás: migración incremental y transparente para los usuarios.

Sin embargo, los estudios y análisis formales realizados hasta el momento, tanto por parte de SPDY como de otras propuestas desarrolladas en paralelo [10,16,19,29] son exiguos, y generalmente provienen del sector industrial, lo que da origen a esta propuesta de investigación académica.

### Líneas de Investigación y Desarrollo

Entre los diferentes trabajos que se están llevando adelante se encuentran tres líneas principales:

- **Estudio y análisis de las debilidades de HTTP** y bajo qué circunstancias ocurren, basándose tanto en trabajos en curso como en experimentación propia. Además, se investiga el protocolo SPDY, los requerimientos para su funcionamiento y algunas de las ideas de las alternativas, tomando en consideración las ventajas y desventajas que presenta frente al protocolo HTTP.
- **Modelización de la red a nivel local y regional.** La intención aquí es poder realizar pruebas y experimentación sobre condiciones de operación de la red representativas de la región ya que todos los estudios actuales reportan resultados sobre redes de condiciones sustancialmente mejores pertenecientes a los principales países.
- **Desarrollo de librerías de código** tanto para dar soporte a los experimentos, documentar ejemplos, y desarrollar los objetivos de las fases posteriores [31]. Se pretende implementar programas para facilitar la migración a SPDY, como servicios intermediarios (proxy) y validadores de conformidad con las especificaciones

propuestas, como también una guía de recomendaciones a tener en cuenta para estudiar e implementar estos cambios.

### Resultados y Objetivos

Se consiguió desarrollar un muxer/demuxer de frames SPDY (versión 2 y 3) con ejemplos básicos en sockets TCP y TLS sobre TCP en el lenguaje Python, con el fin inicial de comprender mejor lo estudiado, más allá de perseguir los objetivos mencionados anteriormente [31].

Basado en los resultados de las etapas iniciales, se propondrán ajustes sobre los protocolos estudiados y se realizarán pruebas de campo (redes locales e Internet), a nivel local y regional. Además, se propondrán e implementarán mecanismos migración a SPDY y/o HTTP/2.0

Algunos resultados iniciales sugieren que SPDY permite disminuir en tiempo de carga total de una página entre un 39%-55% accediendo por un canal cifrado y entre un 27%-60% por un canal no cifrado [30].

Sin embargo, aún no está claro cómo los diversos factores (de la red y de las páginas) inciden en esta mejora. Se espera poder definir un modelo que permita predecir la reducción del tiempo de carga de acuerdo a ciertas características de las páginas, lo que permitiría establecer recomendaciones que apunten a mejorar la performance (como las hay para diseño y posicionamiento, por ejemplo).

### Formación de Recursos Humanos

El presente trabajo corresponde a una beca de investigación del primer autor, quien – además – se encuentra finalizando su maestría en Redes de Datos de la Facultad de Informática de la UNLP.

Complementariamente, se prevé incorporar al menos un pasante alumno al proyecto durante el año en curso y realizar un trabajo final de licenciatura en el área.



## Referencias

- [1] Mike Belshe, "More Bandwidth Doesn't Matter (much)". Google Inc. 2010.  
<https://docs.google.com/viewer?pid=sites&srcid=Y2hyb21pdW0ub3JnfGRldnxneDoxMzcyOWI1N2I4YzI3NzE2>
- [2] Mike Belshe, Roberto Peon, "SPDY Protocol v3", Google Inc., IETF Draft, 2012.
- [3] Yogesh Bhole and Adrian Popescu. 2005. "Measurement and Analysis of HTTP Traffic". *J. Netw. Syst. Manage.* 13, 4 (December 2005), 357-371.
- [4] Michael Butkiewicz, Harsha V. Madhyastha, and Vyas Sekar. 2011. "Understanding website complexity: measurements, metrics, and implications". In Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference (IMC '11). ACM, New York, NY, USA, 313-328.
- [5] Nandita Dukkkipati, Tiziana Refice, Yuchung Cheng, Jerry Chu, Tom Herbert, Amit Agarwal, Arvind Jain, and Natalia Sutin. 2010. "An argument for increasing TCP's initial congestion window". *SIGCOMM Comput. Commun. Rev.* 40, 3 (June 2010), 26-33.
- [6] Roy T. Fielding. "Waka: A replacement for HTTP". 2002.  
[http://gbiv.com/protocols/waka/200211\\_fielding\\_apachec\\_on.ppt](http://gbiv.com/protocols/waka/200211_fielding_apachec_on.ppt)
- [7] The Chromium Projects, "SPDY: An experimental protocol for a faster web". Google Inc., 2009.  
<http://dev.chromium.org/spdy/spdy-whitepaper>
- [8] Preethi Natarajan et al, "Using SCTP as a Transport Layer Protocol for HTTP", Cisco Systems, IETF Draft. 2009.
- [9] IETF, "Httpbis Working Group", 2012.
- [10] IETF, "Httpbis HTTP/2.0 Proposals", 2012.
- [11] IETF Standard, "The TLS Protocol Version 1.0", 1999.
- [12] IETF Standard, "Hypertext Transfer Protocol – HTTP/1.1", 1999.
- [13] A. Langley, "Transport Layer Security (TLS) False Start", Google Inc., IETF Draft, 2010.
- [14] A. Langley, "Transport Layer Security (TLS) Next Protocol Negotiation Extension", Google Inc., IETF Draft, 2012.
- [15] Robert L.R. Mattson, Enhancing HTTP to Improve Page and Object Retrieval Time With Congested Networks. La Trobe University. 2008.
- [16] Robert L. R. Mattson and Somnath Ghosh. 2009. "HTTP-MPLEX: An enhanced hypertext transfer protocol and its performance evaluation". *J. Netw. Comput. Appl.* 32, 4 (July 2009), 925-939.
- [17] Jeffrey C. Mogul. 1995. "The case for persistent-connection HTTP". *SIGCOMM Comput. Commun. Rev.* 25, 4 (October 1995), 299-313. DOI=10.1145/217391.217465.
- [18] Thomas M. Kroeger, Darrell D. E. Long, and Jeffrey C. Mogul. 1997. "Exploring the bounds of web latency reduction from caching and prefetching". In Proceedings of the USENIX Symposium on Internet Technologies and Systems on USENIX Symposium on Internet Technologies and Systems(USITS'97). USENIX Association, Berkeley, CA, USA, 2-2.
- [19] Gabriel Montenegro et al., "HTTP Speed+Mobility", Microsoft Corp., IETF Draft, 2012.
- [20] Jean-Francois Raymond, "Traffic Analysis: Protocols, Attacks, Design Issues, and Open Problems", H. Federrath (Ed.), Designing Privacy Enhancing Technologies, International Workshop on Design Issues in Anonymity and Unobservability, Berkeley, Proceedings, LNCS 2009.
- [21] Steven Rosenberg, Surbhi Dangi, Isuru Warnakulasooriya. "Data and Network Optimization Effect on Web Performance". Carnegie Mellon University, Feb 2012.  
[http://repository.cmu.edu/silicon\\_valley/91/](http://repository.cmu.edu/silicon_valley/91/)
- [22] Marshall T. Rose, "BEEP: The Definitive Guide", 2002. ISBN: 978-0596002442.
- [23] Chris Strom, "The SPDY Book: Making Websites Fly", 2011.
- [24] W. Tarreau et al, "Proposal for a Network-Friendly HTTP Upgrade", IETF Draft, 2012.
- [25] Venkata N. Padmanabhan, Jeffrey C. Mogul, "Improving HTTP latency", *Computer Networks and ISDN Systems*, Volume 28, Issues 1–2, December 1995, Pages 25-35, ISSN 0169-7552, 10.1016/0169-7552(95)00106-1.
- [26] "MUX Overview". W3C. 1998.
- [27] "SMUX Protocol Specification". W3C. 1998.
- [28] Greg White. "Analysis of Google SPDY and TCP Initwnd". Cable Television Laboratories, Inc., 2012.  
[http://www.cablelabs.com/downloads/pubs/Analysis\\_of\\_Google\\_SPDY\\_TCP.pdf](http://www.cablelabs.com/downloads/pubs/Analysis_of_Google_SPDY_TCP.pdf)
- [29] Wenxuan Zhou, Qingxi Li, Matthew Caesar, and P. Brighten Godfrey. 2011. "ASAP: a low-latency transport layer". In Proceedings of the Seventh Conference on emerging Networking EXperiments and Technologies (CoNEXT '11). ACM, New York, NY, USA, Article 20, 12 pages.
- [30] Bryce Thomas, Raja Jurdak, and Ian Atkinson. 2012. SPDYing up the web. *Commun. ACM* 55, 12 (December 2012), 64-73. DOI=10.1145/2380656.2380673
- [31] Marcelo Fidel Fernández, biblioteca Python-SPDY para Python versión 2.6 a 3.3.  
<https://github.com/marcelofernandez/python-spdy>



## Redes Inalámbricas de uso Comunitario

Eduardo Rodríguez, Claudia Deco, Luciana Burzacca, Mauro Pettinari

Departamento de Investigación Institucional,  
Facultad de Química e Ingeniería, Universidad Católica Argentina, 2000 Rosario, Argentina

{erodriguez, cdeco, lburzacca, mauro\_pettinari}@uca.edu.ar

### Resumen

Las redes de malla inalámbricas son un dominio rápidamente creciente y esto trae muchos desafíos. En particular, un desafío difícil e inmediato es el enrutamiento efectivo debido a la volatilidad típica de tráfico en topologías complejas. Trabajos recientes han demostrado que el tráfico inalámbrico es muy variable y difícil de caracterizar. Comprender el impacto de la incertidumbre de la demanda en el ruteo y el diseño de algoritmos de enrutamiento para proporcionar robustez es relativamente un problema de investigación sin explotar. Sin embargo, tiene un gran impacto en el rendimiento de una red y será esencial para su desarrollo en los próximos años. El algoritmo de ruteo usado siempre debería asegurar que la información tome el camino más apropiado de acuerdo a una métrica. Por lo tanto, tomamos esto como nuestro objetivo principal de investigación: Caracterizar y resolver el problema de enrutamiento en redes de malla inalámbricas robustas.

**Palabras Claves:** Redes Malladas Inalámbricas, Redes Mesh, Protocolos.

### Contexto

Esta línea de I+D se está llevando a cabo a través del proyecto del Departamento de Investigación Institucional, Facultad de Química e Ingeniería, Universidad Católica Argentina.

El Proyecto involucrado es:

- Optimización de Redes Inalámbricas de uso Comunitario (2011 – 2014)

Además, los integrantes de esta línea, trabajan en conjunto con otros grupos:

- Asociación Civil Nodo TAU.
- Proyecto TRICALCAR (Tejiendo redes inalámbricas en América Latina y Caribe - [www.wilac.net/tricalcar](http://www.wilac.net/tricalcar)).

### Introducción

Las redes malladas inalámbricas (Wireless Mesh Networks) han tenido un gran éxito en la historia de las ciencias de la computación y de la ingeniería. Sus aplicaciones son numerosas en el dominio industrial, militar y comercial. Son en particular un dominio rápidamente creciente y esto trae muchos desafíos. En particular, un desafío difícil e inmediato es el enrutamiento efectivo debido a la volatilidad típica de tráfico en topologías complejas. Muchos estudios han intentado resolver el problema de enrutamiento mediante métodos heurísticos, pero este enfoque no proporciona los límites de cuán bien se asignan los recursos. Sin embargo, este tipo de investigación generalmente asume que el tráfico de demandas de la red es estático y conocido de antemano. Como resultado, estos algoritmos tienden a sufrir un desempeño pobre. De hecho, trabajos recientes han demostrado que el tráfico inalámbrico es muy variable y difícil de caracterizar. Comprender el impacto de la incertidumbre de la demanda en el ruteo y el diseño de algoritmos de enrutamiento para proporcionar robustez, es relativamente un problema de investigación aún incipiente.

Las redes Mesh abiertas son redes ad-hoc descentralizadas que no se basan en infraestructuras previas, como routers o puntos de

acceso. En su lugar, cada nodo participa en el enrutado, siendo él mismo un router y enviando datos de otros, y de ese modo la determinación de las rutas se hace dinámicamente, basándose en la conectividad que va surgiendo. Para ello, necesitan de protocolos que viabilicen ese comportamiento.

Es de suma importancia el análisis de la performance de diferentes protocolos de comunicación que deben interactuar con diversos dispositivos que hacen al enlace de los nodos de la red a los fines de establecer la integración tecnológica disponible. No menos importante es la determinación de la relación costo / beneficio de una determinada implementación. El conocimiento en tiempo real de la configuración topológica de la red, mediante el uso de distintas herramientas de hardware y software, nos permite el monitoreo del comportamiento y sus alcances. Todo ello posibilita optimizar la red para que brinde un mejor servicio. En general, la optimización se basa en lograr el mejor camino para enrutar los paquetes de datos, sin demoras o con una demora mínima en función de lograr un mejor aprovechamiento de los recursos utilizados.

Una Red Mallada Inalámbrica (Mesh) es una red compuesta por nodos organizados en una topología de malla. Son redes en las cuales la información es pasada entre nodos en una forma de todos contra todos y en una jerarquía plana, en contraste a las redes centralizadas. Toda variación no prevista en el diseño, puede cambiar su topología, afectar a la distribución de carga de la red y al rendimiento general [1].

Las ventajas que presenta frente a otras redes son el bajo costo al utilizar enlaces inalámbricos, la facilidad de aumentar el área de cobertura incluyendo nuevos nodos, la robustez que presenta ante fallos al disponer de rutas alternativas y la capacidad de transmisión que permiten aplicaciones a los usuarios en tiempo real de voz, video y datos. A la hora de incrementar el número de nodos, no es necesario cambiar infraestructuras como en el caso de las redes cableadas, se puede incluir un nuevo nodo en cualquier momento y lugar. Como consecuencia el costo de este tipo de redes inalámbricas es mucho menor que en las redes cableadas, ya que no hay que invertir en materiales de cableado y en estudios enfocados a la unión más óptima de los nodos.

En una red mallada la información atraviesa múltiples saltos y no hay necesidad de una unidad centralizada que controle el modo de transmisión. La comunicación se realiza entre los nodos directamente. Cada nodo puede ser origen y destino de los datos o encaminar la información de otros nodos. Las redes malladas inalámbricas son robustas al tener varios caminos disponibles entre el nodo origen y el destino, de modo que el servicio no se ve afectado por la caída de un nodo o por la ruptura de un enlace. El algoritmo de ruteo usado siempre debería asegurar que la información tome el camino más apropiado de acuerdo a una métrica. Una métrica es el valor por el cual los protocolos determinan cuál ruta tomar o con cuál nodo comunicarse.

Con respecto al hardware para redes malladas, prácticamente cualquier nodo inalámbrico puede convertirse en un nodo Mesh simplemente mediante modificaciones de software.

Una de las debilidades y limitaciones de las redes Mesh es la latencia (el retardo de propagación de los paquetes), que crece con el número de saltos. Los efectos del retardo son dependientes de la aplicación. Por ejemplo los correos electrónicos no son afectados por grandes latencias, mientras que los servicios de voz son muy sensibles a los retardos.

### **Protocolos de Encaminamiento**

La principal función de los protocolos de encaminamiento es seleccionar el camino entre el nodo fuente y destino de una manera rápida y fiable. Las redes malladas inalámbricas pueden utilizar los protocolos de encaminamiento de otras redes ya existentes, pero modificándolos para que funcionen correctamente con ellas. Si se elige esta opción, el protocolo de encaminamiento modificado debe asegurar las principales características que son el número de saltos, el rendimiento, la tolerancia a fallos, el equilibrado de carga, la escalabilidad y el soporte adaptativo.

Podemos clasificar los protocolos de encaminamiento en base al alcance de las transmisiones, al modo que descubren las rutas y en base al algoritmo que implementan.

En base al alcance de las transmisiones, se clasifican en unicast y multicast. Los protocolos unicast transmiten los paquetes de datos uno a uno. El envío de datos se realiza desde un único emisor a un único receptor, mientras que en los

multicast el método de transmisión es de uno-a-muchos, es decir, se envían los paquetes de datos a múltiples destinos simultáneamente. Un caso especial es la transmisión broadcast, donde se envía la información a todos los nodos de la red.

Basándose en el modo que descubren las rutas, hay dos grandes grupos: los que se basan en la topología de la red o los que se basan en la posición de los nodos. A su vez, los protocolos que se basan en la topología se subdividen en reactivos y proactivos. Los protocolos proactivos tienen un conocimiento exhaustivo del estado de la red, de modo que cuando se necesita una ruta, ésta ya es conocida y está lista para usarse de manera inmediata. En escenarios cambiantes no es muy aconsejable, porque se precisa que las tablas de encaminamiento estén actualizadas mediante el envío continuo de mensajes. Esto provoca una sobrecarga de mensajes de control en la red. Los protocolos reactivos sólo obtienen información de encaminamiento cuando es necesario. En consecuencia la sobrecarga de la red es menor que en los protocolos proactivos, mientras que el tiempo en establecer la comunicación aumenta.

También existen protocolos que combinan los anteriores y son los que se denominan protocolos de encaminamiento híbridos. Aprovechan las ventajas de cada uno de los protocolos: utiliza el encaminamiento proactivo cuando los nodos están cerca y utiliza el encaminamiento reactivo cuando los nodos están lejos. También se utiliza el encaminamiento reactivo cuando los caminos son utilizados en pocas ocasiones.

Por último, se pueden clasificar a los protocolos de encaminamiento en base al algoritmo que implementan, que puede ser estado del enlace o vector de distancias. En los protocolos de estado de enlace todos los nodos tienen una tabla con el mapa de red completo. En esta tabla se define el enlace y la distancia para llegar de un nodo a otro y cada nodo envía cada cierto tiempo la información de cómo llegar a sus vecinos. En los protocolos de vector de distancia cada nodo conoce los vecinos conectados a él y los costos de dichos enlaces. Cada cierto tiempo el nodo transmite su tabla de encaminamiento a sus vecinos y éstos recalculan su tabla de encaminamiento si existe nueva información. Los nodos no conocen toda la topología de la red.

## Líneas de investigación y desarrollo

El objetivo general del proyecto es caracterizar y resolver el problema de enrutamiento en redes de malla inalámbricas robustas de manera de optimizar el protocolo BATMAN para su uso en redes mesh inalámbricas comunitarias con uso de VoIP.

Los principales objetivos específicos del presente proyecto son:

- Evaluar los diferentes protocolos utilizados en una red Mesh aptos para servicios de VoIP en una comunidad abierta, tomando como referencia el protocolo BATMAN (Better Approach To Mobile Adhoc Networking)
- Analizar y proponer diferentes maneras de optimizar el protocolo BATMAN.
- Proponer topologías y parámetros concretos a analizar para evaluar la optimización lograda.
- Determinar topologías y/o métodos de acceso alternativos para los casos en que la optimización obtenida no dé los resultados deseados o no pueda aplicarse por condiciones particulares de la implementación.
- Diseñar e implementar una base de datos propia de parámetros que permitan determinar la eficiencia de los protocolos afectados a estas aplicaciones.
- Establecer el impacto medioambiental por el uso de estas redes.
- Fomentar y facilitar la participación de alumnos de la Carrera de Licenciatura en Sistemas y Computación.
- Realizar transferencia tecnológica a una comunidad específica para la implementación de su propia red de conectividad y telefonía.
- Publicar y divulgar los resultados en revistas especializadas, congresos y exposiciones

Para alcanzar estos objetivos se han planteado distintas tareas:

- Actualización del conocimiento y estado del arte de redes inalámbricas de tipo comunitario con y sin VoIP y los protocolos utilizados.

- Relevamiento de implementaciones de redes inalámbricas y análisis de los resultados obtenidos en cada caso.
  - Elección de la topología adecuada y montaje de la red experimental que servirá de base para las investigaciones
  - Propuesta de los métodos a utilizar y los parámetros que se evaluarán en la red experimental. Diseño e implementación de una base de datos propia con estos parámetros.
  - Análisis de tráfico y prestaciones de la red experimental y comparación con los relevamientos efectuados.
  - Diseño de las mejoras e implementación de un prototipo mejorado de protocolo a partir de las conclusiones efectuadas.
  - Propuesta de topologías y/o métodos de acceso alternativos para los casos en que la optimización obtenida no dé los resultados deseados o no pueda aplicarse por condiciones particulares de la implementación.
  - Análisis del impacto medioambiental, mediante la implementación de instrumental adecuado y el diseño e implementación de una base de datos propia para almacenar estas mediciones. Estudio de las consecuencias, contribuciones y aplicaciones de los resultados de dicho análisis.
  - Evaluación de los resultados.
  - Publicación de resultados. Se realizará la publicación de resultados parciales durante el proyecto y de resultados finales al completarse el mismo
- Se efectuó un relevamiento por internet de distintas instalaciones de redes mesh existentes en distintos países entre ellas: Guifi Net, Village Telco, Freifunk.net, CUWiN, The Dharamsala Community Wireless Mesh Network, BuenosAiresLibre, Lugro-mesh, Wireless Ghana, RedLibre, ZittNet, EHAS y WirelessAfrica Group.
  - Se analizaron teóricamente distintos protocolos comparando sus prestaciones y ventajas y en función de esto y del relevamiento de redes existentes en funcionamiento se seleccionaron Batman y OLSR para las pruebas.
  - Se montó una red experimental distribuida en tres edificios del campus de la Universidad a los efectos de tener un campo de pruebas más parecido a la realidad de las redes mesh. En el montaje de esta red se utilizaron equipos de las marcas Linksys (WRT54GL), Ubiquiti (Nonostation 2, Nanostation Loco M2), TP-Link (TL-WR743ND, TLWR842ND). Como firmware se utilizaron distintas versiones de OpenWRT.
  - Actualmente se está analizando el desempeño de la red mesh utilizando protocolo OLSR. Para complementar las pruebas se montaron sobre la red mesh, una central telefónica IP Elastix y con 5 teléfonos internos, tres internos utilizando un ATA (Linksys phone adapter PAP2-NA) y dos por medio de software cliente de centrales IP, también se monto una cámara IP sobre uno de los nodos más alejados.

Las principales ventajas de las redes malladas son el bajo costo ya que se pueden implementar en múltiples dispositivos que están al alcance de todos. Además existen diversos fabricantes que ofrecen soluciones viables para la implementación de este tipo de redes. Otro punto a favor es que las redes malladas inalámbricas son compatibles con las redes existentes. Se analizaron distintos protocolos y se presentaron brevemente sus descripciones para luego compararlos de acuerdo al Tipo de Protocolo, Alcance de trasmisiones, Métrica de ruteo, y si es de Uso Libre o es Propietario. Para elegir el mejor protocolo de encaminamiento hay que estudiar las necesidades de la red y sopesar si se necesita un protocolo que sea rápido en las comunicaciones aunque la transmisión de mensajes de control sea excesiva, o

## Resultados Obtenidos / Esperados

Entre los resultados obtenidos en esta línea de investigación se encuentran:

- Se analizaron principalmente propuestas científicas relacionadas a optimización y prueba de protocolos.



si por el contrario se prefiere una comunicación más lenta pero que no sobrecargue la red con mensajes de control. El dinamismo de la red es crítico para la toma de la decisión, ya que en una red muy dinámica los protocolos proactivos pierden atractivo porque las rutas que descubren están desactualizadas para cuando van a usarse.

## Formación de Recursos Humanos

El equipo de trabajo está integrado por el Magister Eduardo Rodríguez, la Dra. Claudia y los Licenciados Luciana Burzacca y Mauro Pettinari, investigadores de la Universidad Católica Argentina.

Además, para la etapa de implementación y pruebas colaboró también el Licenciado en Sistemas y Computación Santiago Costa. Asimismo, se mantiene abierta la propuesta de tesis de grado y de posgrado, como así la realización de pasantías en el tema.

## Referencias

- [1] I. Akyildiz, X. Wang, W. Wang: "Wireless mesh networks: a survey", "Computer Networks". Vol. 47. No.4 (2005) Pág. 455-487.
- [2] D. Acuña Martínez, R. Roncallo Kelsey: Redes inalámbricas enmalladas metropolitanas. (Octubre 2006) Pág. 46-91.
- [3] Batman <http://www.open-mesh.org/>
- [4] C. Perkins y P. Bhagwat, "Highly Dynamic Destination-Sequenced Distance-Vector Routing (DSDV) for Mobile Computing", In Proceedings of ACM SIGCOMM'94, London, UK, Sep. 1994, pp. 234-244.
- [5] R. Coltun, D. Ferguson, J. Moy, A. Lindem: RFC 5340, OSPF for IPv6. IETF (Julio de 2008).
- [6] J. Chroboczek (2011), The Babel Routing Protocol, RFC 6126
- [7] IEEE 802.11s: Mesh Networking, Extended Service Set (ESS) (July 2011).



## Aplicaciones de Internet de las Cosas

### SIPIA6 - Red de Sensores Inalámbricos con IPv6

Gustavo. Mercado\*, Roberto Borgo<sup>§</sup>, Francisco Gonzalez Antivilo<sup>§</sup>, Carlos Taffernaberry\*, Ana Diedrichs\*, Matías Aguirre\*, María Inés Robles\*, Guillermo Grünwaldt\*, Germán Tabacchi\*, Sebastián Tromer\*, Matías Pecchia\*

\*gridTICS – Grupo UTN de I&D en Tecnologías de la Información y las Comunicaciones  
Departamento de Electrónica - UTN Facultad Regional Mendoza  
Rodríguez 273, Capital - Mendoza  
{gustavo.mercado,carlos.taffernaberry, ana.diedrichs, matias.aguirre}@gridtics.frm.utn.edu.ar

§ Cátedra de Fisiología Vegetal  
Facultad de Ciencias Agrarias – UNCuyo  
Alte Brown 500, Chacras de Coria, Lujan de Cuyo - Mendoza  
rborgo@fca.uncu.edu.ar, fgonzalezantivilo@gmail.com

#### Resumen

El proyecto consiste en el diseño y validación del comportamiento de una red de sensores inalámbricos en la adquisición de parámetros ambientales de parcelas agrícolas, con acceso de forma nativa a Internet, denominada SIPIA6.

Se trata de la continuación de Red SIPIA, actualmente implementada en una parcela agrícola de prueba. La red SIPIA está basada en la norma IEEE 802.15.4, aplicada al entorno agropecuario en el ámbito de agricultura de precisión, con participación de ingenieros e investigadores agrónomos. Gracias a los desarrollos que se han producido en las redes de sensores inalámbricos en los últimos años, especialmente la miniaturización de los dispositivos, han surgido nuevas tendencias en el sector agrícola como la llamada agricultura de precisión. Esta disciplina cubre múltiples prácticas relativas a la múltiple toma de datos para aplicar en la gestión y toma de decisiones de cultivos y animales. Por medio de sensores estratégicamente situados, se realizará un monitoreo de la temperatura y/o la humedad relativa, con el fin de proveer de gran cantidad de datos confiables para la investigación agronómica.

La diferencia principal de la Red SIPIA6 es que sus nodos WSN serán visibles desde Internet y accesibles mediante protocolos y procedimientos estándares. El proyecto permitirá, por lo tanto, extender la visibilidad de una red de sensores agrícola y por lo tanto convertir a SIPIA6 en una aplicación de Internet de las Cosas

#### Palabras Clave

Redes de Sensores, IEEE 802.15.4, Agricultura de Precisión, Internet de las Cosas, 6lowPAN, , Data Logger, Sensado remoto, Microclima

#### Contexto

El proyecto se enmarca en el convenio de cooperación científica/tecnológica entre la Facultad de Ciencias Agrarias de la UNCuyo y la Facultad Regional Mendoza de la Universidad Tecnológica Nacional, con financiamiento de ambas instituciones y es llevado adelante por personal del GRIDTICS y la Cátedra de Fisiología Vegetal.

En la FRM se inserta en el proyecto PID 25/J071 “LIVRES: Análisis y evaluación de características relevantes de las redes de sensores inalámbricos aplicadas al manejo y sensado en agricultura de precisión” y el proyecto PICT2010-22 “RED SIPIA - Red de Sensores inalámbricos para investigación Agronómica”.

En la FCA se inserta en los siguientes proyectos:

-2010 - proyecto “Mecanismo de resistencia a temperaturas subcero en tejidos leñosos de Vitis vinifera cv Malbec. Doctorando: Ing. Agr. González Antivilo.

#### Introducción

Gracias a los avances y reducción de costos en dispositivos electrónicos y de comunicación inalámbrica, es posible construir dispositivos sensores multifuncionales y multipropósito de bajo costo que operan con poca energía, de un tamaño pequeño, y de una capacidad de comunicación a corta distancia. Estos dispositivos se denominan motes [1], y constan de una unidad de procesamiento con un poder de cómputo mínimo, memoria, una unidad de comunicación inalámbrica y uno o varios dispositivos de sensado que capturan parámetros como temperatura, aceleración, humedad, etc. Un conjunto de motes comunicados entre sí es lo que conocemos como una red de sensores inalámbrica (Wireless Sensor Network o WSN) [2].

Una red de sensores está compuesta por varios motes que se encuentran esparcidos en un área determinada. La distribución de los mismos puede ser aleatoria o planeada, lo cual permite su uso en prácticamente en cualquier ambiente físico. Esta característica es provista por un conjunto de protocolos y algoritmos para redes de sensores. Otro de los beneficios que caracteriza a una red de sensores es que sus nodos pueden trabajar de modo cooperativo, aumentando sus posibilidades de aplicación. Además, gracias a la capacidad de cómputo de los mismos, más allá que sea mínima, éstos pueden parcialmente, procesar los datos capturados antes de comunicarlos a la red.

La agricultura de precisión consiste en el uso de sistemas de información basados en diversas tecnologías aplicadas al ámbito de la producción agraria. Algunas de las tecnologías aplicables serían; Redes de Sensores Inalámbricos (WSN), Sistemas de Posicionamiento Global (GPS), evaluación de espectroscopia en el

infrarrojo cercano (NIR), Sistemas de Información Geográfico (SIG). Sistemas que proporcionan medios de observación, evaluación y control de las prácticas agrícolas.

La búsqueda de modelos de cultivos, tanto protegidos (en invernaderos) como en campo abierto, ha sido tradicionalmente un reto puesto que la recolección de información de las diversas variables involucradas es un proceso complejo. Con el advenimiento de las redes de sensores la recolección de información puede precisarse arbitrariamente tanto en el tiempo como en el espacio.

Tradicionalmente estos modelos se han establecido de forma empírica, basados en la experiencia, observación visual e intuición de los agricultores [3], lo que usualmente implica que los modelos están lejos de ser óptimos, siendo una de las limitantes más importantes la escasa resolución espacial alcanzada. Las redes de sensores permiten recopilar información con nuevos niveles de resolución espacial y temporal que brinda el soporte para la elaboración de nuevos modelos o el perfeccionamiento de los existentes.

Ya en el año 2004 Beckwith et. al. [4] establecieron ventajas de utilizar redes de sensores, contra sensores aislados o recolectores de datos (data loggers) en aplicaciones agronómicas, particularmente en el caso de viñedos. Aún en pequeñas extensiones menores a una hectárea las redes requieren menos tiempo de instalación y de recolección de datos, y son además más fáciles de mantener. Estas proveen además capacidad de reaccionar en tiempo real, de acuerdo a los datos recolectados, ya sea disparando alarmas sobre situaciones que deben ser atendidas manualmente, o activando dispositivos que puedan compensar eventos no deseados, por medio de mecanismos de irrigación, sistemas de ventilación, iluminación, etc. Un año antes la Universidad de Carnegie Mellon desarrolló una red de sensores para facilitar las operaciones en un vivero [5], midiendo temperatura y humedad relativa del aire y del suelo y luz, activando actuadores dependiendo de los datos medidos para alterar la circulación de aire, la luminosidad y la temperatura.

En [6] se presenta un panorama de la tecnología de redes inalámbricas de sensores, aptas para ser aplicadas en aplicaciones agrícolas y de la industria alimentaria. Se presentan comparaciones entre estándares inalámbricos de transmisión (WiFi, Bluetooth y ZigBee). También se mencionan los estándares para transductores inteligentes (IEEE1451) y su enlace (IEEE1451.5). En el mencionado documento se clasifican las aplicaciones de redes de sensores en agricultura y la industria alimentaria en 5 grupos, incluyendo el que nos involucra, agricultura de precisión.

La visión detrás de la Internet de las Cosas [7] es que los dispositivos embebidos, también llamados "smart objetos", están cada vez más universalmente conectados a Internet y que son una parte integral de la misma.

La Internet de las cosas, a veces denominada "internet embebida orillera", es un cambio mayúsculo y el mayor desafío a la Internet actual. La misma está "hecha" con dispositivos embebidos conectados a Internet, lo que incluye sensores, máquinas, lectores de RFID y

equipamiento de automatización de edificios, solo para nombrar unas pocas aplicaciones. El exacto tamaño de la Internet de las cosas es difícil de estimar pero se asume que pronto su tamaño excederá en de la Internet actual.

El Wireless Embedded Internet es un subconjunto de la Internet de las cosas y son aquellos dispositivos embebidos de recursos limitados, a menudo operados por baterías y conectados a Internet a través de redes inalámbricas de baja potencia y bajo ancho de banda. 6LoWPAN [8] fue desarrollado para hacer posible Wireless Embedded Internet simplificando las funcionalidades de protocolo de internet IPv6, definiendo un encabezamiento muy compacto y tomando en cuenta la naturaleza de las redes inalámbricas.

#### **El estándar IEEE 802.15.4**

El estándar IEEE 802.15.4 [9] define las características de la capa física y de la capa de control de acceso al medio (MAC) para redes inalámbricas de área personales (WPAN, Wireless Personal Area Networks) de baja tasa de transmisión. Este estándar no establece un nivel de red pero sí plantea parámetros para su implementación. Las ventajas de utilizar el estándar IEEE 802.15.4 es que permite la utilización de dispositivos de fácil instalación que proveen transmisiones confiables a distancias cortas a un precio muy bajo. Por otro lado, el estándar IEEE 802.15.4 permite proporcionar un tiempo de vida razonable al utilizar fuentes de energía limitada (e.j. baterías alcalinas) y al mismo tiempo proporciona una pila de protocolos simple. Las características generales, presentadas son: tasas de transferencias de 250Kb/S, 40Kb/S y 20Kb/S, manejo de redes en estrella y malla (peer-to-peer), direccionamiento corto con 16 bits y extendido con 64 bits, garantía del manejo de las ranuras de tiempo, detección de los niveles de energía recibidos, indicadores de calidad en el enlace así como de conmutación de canales para recibir paquetes, acceso al canal por CSMA/CA. El estándar de comunicaciones IEEE 802.15.4 ofrece dos tipos de dispositivos que participan en la red, estos son: FFD (Full Function Device), dispositivo con todas las funciones y RFD (Reduced Function Device) dispositivo con funciones reducidas. Dependiendo de la aplicación el estándar permite operar en una de dos topologías: la estrella (Star) o punto a punto (peer-to-peer).

#### **El estándar 6LoWPAN**

6LoWPAN es un acrónimo de "IPv6 sobre redes área personal inalámbricas y baja potencia" [8] y también 6LoWPAN es el nombre de un grupo de trabajo en el área de Internet del IETF.

El concepto 6LoWPAN fue originado por la idea de que "el Protocolo de Internet puede y debe ser aplicada incluso a los más pequeños dispositivos", [10] y que los dispositivos de baja potencia con capacidades de procesamiento limitadas deben ser capaces de participar en el Internet de las Cosas.

El objetivo inicial [11] fue definir una capa de adaptación para hacer frente a las exigencias impuestas por IPv6, tales como el aumento del tamaño de la dirección y la

MTU byte de 1280. Se han definido mecanismos de encapsulación y compresión de cabecera que permiten a los paquetes IPv6 ser enviados y recibidos en redes basadas en de IEEE 802.15.4 con MTU más pequeñas. La compresión produce cabeceras veces tan pequeñas de como sólo 4 bytes, mientras que al mismo tiempo permite el uso de diferentes tipos de redes de malla y gestiona la fragmentación y reensamblaje donde sea necesario [12].

El objetivo de la creación de redes IP para la comunicación de radio de baja potencia son las aplicaciones que requieren conectividad inalámbrica a Internet a velocidades de datos bajas en dispositivos con muy limitadas cualidades. Los ejemplos podrían incluir, pero no están limitados a: aplicaciones de automatización y entretenimiento en entornos del hogar, la oficina y la fábrica. También a la adquisición de datos en ambientes agrícolas o naturales.

### Líneas de investigación y desarrollo

El grupo de Redes de Sensores Inalámbricos, perteneciente al grupo UTN GridTICS, se forma por iniciativa de algunos de sus miembros en el año 2008 y comienza con el estudio de la tecnología para la capacitación y luego la adquisición de elementos y dispositivos para la conformación del laboratorio de WSN. El grupo ha realizado publicaciones [13], presentación en congresos [14,15,16,17], cursos de grado y posgrado y asistencia a tesinas de grado [18,19] y tesis de posgrado en curso.

El grupo de tecnología IPv6, perteneciente al grupo UTN GridTICS, se constituye en 2005 y ha tenido una vasta actividad y experiencia y es reconocido como uno de los grupos pioneros en IPv6 de la región. El grupo ha realizado publicaciones [20], presentación en congresos [21,22,23], cursos de grado y posgrado y asistencia a tesinas de grado [24] y tesis de posgrado. Además de participar activamente en las iniciativas de ISOC y de LANIC para la promoción y difusión de IPv6 [25,26,27], siendo también socio activo de la IPv6 Task Force Argentina [28].

La cátedra de Fisiología de la UNC posee una amplia trayectoria en la investigación de respuestas fisiológicas y ecofisiológica de cultivos de gran importancia económica para la región, como son, la vid, ajo y la cebolla. La cátedra y sus integrantes han contribuido al conocimiento de respuestas fisiológicas relacionadas a la calidad de los frutos de vid dependiendo de condiciones ambientales y culturales, y la participación de sus fotoasimilados. Los integrantes del grupo de trabajo acreditan presentaciones y asistencia a congresos [29, 30, 31, 32, 33], y participación en estudios de posgrado [34, 35, 36] y publicaciones [37, 38, 39, 40, 41, 42].

### Objetivos y Resultados

#### Objetivo principal

Diseñar y validar el comportamiento de una red de sensores inalámbricos en la adquisición de parámetros

ambientales de parcelas agrícolas, con acceso de forma nativa a Internet.

#### Objetivo secundario

Adaptar la red de sensores inalámbricos SIPIA para convertirla en una aplicación de Internet de las Cosas

Para establecer los requerimientos para la aplicación de redes de sensores inalámbricos en investigación agrícola:

- Se plantearon escenarios típicos de uso de sensores inalámbricos.

- Se plantearon escenarios de condiciones ambientales a las cuales estarán sometidos los sensores. Se experimentarán las diferentes condiciones de temperatura, humedad, radiación, e incluso condiciones de pH (si estuviese dentro del vegetal) a los cuales podrían estar sometidos los cultivos.

- Se determinaron los requerimientos físicos y mecánicos de los sensores. En base a los escenarios planteados anteriormente determinar los requerimientos físicos para el correcto funcionamiento. Esto incluye aspectos de protección física contra factores como radiación, humedad, fauna, aunque también aspectos de fácil obtención de datos y uso del instrumental.

- Se determinaron los requerimientos del software para la recolección de datos. En base a la experiencia con software de adquisición de datos de sensores se analizarán los requerimientos para la obtención de un software de fácil uso, con funciones de análisis numérico y gráfico.

- Se realizaron la especificación de la red. Para lo cual es necesario especificar con el máximo detalle la red de sensores prototipo que se utilizará.

Para transformar la red SipiA en una aplicación de Internet de las Cosas:

- Se debe definir la tecnología de conectividad a Internet apropiada para la red SIPIA

- Se debe definir el tipo de dispositivo que proporcionará la interfaz entre la WSN y Internet. En primera instancia se propone el uso de un edge router desarrollado sobre una plataforma diferente que haga de pasarela entre la red y la Internet

- Se debe definir la forma en que los datos producidos por los sensores serán visualizados desde Internet. Determinado los protocolos más adecuados para la aplicación de toma de datos de parcelas agrícolas.

#### Avances y resultados preliminares

Se desarrolló un prototipo de red basado en la norma IEEE 802.15.4 [9] para las capas física y enlace. Se prefiere el desarrollo propio para las capas superiores, tomando como prioridad optimizar la red para bajo consumo energético. El objetivo es por lo tanto disminuir el mantenimiento del sistema. Se diseñó e implementó un sistema de gestión que permite prever y determinar el tiempo de agotamiento de baterías, para un mote en particular. Para ello se usan como parámetros el hardware instalado, tipo de baterías y tipo de mote y el protocolo de administración de la energía utilizado.

Se diseñó un sencillo algoritmo de enrutamiento de datos específico para dicho prototipo, buscando maximizar el rendimiento energético del sistema, reduciendo el



overhead. El método antedicho, permite minimizar el procesamiento necesario por parte del mote y reducir el consumo por la transmisión de paquetes pequeños.

Se implementó y ensayó un protocolo de propagación de sincronismo, determinando la mejor relación entre el bajo consumo y el menor error admisible. Se optó por un esquema centralizado, en el cual se resincroniza la red de manera pasiva, habiendo determinado de manera experimental la cantidad mínima de tiempo necesaria entre sincronizaciones sucesivas.

Se implementó un software de gestión en línea. Para el que se estudiaron casos de aplicación y se realizó un relevamiento de requerimientos para el desarrollo de la interfaz humana y de carácter funcional. La versión preliminar prototipo del software para los ensayos en campo tiene las siguientes características:

- Backend desarrollado sobre servidor Debian/Linux, base de datos relacional Mysql y tecnologías Java en el servidor web Apache Tomcat.

- Autenticación segura y conexión cifrada.

- Gestión de la conexión TCP/IP con el Gateway

- Diseño y desarrollo de un protocolo de primitivas para la comunicación con el gateway.

### Formación de Recursos Humanos

Este proyecto de investigación posibilita la colaboración inter-institucional y la ejecución de proyectos conjunto entre grupos I+D de diferentes disciplinas y por lo tanto formar recursos humanos para la proyección académica, científica e industrial de los temas tratados.

En el proyecto participan dos doctorandos con becas ANCYT, un doctorando con becas UTN, dos doctorandos docentes UTN un investigador graduado con becas BINID UTN y seis becarios alumnos con beca UTN

### Referencias

[1] D. Culler and W. Hong. Eds. Special issue on Wireless Sensor Networks. Communications of the ACM, 47(6):pag 30–34, June 2004

[2] Dargie, W. and Poellabauer, C., "Fundamentals of wireless sensor networks: theory and practice", John Wiley and Sons, 2010 ISBN 978-0-470

[3] T. Wark, P. Corke, P. Sikka, L. Klingbeil, Y. Guo, C. Crossman, P. Valencia, D. Swain, , y G. Bishop- Hurley. Transforming agriculture through pervasive wireless sensor networks. Pervasive Computing, pags. 50-57, Abril-Junio 2008

[4] R. Beckwith, D. Teibel, y P. Bowen. Report from the field: results from an agricultural wireless sensor network. In 29th Annual IEEE International Conference on Local Computer Networks, pags. 471-478, November 2004.

[5] W. Zhang, G. Kantor, y S. Singh. Integrated wireless sensor/actuator networks in an agricultural application. In Proceedings of the 2nd International Conference On Embedded Networked Sensor Systems, pag. 317, 2004.

[6] N. Wang, N. Zhang, y M. Wang. Wireless sensors in agriculture and food industry recent development and future perspective. Computers and Electronics in Agriculture, 50(1):1{14, Enero 2006.

[7] Kevin Ashton, "That 'Internet of Things' Thing", RFID Journal, 22 July 2009. Retrieved 8 April 2011

[8] Shelby and Bormann, "6LoWPAN: The Embedded Internet", Wiley, 2009

[9] IEEE 802.15 Working Group for WPAN. <http://www.ieee802.org/15/>

[10] Geoff Mulligan, "The 6LoWPAN architecture", EmNets '07 Proceedings of the 4th workshop on Embedded networked sensors, ACM New York, 2007

[11] G. Montenegro, J. Hui, D. Culler, "Transmission of IPv6 Packets over IEEE 802.15.4 Networks", RFC 4944, IETF, September 2007

[12] N. Kushalnagar, G. Montenegro, C. Schumacher, "IPv6 over Low-Power Wireless Personal Area Networks (6LoWPANs): Overview, Assumptions, Problem Statement, and Goals", RFC 4919, IETF, August 2007

[13] Aguirre M, Egea M, Godoy P, Martínez J, Generale S, Mercado G y Robles J, "Nota de Aplicación NA036: Alarma Anti – Robo inalámbrica ZIG BEE con transceiver MC 13192 contenido en módulo de radio (ANT – ZigBee)", Electrocomponentes SRL, Dic 2008.

[14] G. Mercado, A. Diedrichs y M. Aguirre "The Wireless Embedded Internet", Annals of CASE 2011, ISBN 978-987-9374-69-6, Buenos Aires Marzo 2011

[15] G. Mercado, R. Borgo, F. Gonzalez Antivilo, G. Ortiz Uriburu, A. Diedrichs, P. Farreras, M. Aguirre, F. Battaglia, G. Tabacchi, S. Tromer, "SIPIA Net: Wireless Sensor Network for Agronomical Research", Anales del WICC 2011. XIII Workshop de Investigadores en Ciencias de la Computación, ISBN 978-950-673-892-1, UNR Editora. Editorial de la Universidad Nacional de Rosario, 2011

[16] G. Mercado, R. Borgo, F. González Antivilo, G. Ortiz Uriburu, A. Diedrichs, P. Farreras, M. Aguirre, F. Battaglia, G. Tabacchi, S. Tromer, "Uso de Sensores Inalámbricos como Herramienta para Investigación Agronómica", 1º Simposio Argentino de Viticultura y Enología. Mendoza, Noviembre 2011

[17] G Mercado, R Borgo, F Gonzalez Antivilo, G Ortiz Uriburu, A Diedrichs, S Tromer, N Ledezma, M Aguirre, C Panella, G Tabacchi, J Martí, G Grünwaldt, I Rigoni, G Antón, F Chamorro, R Moreno, S Pérez, "RED DE SENSORES SIPIA Avance del Proyecto, WICC 2012, XIV Workshop de Investigadores en Ciencias de la Computación, Universidad Nacional de Misiones (UNAM) - Posadas - Misiones - Argentina 26 y 27 de Abril de 2012.

[18] J. Martínez, "Leistungsbewertung eines Lokalisierungsalgorithmus in drahtlosen Sensoren", tesina de diplomatura, Technische Universität Dresden - Fakultät Elektrotechnik und Informationstechnik - Institut für Nachrichtentechnik, 2009

[19] G. Bloch, "Implementierung eines Lokalisierungsalgorithmus in drahtlosen Sensoren", tesina de diplomatura, Technische Universität Dresden - Fakultät Elektrotechnik und Informationstechnik - Institut für Nachrichtentechnik, 2009.

[20] Cristian Pérez Monte, María Inés Robles, Gustavo Mercado, Carlos Taffernaberry, Marcela Orbiscay, Sebastián Tobar, Raúl Moralejo y Santiago Pérez,

- “Implementation and Evaluation of Protocols Translating Methods for IPv4 to IPv6 Transition”, *Journal of Computer Science & Technology*, ISSN 1666-6038 Vol. 12 No. 2 August 2012
- [21] Carlos Taffernaberry “6LoWPAN - IPv6 para WSN”, SASE - Simposio Argentino de Sistemas Embebidos 2012, Facultad de Ingeniería de la Universidad de Buenos Aires, Buenos Aires, Agosto 2012.
- [22] Gustavo Mercado, Cristian Pérez Monte, Carlos Taffernaberry, María Inés Robles, Marcela Orbiscay, Sebastián Tobar, Raúl Moralejo, y Santiago Pérez, “Implementación y Evaluación de métodos de Traslación de Protocolos para la transición IPv4-IPv6”. *Anales del CACIC 2011*, ISBN 978-950-34-0756-1, UNLP, La Plata, Octubre 2011.
- [23] C. Taffernaberry, G. Mercado, S. Tobar, C. Pérez Monte, P. Clérigo, I. Robles, M. Orbiscay, S. Pérez, R. Moralejo, “PMIP6: Análisis, Evaluación y Comparación de ambientes Proxy Mobile IP en versión 6, aplicado a Redes de Avanzada”, *Anuales del WICC 2011*. XIII Workshop de Investigadores en Ciencias de la Computación, ISBN 978-950-673-892-1, UNR Editora. Editorial de la Universidad Nacional de Rosario, 2011
- [24] Carlos Tiviroli, Andres Gatti, Gustavo Mercado y Carlos Taffernaberry "QoSIP Meter: Sistema de determinación de condiciones de calidad de servicio en transmisiones de audio/video en tiempo real sobre Internet" *CASE 2012 LIBRO DE TRABAJO* pag. 263-264 ISBN 978-987- 9374-82- Facultad de Ingeniería de la Universidad de Buenos Aires, Buenos Aires, Agosto 2012.
- [25] Robles I, Orbiscay M. "Educatando en IPv6 a través de Conferencias y Talleres Abiertos", *LANIC 18/ LACNOG 2012*, Montevideo, Uruguay, Oct 2012
- [26] G. Mercado y C. Taffernaberry, "6LoWPAN IPv6 for Wireless Sensor Network", *LACNIC XVI LACNOG 2012 ION*, Buenos Aires, Oct 2011
- [27] G. Mercado “6LoWPAN IPv6 for Wireless Sensor Network”, 9º Foro Latinoamericano de IPv6 – FLIP6, *LACNIC XV 2011*, Cancun - Mexico 15 al 20 de mayo de 2011.
- [27] Task Force IPv6 AR [www.ipv6.org.ar](http://www.ipv6.org.ar)
- [28] Cambio climático en la Cordillera de los Andes. Dictado por CCT. Característica: curso de posgrado. Participación: asistente. Ing. Agr. Francisco Gonzalez Antivilo - Ing. Agr. Gisela Ortiz Uriburu
- [29] Workshop Ecofisiología de la Vid. Participación como Organizador. MSc Roberto Borgo
- [30] XI Congreso Latinoamericano de viticultura y enología, Tema: “Partición de fotoasimilados, en vides con diferente potencial vegetativo, con afecciones en la relación fuentes/destinos. Autores: Borgo, R; Gonzalez Antivilo, F; Gamboa, D; Cavagnaro, B; Perez Peña, J; Solanes, Lugar: Mendoza, Argentina.
- [31] SeCyT(U.N.Cuyo2007-2009) Tema: Aspectos fisiológicos y manejo de viñedos afectados por granizo. (Roberto Borgo:Co-Director).
- [32] SeCyT (ANPCyT) 2007 “Estudio exploratorio sobre el impacto del Cambio Climático en cultivos agrícolas de la Región Cuyo” del Programa de Recursos naturales, medio ambiente y Prevención de catástrofes (participante por INTA).
- [33] SeCTyP (UNCuyo). 2009/11. Efecto de altas temperaturas, estrés hídrico y ácido abscísico sobre el crecimiento, producción, composición de bayas y de los vinos obtenidos en *Vitis vinifera L. cv Malbec*
- [34] Proyecto “Estudio del efecto de las altas temperaturas sobre la fisiología y la calidad de uva de *Vitis vinifera var Malbec*”. Doctorando: Ing. Agr. Ortiz Uriburu.
- [35] Proyecto “Mecanismo de resistencia a temperaturas subcero en tejidos leñosos de *Vitis vinifera cv Malbec*”. Doctorando: Ing. Agr. González Antivilo.
- [36] Ajuste de un método colorimétrico para la determinación de almidón en tejidos leñosos (en evaluación *Revista Enología*).
- [37] Galñmarini, C.R.; R. Borgo; J. Gaviola y R.M. Tizio. 1988. "Efecto del ácido giberélico (AG3) sobre la producción de semillas de zanahoria (*Daucus carota L.*) cv. Flackkeé. I- Efecto de diferentes concentraciones y épocas de aplicación sobre la duración del ciclo vegetativo, rendimiento y calidad de semilla.
- [38] BORG, R.; O.M.S. de FERNANDEZ; J.B. CAVAGNARO y G. ESTEVEZ 1992. "DESCANUTADO: Comentarios sobre posibles causas de la disparidad de resultados experimentales". *INFORAJO Suplemento 1 setiembre-diciembre 1992*.
- [39] MAKUCH, M.A. y R. BORG. 1994 "Influencia del estado de madurez y almacenamiento de frutos de *Cucurbita moschata (Dutch.) cv Paquito-INTA*, sobre la calidad de la semilla". *Revista RIA, Vol 25 (1): 97-106*.
- [40] STAHLSHMIDT, O.M.; J.B. CAVAGNARO y R. BORG. 1994. "Influence of planting date and seed cloves size on leaf area and yield of two garlic cultivars (*Allium sativum L.*)". In *Proceedings of the International Symposium on Edible Alliaceae ACTA HORTICULTURAE Nro 433 519-527. 1998*
- [41] STAHLSCHMID, O.M.; J.B.CAVAGNARO y R. BORG" Growth- analysis of three garlic (*Allium sativum L.*) cultivars with differences in yield". Agosto 1994. In *Proceedings of the International Symposium on Edible Alliaceae ACTA HORTICULTURAE Nro 433 427-435. 1998*
- [42] VALENTINE AJ, MORTIMER PE, LINTNAAR M, BORG R "Photosynthetic and nutritional responses of drought stressed arbuscular mycorrhizal grapevines", *SYMBIOSIS JOURNAL* Vol. 41, No. 3 (2006), pp. 127-134



## Desarrollo de Aplicaciones nativas para IPv6

Daniel Giulianelli, Rocío Rodríguez, Pablo Vera, María Antonella Cornejo

GIDFIS (Grupo de Investigación, Desarrollo y Formación en Innovación de Software)

Departamento de Ingeniería e Investigaciones Tecnológicas

Universidad Nacional de La Matanza

Florencio Varela 103, San Justo, Buenos Aires, Argentina

4480-8900 (interno 8751)

dgiulian@ing.unlam.edu.ar; rrodriguez@ing.unlam.edu.ar;

pvera@ing.unlam.edu.ar; mcornejo@ing.unlam.edu.ar

### Resumen

Durante bastante tiempo la preocupación fue proveer a la Universidad Nacional de La Matanza de conectividad IPv6, para lo cual el enfoque estaba puesto netamente en cuestiones de Hardware. Contando con dicha conexión, luego el eje central se transformó en implementar software que aprovechara verdaderamente este protocolo, el cual generalmente atiende cuestiones físicas (manejo de red, control de paquetes, etc.). En esta línea de investigación se pretende desarrollar mediante frameworks particulares, aplicaciones más orientadas a servicios, que aprovechen las características intrínsecas que brinda el protocolo (por ejemplo: calidad de servicio, posibilidad de multicasting, etc).

**Palabras Clave:** desarrollo nativo, aplicaciones, IPV6, redes,

### 1. Contexto

En la Universidad Nacional de La Matanza actualmente hay tres líneas de I+D (Investigación y Desarrollo) en esta temática, las dos primeras orientadas netamente al área de redes y la tercera más orientada a alto nivel,

correspondiente al área de Innovación de Software, cuyos objetivos son:

1. Proveer a más sectores de la universidad conectividad IPv6.
2. Coordinar a alumnos avanzados en la Orientación de Redes para colaborar con el IRTF (Internet Research Task Force) [1] realizando pruebas para futuros protocolos.
3. Formación en desarrollo nativo para IPv6 y construcción de soluciones de software que permitan aprovechar las características del protocolo.

La tercera línea es la que lleva a cabo el GIDFIS (Grupo de Investigación, Desarrollo y Formación en Innovación del Software) con un proyecto que tiene un tiempo previsto de dos años (2013 - 2014) y cuenta con financiamiento mediante el programa PROINCE (PROgrama de INCentivos).

### 2. Introducción

En el 2007 el LACNIC (Latin American and Caribbean Internet Addresses Registry) [2] mencionaba al “2011 como el año en el que se recomienda a todos los proveedores de Internet de la región a tener bloques de

direcciones IPv6, ya en uso, para servicios de producción” [3].

Sin embargo en la actualidad aún muchos usuarios y proveedores de servicios trabajan bajo IPv4, otros están implementando mecanismos de transición: Doble Pila, Entubamiento (Tunneling) ó Traducción de Encabezados. Para poder trabajar internamente con IPv4 y salir al exterior con una IPv6.

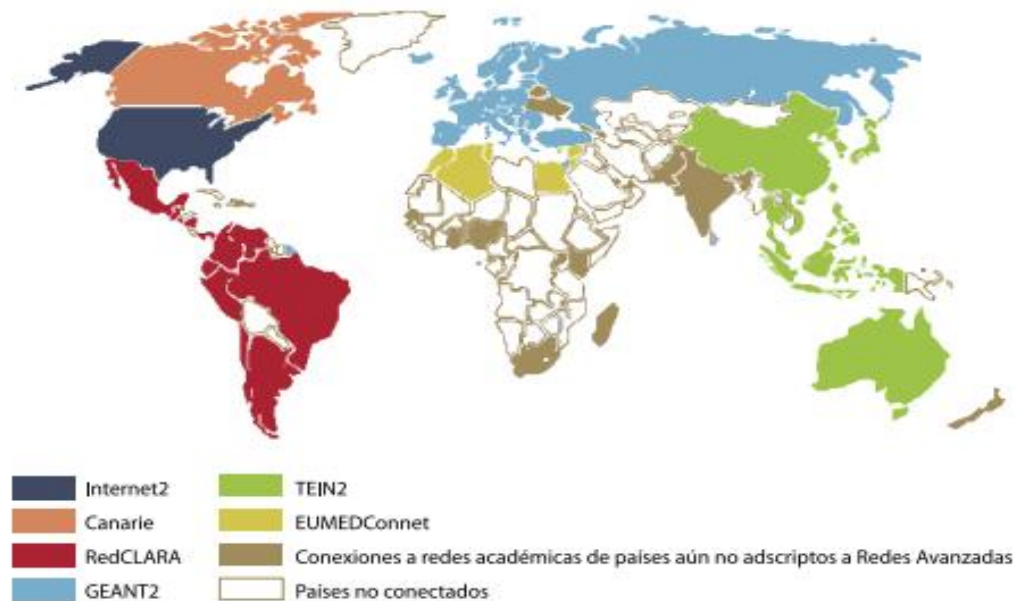
Poco a poco IPv6 irá reemplazando la conectividad del protocolo IPv4, la dificultad básica reside en poder actualizar todo el hardware que sólo puede funcionar con el protocolo anterior. Esta actualización lógicamente se va produciendo en forma paulatina (tomándose mayor tiempo que el planificado, lo que provoca que ya haya áreas en las que se han agotado las direcciones IPv4 asignadas [3]).

“Actualmente los sectores científicos y de educación poseen redes físicas que los vinculan, la mayoría de

ellas con características avanzadas... e IPv6 en forma nativa” [4]. En el mapa de la figura 1 (publicado por el Internet Society) puede observarse el cubrimiento de redes, la mayoría de ellas con soporte a IPv6 desde hace años.

Cuando efectivamente IPv6 sea el protocolo de uso tradicional, surgirá el siguiente interrogante: ¿cómo desde las aplicaciones se pueden aprovechar las ventajas que este protocolo ofrece?. A continuación se enumeran las principales ventajas de IPv6:

1. Mayor cantidad de direcciones: “El protocolo IPv4, dispone solo de 32 bits de direcciones proporcionando un espacio teórico de  $2^{32}$  (aproximadamente cuatro mil millones) interfaces de red únicas globalmente direccionables. IPv6 en cambio tiene direcciones de 128 bits y por tanto puede direccionar  $2^{128}$  interfaces de red (340.282.366.920.938.463.374.607.431.768.211.456).” [5]



**Figura 1.** Redes Académicas en el Mundo

2. Multicast: IPv6 no emplea la técnica de difusión (broadcast) para comunicar con todos los nodos en un segmento de red. Esa funcionalidad se realiza empleando multidifusión (multicast) [6].

La multidifusión IP, consiste en enviar datagramas IP a un grupo de receptores interesados, previa suscripción a un grupo multicast, empleando una única transmisión (tal como se ejemplifica en la figura siguiente). Puede funcionar en comunicaciones de los tipos uno-variados o varios-variados.

3. Seguridad Incorporada: IPv6 incluye de forma nativa IPSEC, que permite la autenticación, encriptación y compresión del tráfico IP. Esto permite utilizar seguridad sin tener que configurar algo particular en cada aplicación[7].
4. Aplicaciones: IPv6 permite el uso de jumbogramas, paquetes de datos de mayor tamaño (hasta 64 bits). Para dar mejor soporte a tráfico en tiempo real (ej. videoconferencia), IPv6 incluye etiquetado de flujos en sus especificaciones. Con este mecanismo los encaminadores o routers pueden reconocer a qué flujo extremo a extremo pertenecen los paquetes que se transmiten.
5. Plug and Play: IPv6 incluye en su estándar el mecanismo "plug and play", lo cual facilita a los usuarios la conexión de sus equipos a la red. La configuración se realiza automáticamente. Esto permite que al conectar una máquina a una red IPv6, se le asigne automáticamente una (o varias) direcciones IPv6.
6. Movilidad: IPv6 incluye mecanismos de movilidad más eficientes y

robustos lo cual beneficiará no sólo a los usuarios de telefonía y dispositivos móviles, sino que también permitirá (por ejemplo) tener buenas conexiones a internet durante los vuelos de avión.

7. Extensibilidad: IPv6 ha sido diseñado para ser extensible y ofrece soporte optimizado para nuevas opciones y agregados, permitiendo introducir mejoras en el futuro.
8. VOIP: Dos de los problemas actuales de los servicios de Voz sobre IP (VoIP) son QoS y NAT (Network Address Translation).

Las comunicaciones pueden resultar en baja calidad de voz (QoS), y presentar dificultad para atravesar firewalls (NAT).

Al incorporar IPv6 una gran cantidad de direcciones, no será necesario utilizar NAT, y sus nuevas capacidades de Plug and Play, seguridad, y QoS implicarán mejores conexiones de voz.

LACNIC (Latin American and Caribbean Internet Addresses Registry) indica que "la adopción temprana de IPv6 por la comunidad académica ha tenido como fin, por un lado la experimentación e investigación y por otro la formación de recursos humanos en el tema. A su vez, algunas necesidades propias de este sector se ven beneficiadas con características disponibles en este protocolo" [8]. Este organismo señala algunos ejemplos a nivel aplicación:

- La necesidad de contar con direcciones públicamente alcanzables, que permitan la interacción entre pares (en aplicaciones "peer to peer" como videoconferencia, operación remota de instrumentos, grids, etc.)

- Características como multicast, necesario en aplicaciones como access grid y otras que requieren optimizar el uso del ancho de banda.
- Disponibilidad de IPSec como parte del stack, lo que facilita el despliegue de aplicaciones que requieren seguridad de extremo a extremo, como disponibilidad de recursos en malla (grids).
- Las nuevas posibilidades que brindan las características de QoS incorporadas al protocolo.

Las nuevas características del protocolo permiten generar aplicaciones que puedan incorporarlas en forma nativa, permitiendo una importante innovación a nivel de software.

### 3. Línea de Investigación y Desarrollo

A partir de la presente línea de investigación y desarrollo tiene las siguientes tareas asociadas:

- Análisis exhaustivo de las características de IPv6
- Implementación de Aplicaciones ya existentes
- Aprendizaje de desarrollo en framework para generar aplicaciones nativas.
- Implementación de Aplicaciones desarrolladas por el grupo de I+D.

### 4. Resultados y Objetivos

Esta línea de investigación ha comenzado en el presente año, con lo cual los resultados son proyecciones esperadas en base a los objetivos planteados. Tal como se planteó en este artículo: Cuando

efectivamente IPv6 sea el protocolo de uso tradicional, surgirá un siguiente interrogante que es: ¿cómo desde las aplicaciones se pueden aprovechar las ventajas que este protocolo ofrece?

La importancia del proyecto reside en el conocimiento que se profundizará sobre IPv6 con un alto interés en la transferencia del mismo tanto en el área docente como con otros equipos externos de investigación. En particular para la Universidad Nacional de La Matanza, será posible a través del presente proyecto sentar bases de aplicaciones gratuitas para ser utilizadas en IPv6. Se espera poder transferir el conocimiento tanto internamente a docentes y alumnos; como a pares fuera del ámbito de la UNLaM.

### 5. Formación de Recursos Humanos

El grupo de investigación se encuentra conformado por 10 Docentes (4 de ellos graduados de la propia universidad) y 2 Alumnos de la Universidad que se encuentran formándose en actividades de I+ D. Se prevé incorporar en el segundo año más alumnos al proyecto.

Actualmente se cuenta con 2 tesis una de grado y otra de postgrado en elaboración dentro de la temática de la presente línea.

### 6. Referencias

- [1] Internet Research Task Force  
<http://irtf.org/>
- [2] LANIC (Latin American and Caribbean Internet Addresses Registry)  
<http://www.lacnic.net/es/web/lacnic/inicio>



- [3] AZAEL Fernández Alcántara, “Direcciones IPv4 ¿recurso de Internet en Agotamiento?”. Universidad Nacional Autónoma de México. (2007)  
<http://www.enterate.unam.mx/Articulos/2007/junio/art1.html>
- [4] O’ FLAHERTY y otros. Internet Society. Capítulo Argentina. “IPv6 para Todos. Guía de uso y aplicación para diversos entornos”. ISBN 978-987-25392-1-4, Pág. 109 a 118 (2009).  
<http://www.ipv6tf.org/pdf/ipv6paratodos.pdf>
- [5] PALET Jordi. “¿Qué es IPv6? – Portal de Transición a IPv6 de América Latina y el Caribe.  
<http://portalipv6.lacnic.net/es/ipv6/novedades/qu-es-ipv6>
- [6] “Introducción a IPv6: Multicast en IPv6”. Portal IPv6 Cuba (2003).  
<http://www.cu.ipv6tf.org/multicastipv6.htm>
- [7] FEYRER Hubert, O’ Reilly. “The future of the Internet” (2001).  
[http://onlamp.com/pub/a/onlamp/2001/05/24/ipv6\\_tutorial.html](http://onlamp.com/pub/a/onlamp/2001/05/24/ipv6_tutorial.html)
- [8] CICILEO Guillermo, “IPv6 en el Ambiente Académico”. Portal de Transición a IPv6 de América Latina y el Caribe.  
<http://portalipv6.lacnic.net/es/ipv6/ipv6-en/ambiente-acad-mico-0>

# Administración de QoS en ambientes de redes de servicios convergentes

María Murazzo<sup>1#</sup>, Nelson Rodríguez<sup>2#</sup>, Ricardo Vergara<sup>3\*</sup>, Franco Carrizo<sup>4\*\*</sup>, Facundo Gonzalez<sup>5\*\*</sup>, Enzo Grosso<sup>6\*\*</sup>

<sup>#</sup> Docentes e Investigadores - Departamento de Informática – Instituto de Informática

<sup>\*</sup> Egresado Licenciatura Sistemas de Información - Departamento de Informática

<sup>\*\*</sup> Alumnos Avanzados Licenciatura en Ciencias de la Computación - Departamento de Informática  
Facultad de Ciencias Exactas, Físicas y Naturales, Universidad Nacional de San Juan, San Juan, Argentina

<sup>1</sup> [marite@unsj-cuim.edu.ar](mailto:marite@unsj-cuim.edu.ar)

<sup>2</sup> [nelson@iinfo.unsj.edu.ar](mailto:nelson@iinfo.unsj.edu.ar)

<sup>3</sup> [ricardo.ed.vergara@gmail.com](mailto:ricardo.ed.vergara@gmail.com)

<sup>4</sup> [francosebastian.23@gmail.com](mailto:francosebastian.23@gmail.com)

<sup>5</sup> [facu\\_jgg@hotmail.com](mailto:facu_jgg@hotmail.com)

<sup>6</sup> [enzodanielgrosso@hotmail.com](mailto:enzodanielgrosso@hotmail.com)

## CONTEXTO

El presente trabajo se encuadra dentro del área de I/D *Arquitectura, Redes y Sistemas Operativos* y se enmarca dentro del proyecto de investigación “*Implantación de un ambiente de cloud computing para integración de recursos*”, código 21E877.

## RESUMEN

La calidad de servicio (QoS) se ha convertido en un factor muy importante dentro de los servicios convergentes actuales. De esta manera es posible trabajar sobre una plataforma de red cableada, inalámbrica o móvil, dándole soporte de funcionamiento adecuado a servicios emergentes, como voz, video o datos.

En este tipo de ambientes donde convergen diferentes tipos de redes y de aplicaciones es necesario que se establezcan mecanismos que permitan proveer de niveles de QoS adecuados para las aplicaciones que lo necesiten.

Estos mecanismos de QoS se pueden implementar mediante la inclusión de extensiones de QoS a los protocolos existentes o mediante una gestión de tráfico que permita priorizar el tráfico proveniente de aplicaciones sensitivas.

El presente trabajo pretende evaluar el impacto de estas dos opciones en ambientes de red cableadas e inalámbricas donde coexisten aplicaciones con y sin restricciones de QoS.

**Palabras clave:** QoS, ancho de banda, retardo, NS 2, MANET

## 1. INTRODUCCION

Las actuales redes de telecomunicación se caracterizan por un constante incremento del número, complejidad y heterogeneidad de los recursos que las componen.

El tráfico multimedia, como el utilizado en telefonía IP o videoconferencia, puede ser extremadamente sensible a los retardos y puede crear demandas de QoS únicas sobre las redes que los transportan.

Cuando los paquetes son entregados usando el modelo de mejor esfuerzo, estos no arriban en orden, en una manera oportuna. El resultado son imágenes no claras, desiguales y movimientos lentos, y el sonido no se lo obtiene sincronizado con la imagen.

Los aspectos críticos que causan la mayor parte de problemas en aplicaciones con grandes restricciones de recursos son: *falta de ancho de banda, retardo extremo a extremo y pérdida de paquetes*.

Con respecto a la *falta de ancho de banda*, la mejor opción para contrarrestar este problema, es clasificar el tráfico dentro de clases de QoS y priorizar tráfico de acuerdo a la importancia del mismo.

El *retardo extremo a extremo*, es el tiempo tomado por un paquete en alcanzar el punto final de recepción después de ser transmitido desde un punto de envío. Una forma de disminuir este retardo es dándole a los paquetes pertenecientes a aplicaciones sensitivas, cierta prioridad para que en el camino extremo a extremo sean tratado de manera más ágil.

Por último, y en relación a la *perdida de paquetes*, los paquetes pueden ser descartados cuando un enlace está congestionado. Esta problemática se puede paliar mediante un esquema de scheduler de tráfico que permita proporcionar un mejor servicio a paquetes pertenecientes a aplicaciones sensibles [1].

La QoS, es un término usado para definir la capacidad de una red para proveer diferentes niveles de servicio a los distintos tipos de tráfico. Permite que los administradores de una red puedan asignarle a un determinado tráfico prioridad sobre otro y, de esta forma, garantizar que un mínimo nivel de servicio le será provisto.

Debido al desarrollo de estos nuevos tipos de aplicaciones (streaming, Voz sobre IP, videoconferencia, etc.), la necesidad de implementar técnicas de calidad de servicio se ha vuelto más evidente.

Aplicando técnicas de Calidad de Servicio se puede proveer un servicio más acorde al tipo de tráfico y de esta manera permitir:

- Dar prioridad a ciertas aplicaciones de nivel crítico en la red.
- Maximizar el uso de la infraestructura de la red.

- Proveer una mejor performance a aplicaciones sensitivas al delay como son las de voz y video.
- Responder a cambios en los flujos del tráfico de red.

Al aplicar técnicas de QoS, el administrador de la red puede tener control sobre los diferentes parámetros que definen las características de un tráfico en particular (retardo extremo a extremo, latencia, variaciones de latencia, perdida de paquetes, ancho de banda).

El problema de la administración de QoS, está prácticamente resuelto en redes fijas, pero esto no sucede en redes inalámbricas y específicamente en redes MANET (Mobile Ad-hoc NETwork) cuyas características de las hacen necesario un nuevo estudio para afrontar este problema.

La topología dinámica, la naturaleza multihop y los escasos recursos de los nodos hacen necesario que los mecanismos de provisión de QoS sean lo más ligeros posibles, en cuanto a carga de procesamiento, como de recursos de red (ancho de banda), para evitar que el throughput o capacidad disponible por nodo se reduzca drásticamente [2].

De todo lo enunciado se puede llegar a la conclusión que la única manera de poder lograr la coexistencia de aplicaciones con diferentes niveles de requerimientos de recursos es realizando una adecuada administración del tráfico mediante la priorización implícita o explícita de los paquetes de datos.

Esta priorización permitirá que ciertos flujos de datos puedan ser tratados de forma preferencial logrando maximizar el uso del ancho de banda, minimizar el retardo extremo a extremo y minimizar la perdida de paquetes.

Esta priorización se logra mediante la implementación de mecanismos de QoS que permita una gestión de los flujos de tráfico [3].

## 2. LINEAS DE INVESTIGACION y DESARROLLO

El presente trabajo posee dos líneas de investigación.

La primera línea de investigación, tiene como objetivo principal el análisis del comportamiento de las redes MANET frente a necesidades de requerimiento de QoS. Para ello, se propone trabajar en el mejoramiento de las prestaciones de los protocolos de ruteo reactivos (DSR y AODV) para MANET mediante el uso del estándar 802.11e.

La segunda línea de investigación, tiene como objetivo principal la administración y monitoreo de ancho de banda en redes cableadas e inalámbricas. Esto permitirá dividir y administrar el tráfico de datos, permitiendo la priorización del tráfico que requiera niveles estrictos de recursos.

## 3. RESULTADOS OBTENIDOS/ESPERADOS

De las dos líneas de investigación propuestas solo de la primera ya existen resultados debido a que se está trabajando desde 2009.

Con respecto a la segunda línea de investigación recién este año se está empezando con las tareas de investigación.

### 3.1- Resultados Obtenidos:

Para la primera línea de investigación, se realizaron estudios comparativos de la performance de los protocolos de ruteo para MANET más utilizados. Los protocolos seleccionados fueron DSR y AODV.

Para las pruebas se utilizó el simulador NS-2. Se trabajó con escenarios de trabajos para simular el comportamiento de los protocolos de ruteo reactivos mediante la priorización basada en 802.11e.

Se trabajó con un campo de simulación de 500 X 500 metros, con 20, 30, 50 y 100 nodos durante un periodo de simulación de 2000 seg. Se trabajó con dos modelos de

movilidad: Random Waypoint Mobility Model (RWPM), y Random Walk Mobility Model (RWKM). Se analizaron el retardo extremo a extremo y la sobrecarga de la red.

Para el caso del análisis del retardo extremo a extremo, las figuras 1 y 2 muestran el comportamiento del protocolo AODV con y sin QoS para los dos modelos de movilidad propuestos en el escenario de 100 nodos.

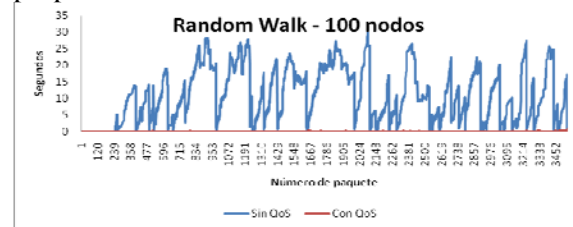


Figura 1: retardo extremo a extremo para RWPM

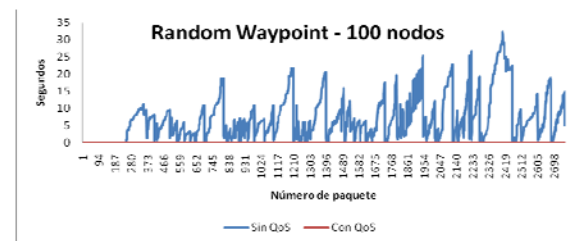


Figura 2: retardo extremo a extremo para RWKM

Considerando el parámetro de la sobrecarga generada por el protocolo de ruteo, los resultados se muestran en las figuras 3 y 4, para el escenario de 100 nodos.

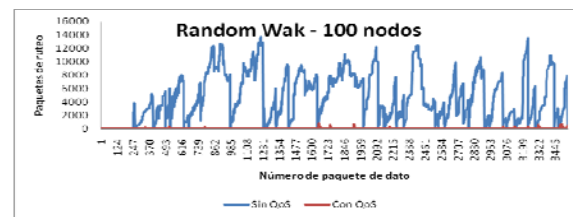


Figura 1: Sobrecarga con RWKM

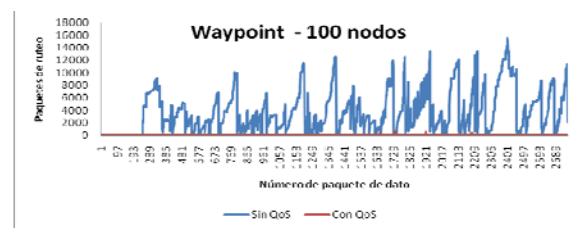


Figura 4: Sobrecarga con RWPM

En líneas generales se puede concluir que la implementación del estándar 802.11e



produce mejores resultados en las redes que lo implementan.

Cualquiera sea el tipo de movilidad, y la cantidad de nodos, el retardo es siempre más bajo y más estable cuando el tráfico pertenece a una red con 802.11e. Aún cuando la granularidad es alta el retardo promedio en las redes con QoS es menor a 1 segundo.

Siempre que la granularidad de la red sea menor a 100 nodos, el retardo promedio será muy bajo, no importa qué tipo de movimiento tienen los nodos.

En cuanto a la sobrecarga, el estándar 802.11e marca una gran diferencia en la sobrecarga entre las redes que lo implementan de aquellas que no lo hacen. Mientras que la granularidad sea baja la sobrecarga se mantiene estable y no alcanza grandes valores cuando se implementa el estándar de QoS a nivel de MAC.

Cuando la granularidad de la red es baja, menor a 50 nodos, y se implementa QoS, la sobrecarga de la red es estable y se mantiene en bajos niveles. A partir de los 50 nodos, y si las aplicaciones son del tipo CBR, la sobrecarga se vuelve inestable y alta.

Esto permite concluir que en el caso de redes con alta granularidad, las aplicaciones sensibles al ancho de banda se verán seriamente afectadas, debido a que la sobrecarga no sólo es alta sino que además es inestable.

Las redes con alta granularidad, sin importar el tipo de movimiento que tengan los nodos, no son aptas para aplicaciones sensibles al ancho de banda, debido a que existe una gran sobrecarga de ruteo, y esta es muy inestable.

De lo anterior se ha podido concluir que los entornos móviles con alta granularidad no son aptos para aplicaciones como video llamadas, juegos en línea, streaming de video

HD, etc., que son, como se mencionó, sensibles al ancho de banda.

El estándar 802.11e funciona bien cuando la granularidad es baja, no mayor a 50 nodos y sólo cuando las aplicaciones que compiten por los recursos de la red no requieren confirmación de recepción y reenvío como lo son aquellas que funcionan bajo protocolos como TCP.

### 3.2- Resultados Esperados:

Para la primera línea de investigación, se propone;

- Implementación de QoS en niveles superiores al de la capa MAC. Como por ejemplo en la capa de red.
- Ampliar las simulaciones usando otros protocolos reactivos como DSR.
- Probar la QoS cuando existen celdas con distintos protocolos ruteo, o redes con distintos requerimientos de seguridad.
- Estudio e implementación de distintos modelos de movilidad es un gran desafío hoy en día, debido a que se hace difícil encontrar generadores de movimiento bajo distintos modelos de movilidad.

Para la segunda línea de investigación, se plantea trabajar en el Laboratorio de redes de la FCEFyN de la UNSJ, instalando Mikrotik para realizar administración de ancho de banda y la herramienta Mangle para la priorización de tráfico.

## 4. FORMACION DE RECURSOS HUMANOS

Dentro de los temas involucrados en esta línea de investigación se están desarrollando actualmente dos tesis de maestría, una de las cuales corresponde a la Maestría en Redes de datos de la UNLP y tres tesinas de grado de licenciatura.

También participan en el desarrollo de las investigaciones tres alumnos avanzados de la licenciatura en calidad de becarios.

## 5. BIBLIOGRAFIA

- [1] Kim, Anbin. *QoS support for advanced multimedia systems*. Information Networking (ICOIN), 2012 International Conference. Page(s): 453 - 456
- [2] Marrone, Robles, Murazzo, Rodríguez, Vergara. *Administración de QoS en MANET*. WICC 2011.
- [3] Robert Wójcik. *Flow Oriented Approaches to QoS Assurance*. Journal ACM Computing Surveys (CSUR). Volume 44 Issue 1, January 2012 . Article No. 5.

## Sistemas Embebidos en Red con Requerimientos de Tiempo Real

Guillermo Friedrich, Guillermo Reggiani, Ricardo Cayssials, Christian Galasso, Sergio Pellegrino,  
Lorena Cofre y Gabriela Velasquez

*Grupo de I+D en Sistemas y Tecnologías de la Información y las Comunicaciones (Grupo SiTIC)  
Universidad Tecnológica Nacional – Facultad Regional Bahía Blanca –  
Departamento de Ingeniería Electrónica  
11 de Abril 461  
(8000) Bahía Blanca*

*{gfried, ghreggiani, rcayssials, spellgrino}@frbb.utn.edu.ar,  
christian\_galasso81@yahoo.com.ar, {stonietti, cofrelorena}@gmail.com,  
gabrielavelasquez2112@hotmail.com*

### Resumen

*Los sistemas embebidos están presentes en la mayoría de los dispositivos con los que interactuamos en la vida cotidiana: desde teléfonos celulares, computadoras de mano, electrodomésticos, juguetes, hasta aplicaciones aeroespaciales, industriales, instrumental médico, equipamiento del automóvil, etc.*

*Por la naturaleza de las aplicaciones y sus inherentes ventajas, existe un creciente interés en utilizar redes inalámbricas (WLAN/WPAN) para la comunicación en redes de sistemas embebidos. En el marco de este proyecto, y siguiendo la línea de trabajo de los anteriores Proyectos de I+D (PID) homologados: 25/B021 y 25/B024, se propone modelar, ensayar, simular y plantear esquemas en los niveles de Control de Acceso al Medio (MAC) inalámbrico y de enrutamiento, para que en función de las condiciones particulares y restricciones impuestas por el entorno de operación (restricciones temporales, calidad de servicio, robustez, consumo, movilidad, distribución geográfica, etc.) se puedan mejorar determinados aspectos de su desempeño, tales como factor de utilización, retardo, "jitter", pérdida de paquetes, tasa de transferencia, etc. Se propone continuar utilizando simuladores consolidados, como ns-2[1], y otras herramientas de hardware y software que permitan modificar y/o implementar algoritmos sobre sistemas embebidos con el fin de realizar ensayos sobre los medios físicos. Como parte del trabajo se tiene previsto el desarrollo, a nivel de prototipo, de interfaces de red inalámbrica que implementen alguno de los modelos propuestos.*

**Palabras Clave:** Sistemas Embebidos, Redes Inalámbricas, ns-2, MAC inalámbrico.

### Contexto

La línea de investigación presentada se encuentra en ejecución en el marco del proyecto titulado: “*Sistemas embebidos en red con requerimientos de tiempo real. Modelado, simulación e implementación práctica de protocolos de comunicación*”. El mismo es financiado por la Universidad Tecnológica Nacional, está incorporado al Programa de Incentivos y su ámbito de realización es el Grupo SiTIC (Grupo de I+D en Sistemas y Tecnologías de la Información y las Comunicaciones), de la UTN - Facultad Regional Bahía Blanca.

### Introducción

La investigación y desarrollo sobre redes inalámbricas continúa siendo un área muy activa, por la aparición de nuevos escenarios de aplicación, por lo general basados en dispositivos embebidos móviles.

Las redes inalámbricas presentan diversos desafíos: mejorar su desempeño para requerimientos específicos (performance para tráfico multimedia; robustez y tolerancia a fallas para ambientes industriales); reconfiguración de la topología para redes de sensores; privacidad y seguridad para aplicaciones administrativas, financieras y de gestión; bajo consumo; optimizar el tamaño de tramas en función de las condiciones del canal; arquitectura multi-hop en redes ad-hoc para favorecer la movilidad[2].

Se puede ver la constante aparición de nuevos estándares o modificaciones a los existentes, a fin de adaptarse a nuevos requerimientos impuestos a las redes

debido a la evolución tecnológica. Hay cada vez más aplicaciones que requieren del uso de redes, con exigencias y restricciones específicas, que obligan a una evolución permanente de los estándares y tecnologías de red. Cabe mencionar los estándares 802.22 (WRAN: Wireless Regional Area Network), 802.11p (WAVE: Wireless Access Vehicular Environment) entre otros.

Por otra parte, se puede ver la proliferación de trabajos de investigación y desarrollo orientados a optimizar el uso de distintos tipos de redes [3][4] para operar en diferentes contextos, como las aplicaciones que requieren una Calidad de Servicio (QoS) confiable que van desde la voz sobre IP (VoIP) hasta Sistemas de Control basados en Red (NCS)[5]. Uno de los principales requerimientos para tales tipos de aplicaciones es disponer de vínculos de comunicación confiables. Por ejemplo, en sistemas de automatización en tiempo real (RT) los datos de control deben transmitirse periódicamente entre sensores, controladores y actuadores cumpliendo con estrictos tiempos de transferencia.

El estándar IEEE 802.11e intenta ofrecer un nivel diferenciado de calidad de servicio para dar soporte a distintos tipos de aplicaciones [6]. La mayoría de los estudios han evaluado el comportamiento de esta norma teniendo en cuenta las necesidades típicas del tráfico multimedia (voz y video). Sin embargo, cuando los servicios de comunicaciones se utilizan para soportar aplicaciones de RT, se deben considerar requerimientos de comunicación específicos y más estrictos [7][8].

El estándar 802.11p[9] establece un marco de referencia para el desarrollo de redes y aplicaciones de comunicaciones inter-vehicular, en las que el tiempo de establecimiento de conexión entre vehículos, previo a la habilitación de la comunicación, debe ser muy acotado.

La inserción del estándar 802.15.4 (ZigBee) en ambiente industriales "sucios", donde se requiere desempeño de tiempo real de alta calidad, diagramación determinística, confiabilidad y bajo consumo, requiere de cambios en el control de acceso al medio, basado en CSMA/CA, para evitar consecuencias desastrosas en el campo industrial.

## Líneas de Investigación y Desarrollo

El Proyecto propuesto es una continuación del PID 25/B021 "Redes, protocolos y sistemas para la comunicación de datos. Estudio, modelado, simulación y desarrollo para aplicaciones con elevadas exigencias" y del PID 25/B024 "Redes y protocolos para comunicación de datos en aplicaciones de instrumentación y control: Modelado, simulación y desarrollos experimentales".

Las tecnologías inalámbricas son una buena elección para trabajar en ambientes industriales, donde es necesario interconectar sistemas móviles o bien se desea evitar el cableado de sensores y controladores en planta.

Sin embargo estas tecnologías presentan problemas de confiabilidad y temporizado inherentes a las características de los canales de radio, a los mecanismos de acceso al medio, etc. El estándar 802.11e provee dos alternativas de acceso al medio (EDCA y HCCA) con cuatro niveles de Calidad de Servicio (QoS) diferenciados. En el PID 25/B021 se propuso un mecanismo de control de acceso al medio, denominado WRTMAC (Control de Acceso al Medio Inalámbrico para Tiempo Real)[10], desarrollado a partir del esquema EDCA del estándar 802.11e, optimizado empleando clases de prioridades. El manejo de los espacios entre tramas para arbitraje (AIFS) fue modificado a fin de que el tiempo de respuesta de la red sea predecible. Esto provee un mecanismo de control de acceso al medio (MAC) libre de colisiones en redes inalámbricas. En la última etapa del proyecto se realizaron trabajos comparativos de WRTMAC con respecto a EDCA, simulado en ns-2. El PID 25/B024 propone configurar los parámetros de EDCA para eliminar los factores probabilísticos y garantizar una latencia máxima de transmisión, a fin de que sea apto para implementar sistemas de control basado en red (NCS) sobre redes inalámbricas (WNCS). El nuevo protocolo fue denominado RT-EDCA[11][12].

## Resultados y Objetivos

La evaluación del desempeño del modelo RT-EDCA fue realizada por medio de simulaciones. Se utilizaron mensajes periódicos de pequeño tamaño similar a las tramas empleadas en los sistemas de control sobre redes inalámbricas (WNCS).

En la Fig. 1 se presentan resultados comparativos obtenidos en las simulaciones, tanto para RT-EDCA como para EDCA estándar. Se observa que RT-EDCA presenta un mejor desempeño hasta 28 nodos (con paquetes de 50 bytes) y hasta 40 nodos (con paquetes de 500 bytes). A partir de allí es superado por EDCA, debido a que los valores crecientes de AIFS aumentan la duración del ciclo de transmisión de cada trama.

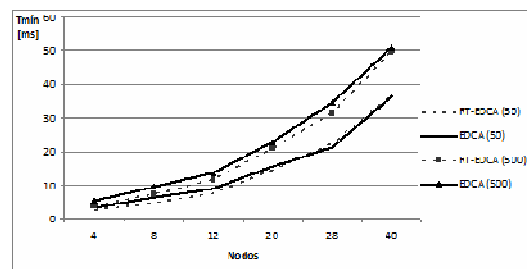


Fig. 1. Período mínimo comparativo entre RT-EDCA vs EDCA

Las características determinísticas de RT-EDCA mejoran notablemente el desempeño respecto a EDCA



(hasta un cierto número de nodos) al eliminar las colisiones.

Producción del Grupo:

- "Supertrama de Duración Variable en MAC 802.15.3 Utilizando MCTA Dinámicos", Reggiani, O. Alimenti, R. Cayssials y E. Ferro, Clei 2009, Pelotas, Brasil, ISBN: 857669247-3, del 22 al 25 de Septiembre de 2009.

- "Improvements in the Superframes Generation in MAC 802.15.3 Using Dynamic MCTA", Reggiani, O. Alimenti, R. Cayssials y E. Ferro, WIP 27th Edition The Symposium on Computer Networks and Distributed Systems (SBRC), Recife, ISBN:978-85-7669-226-3, pp:153-158, May 25 – 29, 2009, Brasil.

- "Un modelo para el análisis de la confiabilidad de Ethernet Industrial en topología de anillo", Guillermo R. Friedrich y Jorge R. Ardenghi, RIAII Revista Iberoamericana de Automática e Informática Industrial, ISSN: 1697-7912 (edición impresa), ISSN 1697-7920 (edición web), vol 6, N° 3, pp. 101-109, julio de 2009.

- "Propuesta en MAC de Redes de Area Personal (WPAN) para Optimizar el Tiempo de Respuesta", G. Reggiani, O. Alimenti y G. Friedrich, XV Congreso Argentino de Ciencias de la Computación - CACIC 2009 - III Workshop de Arquitecturas, Redes y Sistemas Operativos (WARSO) 2009, pp:, ISBN: 978-897-24068-4-1, San Salvador de Jujuy - Jujuy, Argentina, 5 al 9 de octubre de 2009.

- "A Free-Collision MAC Proposal for 802.11 Networks", O. Alimenti, G. Friedrich and G. Reggiani, 28th Edition The Symposium on Computer Networks and Distributed Systems (SBRC 2010) and 12th Brazilian Workshop on Real-Time and Embedded Systems (WTR 2010), Gramado, ISSN: 2177-496X, pp: 89 – 100 (cd), May 24 – 28, 2010, Brasil.

- "Una Variante al Control de Acceso al Medio de 802.11e para Tiempo Real", G. Friedrich, O. Alimenti y G. Reggiani, XVII International Congress Of Electronic, Electrical And Systems Engineering" XVII INTERCON 2010. ISBN: 9 al 13 de agosto de 2010. Puno, Perú.

- "WRTMAC: A MAC Proposal for 802.11 Networks in Factory Automation", G. Friedrich, O. Alimenti and G. Reggiani, 15th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA '2010), ISBN:978-4244-6849-2. IEEE Catalog Number CFP10ETF-CDR. September 13-16, 2010, Bilbao, Spain. Publicado en IEEE Xplore Digital Library: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5641191>.

- "Evaluación de una Variante de Control de Acceso al Medio Inalámbrico para Tiempo Real Basada en 802.11e", G. Friedrich, O. Alimenti, G. Reggiani, F. Maidana, S. Tonietti y D. Gómez De Marco, XVI Congreso Argentino de Ciencias de la Computación -

CACIC 2010 - III Workshop de Arquitecturas, Redes y Sistemas Operativos (WARSO) 2010, Buenos Aires - Argentina, 18 al 22 de octubre de 2010.

- "Una Adaptación del Protocolo EDCA para Tiempo Real", Omar Alimenti, Guillermo Friedrich, Guillermo Reggiani, Santiago Tonietti, Federico Maidana, Damián Gómez De Marco, Brazilian Symposium on Computing System Engineering (SBESC) and XIII Real-time Systems Workshop (WTR 2011), Florianópolis, ISSN: 978-0-7695-4641-4/11, pp: 145-150, Nov 7 – 11, 2011, Brasil. Publicado en IEEE Xplore Digital Library: [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=6114849](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6114849).

- "Redes Inalámbricas y Protocolos para Comunicación de Datos en Aplicaciones de Instrumentación y Control", Omar Alimenti, Guillermo Friedrich, Guillermo Reggiani, Ricardo Cayssials, Christian Galasso, Damián Gomez de Marco, Federico Maidana y Santiago Tonietti, XIII Workshop de Investigadores en Ciencias de la Computación (WICC 2011), Rosario, Santa Fe, los días 5 y 6 de mayo de 2011.

- "Comunicaciones en Tiempo Real adecuando el Protocolo 802.11e (EDCA)", Omar Alimenti, Guillermo Friedrich, Guillermo Reggiani y Santiago Tonietti, 40° Jornadas Argentinas de Informática, 12th AST Argentine Symposium on Technology, ISSN: 1850-2806, pp: 43-54, Córdoba, Argentina, 29/08 al 02/09 de 2011.

- "Análisis de Generación de Supertramas en MAC 802.15.3", G. Reggiani, Lorenzo De Pascuale, O. Alimenti y G. Friedrich, XVII Congreso Argentino de Ciencias de la Computación - CACIC 2011 - VI Workshop de Arquitecturas, Redes y Sistemas Operativos (WARSO) 2011, pp 1066-1075, ISBN 978-950-34-0756-1, La Plata, Argentina, 10 al 14 de octubre de 2011.

- "Adaptación de ns-2 para una Variante del Protocolo 802.11e (EDCA)", Guillermo Reggiani, Omar Alimenti, Guillermo Friedrich, Santiago Tonietti, Federico Maidana, Damian Gomez de Marco, XVII Congreso Argentino de Ciencias de la Computación - CACIC 2011 - VI Workshop de Arquitecturas, Redes y Sistemas Operativos (WARSO) 2011, 1056-1065, ISBN 978-950-34-0756-1, La Plata, Argentina, 10 al 14 de octubre de 2011.

- "Protocolos para Redes Inalámbricas: Simulaciones y modelaje experimental", Omar Alimenti, Guillermo Friedrich, Guillermo Reggiani, Ricardo Cayssials, Christian Galasso, Damián Gomez de Marco, Federico Maidana y Santiago Tonietti, XIV Workshop de Investigadores en Ciencias de la Computación (WICC 2012), Rosario, Santa Fe, los días 26 y 27 de abril de 2012.

- "Análisis del Tiempo de Respuesta en entorno de Tiempo Real sobre el MAC 802.11e", O. Alimenti, G. Friedrich, G. Reggiani, S. Tonietti, G. Velazquez y L. Cofre, 41° Jornadas Argentinas de Informática, 13th

AST 2012 Argentine Symposium on Technology, ISSN 1850-2806 pp 132-142. La Plata, Argentina, 27/08 al 31/08 de 2012.

- "Evaluación de EDCA 802.11e en Tiempo Real con Agrupamiento por Clases de Prioridad", O. Alimenti, G. Reggiani, G. Friedrich, S. Tonietti, G. Velazquez y L. Cofre, XVIII Congreso Argentino de Ciencias de la Computación - CACIC 2012 -VII Workshop de Arquitecturas, Redes y Sistemas Operativos (WARSO) 2012, ISBN 978-987-1648-34-4, Bahía Blanca, Argentina, 8 al 12 de octubre de 2012.

Los objetivos del proyecto son:

Objetivos generales:

- Adaptar estándares y tecnologías de control de acceso al medio inalámbrico existentes, para su aplicación en sistemas de instrumentación y control.

- Desarrollar propuestas de nuevos mecanismos de control de acceso al medio, tratando de aprovechar elementos de las tecnologías existentes.

- Implementar y ensayar uno o más de los mecanismos de control de acceso al medio desarrollados y/o adaptados.

Objetivos específicos:

- Explorar los estándares y tecnologías de control de acceso al medio inalámbrico existentes.

- Identificar fortalezas y debilidades en el contexto del campo de aplicación en el que se lo pretende aplicar.

- Proponer modificaciones a los mecanismos existentes o bien introducir nuevos, tratando de aprovechar y reutilizar elementos de tecnologías existentes.

- Realizar algún desarrollo experimental de hardware y software (firmware) que implemente la capa física y MAC inalámbrica, tanto de algún estándar existente como así también de las variantes propuestas.

- Evaluar los modelos desarrollados mediante simulaciones y ensayos experimentales.

## Resultados esperados

La contribución que se pretende realizar desde el punto de vista científico-tecnológico está dada por el desarrollo de variantes a protocolos de control de acceso al medio inalámbrico y enrutamiento, con el fin de mejorar su desempeño en determinados escenarios de interés práctico como ser redes de sensores, NCS, electrónica vehicular, VoIP, etc. En este proyecto se pretende realizar un importante esfuerzo en desarrollos experimentales y a partir de los resultados que se obtengan es de esperar una posible transferencia, aplicable a sistemas embebidos de instrumentación y control, a solucionar problemas de consumo energético, a electrónica vehicular entre otros.

Otra contribución esperada en lo referente a la transferencia al medio, es la posibilidad de introducir modificaciones sobre equipamiento de red inalámbrica estándar, a fin de mejorar su desempeño.

Asimismo, se espera efectuar alguna transferencia con impacto favorable a nivel regional y/o nacional, proponiendo aplicaciones de esta tecnología para resolver problemas de monitoreo y control ambiental, control de riego, etc. En tal sentido, hay en marcha un convenio de colaboración con el Grupo LIHANDO de la UTN Facultad Regional Mendoza, que investiga sobre la posibilidad de favorecer las precipitaciones pluviales bajo ciertas condiciones atmosféricas. La participación del Grupo SiTIC está dada en el diseño y desarrollo de las estrategias, equipamiento y sistemas para el control y la comunicación de los equipos de campo.

## Formación de Recursos Humanos

En cuanto a la formación de recursos humanos, en el marco del proyecto está prevista la realización de, al menos, tres tesis de la Maestría en Redes de Datos de la Universidad Nacional de La Plata. Las mismas se encuentran actualmente en la etapa de definición y presentación de los planes de tesis.

También están prevista la participación de alumnos avanzados de la carrera de Ingeniería Electrónica, esperándose la realización de, al menos, cuatro proyectos finales de carrera. Estos alumnos ya han venido participando en proyectos anteriores y son coautores de distintos trabajos ya presentados y/o publicados.

## Referencias

- [1] The Network Simulator- ns-2, April 28, 2007. <<http://www.isi.edu/nsnam/ns/index.html>>
- [2] "Delay Analysis of IEEE 802.11 PCF MAC based Wireless Networks", Biplab Sikdar. Global Telecommunications Conference, 2005. GLOBECOM '05. IEEE.
- [3] "Performance Evaluation of IEEE 802.11 Wireless Networks for Real-time Networked Control Systems", Li Gui, Yu-Chu Tian, Colin Fidge. Proceedings The 2007 International Conference on Embedded Systems and Applications, Monte Carlo Resort, Las Vegas, Nevada, USA.
- [4] "Control and Communication Challenges in Networked Real-Time Systems", Baillieul, J.; Antsaklis, P.J. Proceedings of the IEEE Issue Date: Jan. 2007 Volume: 95 Issue: 1, page(s): 9 – 28, ISSN: 0018-9219.
- [5] "A Survey of Recent Results in Networked Control Systems", Hespanha, J.P. Naghshtabrizi, P. Yonggang Xu, page(s): 138 – 162. Proceedings of the IEEE Issue Date: Jan. 2007 Volume: 95, ISSN: 0018-9219.
- [6] "How to use the IEEE 802.11e EDCA Protocol when Dealing with Real-Time Communication", Ricardo Moraes, Francisco Vasques, Paulo Portugal, José Alberto

- Fonseca. 11th Brazilian Workshop on Real-Time and Embedded Systems (WRT 2009).
- [7] "A Survey on Real-Time MAC Protocols in Wireless Sensor Networks", Zheng Teng, Ki-II Kim. Communications and Network, 2010, 2, 104-112.
- [8] "A Deterministic Scheduling Mechanism for Industrial Wireless Networks", TongBo Li, Heng Wang, Ping Wang and Yeon Kim, IC EMI'2009.
- [9] "IEEE 802.11p: Towards an International Standard for Wireless Access in Vehicular Environments", Daniel Jiang, Luca Delgrossi, ISBN: 978-1-4244-1645-5/08/©2008 IEEE.
- [10] "WRTMAC: A MAC Proposal for 802.11 Networks in Factory Automation", G. Friedrich, O. Alimenti y G. Reggiani, 15th IEEE International Conference on Emerging Technologies and Factory Automation – ETFA 2010 (UPV/EHU) ISBN: 978-1-4244-6849-2 – 13 al 16 de Setiembre de 2010 – Bilbao – España.
- [11] "Evaluación de una Variante de Control de Acceso al Medio Inalámbrico para Tiempo Real Basada en 802.11e", O. Alimenti, G. Friedrich, G. Reggiani, F. Maidana, S. Tonietti y D. Gomez de Marco, XVI Congreso Argentino de Ciencias de la Computación - CACIC 2010 – III Workshop de Arquitecturas, Redes y Sistemas Operativos (WARSO) 2010, Buenos Aires, Argentina, 18 al 22 de octubre de 2010.
- [12] Wietholter S., Hoene C., "Design and Verification of an IEEE 802.11e EDCA", Technical Report TKN-03-19, November 2003, Berlin

# Despliegue de MANETs para M-learning en zonas de recursos limitados

Sergio M. Rocabado<sup>1</sup>, Mario C. Montalvetti<sup>2</sup>, Susana I. Herrera<sup>2</sup>, Daniel Arias Figueroa<sup>1</sup>

(1) *Centro de Investigación y Desarrollo en Informática Aplicada (CIDIA),  
Facultad de Ciencias Exactas, Universidad Nacional de Salta*  
[srocabad@cidia.unsa.edu.ar](mailto:srocabad@cidia.unsa.edu.ar), [daaf@cidia.unsa.edu.ar](mailto:daaf@cidia.unsa.edu.ar)

(2) *Instituto de Investigación en Informática y Sistemas de Información,  
Facultad de Ciencias Exactas y Tecnologías, Universidad Nacional de Santiago del Estero*  
[mmontalv@unse.edu.ar](mailto:mmontalv@unse.edu.ar), [sherrera@unse.edu.ar](mailto:sherrera@unse.edu.ar)

## Resumen

Las zonas rurales de recursos limitados del país se caracterizan, entre otros aspectos, por su baja densidad demográfica, cobertura de red celular muy limitada y carencia de servicio de distribución de energía eléctrica. Los habitantes de estas zonas utilizan energías alternativas, como paneles solares y grupos electrógenos, para cubrir necesidades energéticas elementales. La región Noroeste de Argentina (NOA) posee numerosas zonas de este tipo, donde los pobladores son personas de bajos recursos y tienen pocas posibilidades de educación en su entorno; se garantiza la educación primaria pero son pocas las escuelas secundarias. En este contexto, el aprendizaje mediado por tecnologías es prácticamente nulo.

En este trabajo se propone realizar una investigación sobre el despliegue de MANETs, seguras y de bajo consumo, que permitan implementar estrategias de m-learning en estas zonas.

Esta propuesta se basa en la figura de un profesor itinerante que imparte educación secundaria en zonas rurales utilizando objetos de aprendizaje almacenados en un servidor de recursos m-learning. Los objetos son accedidos desde teléfonos celulares sencillos que se auto organizan entre sí, formando una MANET, para optimizar el uso de recursos.

**Palabras clave:** MANET, m-Learning, Bluetooth, GPRS, seguridad, redes.

## 1 Contexto

El presente trabajo se lleva a cabo en el marco de un proyecto de investigación que tiene como objetivo realizar investigación aplicada sobre nuevas tecnologías que permitan mejorar la calidad de las aplicaciones móviles.

El equipo de investigación está conformado por docentes de diferentes universidades nacionales del NOA:

Instituto de Investigación en Informática y Sistemas de Información de la Facultad de Ciencias Exactas de la Universidad Nacional de Santiago del Estero (UNSE), Facultad de Humanidades de la Universidad Nacional de Santiago del Estero (UNSE), Departamento de Sistemas de la Facultad de Tecnologías y Ciencias Aplicadas de la Universidad Nacional de Catamarca (UNCA) y Departamento de Informática de la Facultad de Ciencias Exactas de la Universidad Nacional de Salta (UNSA).

Además, el proyecto cuenta con el asesoramiento de dos investigadoras de Institutos de Investigación de la Facultad de Informática de la Universidad Nacional de La Plata (UNLP). La asesora en Sistemas Móviles pertenece al Laboratorio de Investigación y Formación en Informática Avanzada (LIFIA) de la UNLP, mientras que la asesora en educación a distancia pertenece al Instituto de Investigaciones en Informática (LIDI) de la UNLP.

Está financiado por el Consejo de Ciencia y Técnica de la Universidad Nacional de Santiago del Estero, durante el período Enero 2012-Diciembre 2015.

## 2 Introducción

Una red móvil ad-hoc o MANET (del inglés Mobile Ad-hoc Networks) [1] es una colección de nodos inalámbricos móviles que se comunican de manera espontánea y auto organizada constituyendo una red temporal sin la ayuda de ninguna infraestructura preestablecida (como puntos de acceso WiFi o torres de estaciones base celulares con antenas 2G, 3G o 4G) ni administración centralizada.

Una de las principales ventajas de una MANET es la posibilidad de integrarla a una red de infraestructura con diferentes fines, entre otros podemos mencionar el acceso a aplicaciones y recursos M-learning de una organización desde un dispositivo móvil [2] [5].

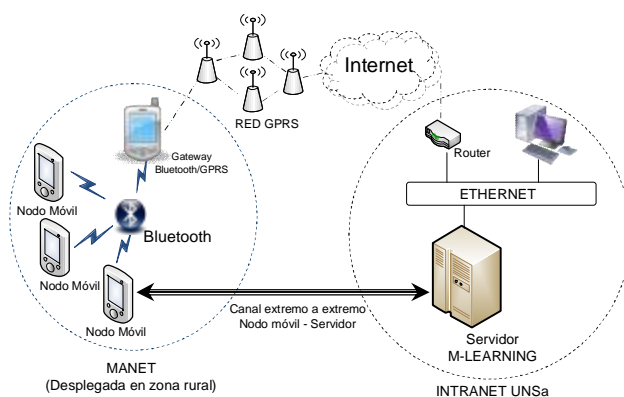
En este trabajo desplegamos una MANET en una zona rural de recursos limitados, e integramos la misma a la red de infraestructura de una organización (Intranet)



utilizando los servicios de la red celular. De esta manera se posibilita el acceso de los nodos ad hoc a recursos M-learning almacenados en un servidor de la intranet.

El despliegue y la integración de la MANET se realizaron considerando los siguientes inconvenientes y limitaciones:

- La energía en la zona de despliegue es escasa, lo que dificulta la capacidad de recarga de los dispositivos que forman parte de la MANET.
- Las redes celulares en zonas remotas no brindan servicios de tercera (3G) o cuarta generación (4G), solo se dispone de tecnología 2G (GSM/GPRS) [4] que proporciona un ancho de banda limitado y variable.
- La mayor parte de los dispositivos móviles utilizados en zonas rurales son equipos de características básicas y funcionalidades limitadas, que incorporan tecnologías como Bluetooth y 2G en lugar de WiFi y 3G.
- Las MANETs y las redes celulares utilizan un medio compartido (aire) para transmitir los datos y se encuentran expuestas a “ataques” o accesos no autorizados. Se requiere entonces la utilización de canales de comunicación “seguros”.
- La implementación de niveles de seguridad elevados implica un incremento del consumo de ancho de banda y de la energía en los nodos móviles [8].



**Figura 1.** Escenario de prueba

En la figura 1 se observa la representación gráfica del escenario que implementamos para realizar las pruebas y mediciones. En el mismo se conecta una MANET, desplegada en zona rural, a una Intranet a través de la red celular. Los dispositivos móviles (nodos) de la MANET se conectan al servidor de recursos M-learning utilizando un canal lógico extremo a extremo. El tráfico entre el nodo móvil y el servidor se gestiona a través de uno de los nodos que actúa como Gateway entre la MANET y la red celular. Este nodo es el encargado de enviar los paquetes de datos hacia los routers de la red celular; desde donde y a través de Internet son direccionados a la intranet para ser entregados al servidor de recursos.

El despliegue de las MANETs se puede realizar utilizando alguno de los siguientes estándares de comunicaciones inalámbricas de corto alcance:

Bluetooth (IEEE 802.15.1), Ultra-wideband (UWB, IEEE 802.15.3), ZigBee (IEEE 802.15.4) y WiFi (IEEE 802.11) [6].

La integración de la MANET a la red de infraestructura, requiere el uso de alguna de las siguientes tecnologías de red celular: 2G (GSM), 2.5G (GPRS), 3G (UMTS, HSDPA y HSUPA) y 4G (LTE) [5].

Para la construcción de nuestro escenario de pruebas elegimos Bluetooth [3] para el despliegue de la MANET y GPRS [4] para la integración a la red de infraestructura. La fundamentación se presenta en un trabajo previo del grupo [7].

En Argentina, el m-learning se ha convertido en una tendencia vinculada a las propuestas educativas [9, 10, 11, 12].

El m-learning se puede presentar en tres modos diferentes, y a su vez cada uno de ellos involucra diversas estrategias de aprendizaje mediadas por tecnologías móviles. Según tal, los modos son los siguientes [13]: recuperación de información, recopilación y análisis de información y comunicación, interacción y colaboración en redes.

Específicamente, esta investigación trata de aportar soluciones a la educación secundaria en zonas desfavorables en las cuales están vigentes los proyectos de itinerancia para garantizar los dos primeros años de secundaria en zonas donde no existen establecimientos educativos de este nivel. Estos proyectos financian la figura de “profesores itinerantes” que dan clases de asignaturas específicas de 1° y 2° año de la secundaria, en las escuelas primarias de una determinada zona rural, donde 7 de cada 10 escuelas no tienen acceso a la red de distribución de energía eléctrica y utilizan paneles solares o grupos electrógenos para cubrir sus necesidades energéticas.

El profesor itinerante puede utilizar la MANET y el escenario presentado para aplicar estrategias de m-learning. Estas estrategias corresponderían al 1° modo, es decir, los alumnos acceden, mediante los dispositivos móviles que forman parte de la MANET, a objetos de aprendizaje de un área de conocimiento específica.

### 3 Líneas de investigación y desarrollo

Las principales líneas de investigación de esta propuesta son:

Despliegue de MANETs en zonas rurales con recursos limitados e integración de las mismas a redes de infraestructura.

Establecimiento de canales extremo a extremo (no seguros y seguros) entre un dispositivo móvil, que forma parte de la MANET en zona rural, y un servidor con recursos de m-learning instalado en una red de infraestructura.

Estudio del impacto de la seguridad en el rendimiento de los dispositivos móviles, dado que el establecimiento de canales seguros implica un consumo adicional de energía (algoritmos de encriptación) y ancho de banda (transmisión de encabezados adicionales) [8].

Adaptación de aplicaciones y recursos m-learning para optimizar el consumo de energía y el aprovechamiento de ancho de banda en los dispositivos móviles.

#### 4 Resultados y Objetivos

El objetivo general que guía esta investigación aplicada es:

Brindar el acceso a recursos m-learning a personas que se encuentran en zonas rurales de recursos limitados, mediante el despliegue e integración de MANETs.

Para ello, se persiguen los siguientes objetivos específicos:

- Implementar escenarios que sirvan de soporte para la distribución de contenidos M-learning utilizando dispositivos móviles ubicados en zonas rurales.

- Estudiar tecnologías para el despliegue y formación de MANETs, y seleccionar la que mejor se adapte a zonas de recursos limitados.

- Estudiar tecnologías para la integración de MANETs a redes de infraestructura y seleccionar la adecuada para realizar la integración desde zonas rurales.

- Estudiar protocolos de seguridad extremo a extremo (nodo móvil a servidor) y seleccionar un nivel de seguridad óptimo sin comprometer el consumo de recursos de los dispositivos.

Los resultados preliminares de nuestra investigación fueron publicados en CACIC 2012: "Integración Segura de MANETs, desplegadas en zonas de recursos limitados, a Redes de Infraestructura" [7].

#### 5 Formación de recursos humanos

La propuesta involucra la integración de los conocimientos en esta área por parte de investigadores de UNSa (que tiene antecedentes en el tema) y de UNSE (investigadores que recién se inician en este tema).

El proyecto cuenta con la participación de estudiantes avanzados de carreras de grado en Informática, pertenecientes a la UNSE y se espera incorporar alumnos de la UNSa. Estos reciben formación en el área redes informáticas y experiencia en el desarrollo de investigaciones.

Uno de los investigadores de esta línea está desarrollando sus tesis de Maestría en Redes (UNLP); y dos alumnos están desarrollando su trabajo final de

grado de Licenciatura en Sistemas de Información (UNSE).

#### Referencias

1. IETF MANET Active Work Group, <http://tools.ietf.org/wg/manet>
2. IEEE 802.15 WPAN task 1, <http://www.ieee802.org/15/pub/TG1.html>
3. Bluetooth Special Interest Group: "Bluetooth Profiles Specification Version 1.1", en Specification of the Bluetooth System, tomo 2, Febrero 2001.
4. ETSI EN 301 344, Digital cellular telecommunications system, General Packet Radio Service (GPRS), Service description, V7.4.1, 2000.
5. Carlos de Morais Cordeiro and Dharma Prakash Agrawal. Integrating MANETs, WLANs, and Cellular Networks. En su: Ad Hoc and Sensor Networks - Theory and Applications. 2nd Ed. Singapore: World Scientific Publishing, 2011. pp. 587-620. ISBN: 978-9814338899.
6. Carlos de Morais Cordeiro and Dharma Prakash Agrawal. Wireless PANs. En su: Ad Hoc and Sensor Networks - Theory and Applications. 2nd Ed. Singapore: World Scientific Publishing, 2011. pp. 196-258. ISBN: 978-9814338899.
7. Sergio H. Rocabado Moreno, Javier Díaz, Ernesto Sánchez y Daniel Arias Figueroa. Integración Segura de MANETs, desplegadas en zonas de recursos limitados, a Redes de Infraestructura. CACIC 2012. Bahía Blanca, Nov. 2012.
8. P. Ni, Z. Li: "Energy cost analysis of IPSec on handheld devices", Elsevier (2004)
9. Cukierman, U., Virgili, J. La Tecnología educativa al servicio de la educación tecnológica. UTN, Bs.As., 2010.
10. Cukierman, U. y Otros. Una experiencia de uso de celulares en un curso de articulación escuela media y universidad en modalidad a distancia. VirtualEduca, Brasil, 2007.
11. Lion, C. Imaginar con Tecnologías: Relaciones entre tecnologías y conocimiento. Buenos Aires, 2007.
12. UNESCO. Declaración mundial sobre la educación superior en el siglo XXI: visión y acción. París, 2009.
13. Woodill, G. The mobile learning edge. Ed. Mc Graw Hill, 2011.

## SARA Six

### Análisis, Implementación y Evaluación de Servicios Colaborativos Competitivos Aplicados a Redes de Avanzada

Gustavo. Mercado, Carlos Taffernaberry, Raúl Moralejo, Cristian Pérez Monte, María Inés Robles,  
Marcela Orbiscay, Sebastián Tobar, Patricia Clérigo, Carlos Tiviroli

gridTICS – Grupo UTN de I&D en Tecnologías de la Información y las Comunicaciones  
Departamento de Electrónica - UTN Facultad Regional Mendoza  
Rodríguez 273, Capital - Mendoza  
{gustavo.mercado,carlos.taffernaberry,}@gridtics.frm.utn.edu.ar

#### Resumen

El uso de redes académicas avanzadas está produciendo cambios importantes en el ámbito de la investigación y la educación, otorgando nuevas herramientas que nos acercan cada vez más a otras comunidades científicas y educativas del mundo. Esto permite una mayor interacción y apoyo entre investigadores, colaboración en investigación científica avanzada, etc.

Las redes de avanzada de la región carecen de las características y bondades de las redes avanzadas de orden mundial debido a varios factores. Estos elementos hacen que haya falta de oportunidades para los investigadores y académicos, poca competitividad en I+D, poca interacción con otras comunidades de científicos e investigadores, desaprovechamiento de las infraestructura disponible, etc.

En este sentido, el proyecto “Análisis, implementación y evaluación de servicios colaborativos competitivos aplicados a redes de avanzada” pretende ser un ambiente de desarrollo que permita:

El ESTUDIO de redes Avanzadas, comprendiendo cabalmente sus funciones, objetivos y alcances,  
El DISEÑO y actualización de servicios y aplicaciones para que puedan operar en redes académicas avanzadas,  
El DESARROLLO e IMPLEMENTACIÓN de nuevos servicios y aplicaciones en redes avanzadas en la región,  
El ENSAYO de las funcionalidades como Calidad de Servicio, IPv6 y VoIP sobre redes avanzadas regionales  
La CAPACITACIÓN y la DIFUSIÓN para ayudar a la comunidad Académica, científica y educativa regional a comprender y utilizar nuevas tecnología, permitiendo de esta manera eliminar las deficiencias enumeradas anteriormente.

#### Palabras Clave

Redes de Avanzada, IPv6, VoIP, Movilidad IP, Qos, GPU

#### Contexto

La presente investigación tiene sostén financiero en el proyecto UTN FRM 25/J069 “SARA Six: Análisis, implementación y evaluación de Servicios colaborativos competitivos Aplicados a Red de Avanzada experimental de la región”, 2010 – 2012, Director: Gustavo Mercado.

Tiene como antecedente el proyecto 25/J052 “CODAREC6 INTRANET Diseño y simulación de la implementación de tecnologías y procedimientos de

transición del protocolo IPv6 en INTRANETS usando el “ip6 test bed” 2007-2009 Director: Ing. Gustavo Mercado.

Además ha generado el proyecto PICTOPRO “PMIP6; Análisis, Evaluación y Comparación de ambientes Proxy Mobile IP en versión 6, aplicado a Redes de Avanzada”, 2011-2012, Director: Taffernaberry, Juan Carlos

#### Introducción

La revolución del siglo XX produjo la aparición de Internet, la cual introdujo cambios en todos los estados de nuestras vidas. Sin embargo, a poco tiempo de su incorporación en nuestro que hacer, se fue sintiendo un vacío e insatisfacción en la comunidad científica y educativa, en la academia y en la investigación [1]. La solución casi no tardó, y fueron los avances en infraestructura tecnológica desarrollados por las redes avanzadas lo que reinstaló la esperanza en las comunidades de investigación, científicas y académicas, pues pusieron a su disposición a través de una gran infraestructura tecnológica condiciones de uso casi en exclusividad de herramientas y aplicaciones que les permitieron el desarrollo e incremento de sus actividades, lo que era insostenible a partir de la Internet comercial.

Hoy estas redes son conocidas como redes académicas avanzadas o redes de investigación [2], y su característica principal es que permiten trabajar a la comunidad de investigadores y académicos mediante mecanismos (herramientas o aplicaciones) de colaboración, para compartir información y recursos a través de una serie de redes interconectadas.

Utilizando las herramientas y aplicaciones desarrolladas en las redes académicas avanzadas se dio un fuerte impulso a la investigación y a la educación, y los investigadores y académicos llevan a cabo un sin número de proyectos afines que están en la frontera del conocimiento, y que han sido posible gracias a la rápida evolución de las tecnologías de telecomunicación, particularmente, las de intercambio y comunicación de datos.

En este contexto se desarrollan y mejoran métodos de trabajo e investigación que refuerzan la colaboración entre equipos, técnicos, académicos y científicos, ubicados en lugares dispersos geográficamente. Desde las videoconferencias de alta calidad, combinadas con herramientas que generan espacios virtuales comunes de trabajo, hasta la formación de Mallas (Grid) que conforman centros, institutos virtuales de investigación,

en los cuales se construyen y combinan nuevas formas de hacer investigación, ciencia, tecnología, arte y educación que permiten reinventar cada día el conocimiento del futuro [3].

Una aplicación es toda aquella herramienta que se desarrolla, se construye y utiliza sobre la red para el desarrollo de la ciencia, la educación y la investigación, por lo mismo, requiere para su funcionamiento de estas redes avanzadas. Así, todas las herramientas y servicios que operan a través de las redes avanzadas son, a su vez, productos de aplicaciones que han sido desarrolladas por expertos y marcan una gran diferencia en el cómo se llevan a cabo hoy los procesos de enseñanza y aprendizaje e investigación.

Por otro lado el protocolo IPV4 comienza a dar señales de debilidad [4]. Después de 20 años, la versión 4 del protocolo de Internet (IP) ya no puede seguir brindando respuestas adecuadas, sobretodo en cuanto al paulatino agotamiento de las direcciones IP disponibles, un proceso que culminará en unos pocos años, al ritmo actual de crecimiento de la red.

Ante el enorme crecimiento de usuarios de Internet, que hoy tienen exigencias distintas a las de hace unos años, las poco más de cuatro mil millones de direcciones en todo el mundo que posibilita el IPv4 se han vuelto insuficientes. El tiempo de vida de IPv4 fue extendido por 10 años gracias a técnicas tales como reutilización de direcciones con traducción (NAT) y uso temporal de asignaciones (CIDR); este fue el período necesario para que IPv6, sucesor de la versión 4, crezca y se establezca definitivamente [5].

Finalmente en 1992 la Internet Engineering Task Force (IETF), que desarrolla los protocolos estándar para Internet, convocó a la comunidad de investigadores a estudiar alternativas para el IPv4. El resultado llegó en 1995 y se llamó IPv6 (Internet Protocol versión 6) [6]. Si bien por estos días IPv6 es especialmente atractivo para los pioneros en los sectores de redes inalámbricas, de juegos, de uso doméstico, redes de investigación nacional conectadas a nivel mundial, organismos militares y gobierno, una vez estandarizado, entre 2005 y 2008, ofrecerá importantes beneficios a las empresas. Por su parte, los proveedores de servicios Internet (ISPs) móviles, locales y regionales ya están evaluando los productos y plataformas derivados de la tecnología IPv6, para que se puedan comercializar en un futuro inmediato.

Algunas características de IPv6 son:

- Capacidad de direccionamiento expandida: en IPv6, las direcciones se componen de 16 octetos (8 bits de la forma 00001111), es decir, 128 bits. Esto da lugar a  $2^{128}$  (340 sextillones) de direcciones IP. Para apreciar la magnitud de esta cifra, puede decirse que equivale a 1,1 mol de direcciones por cada metro cuadrado de superficie terrestre (1 mol = 602000 trillones, el llamado Número de Avogadro). Si cada una ocupara 1 mm<sup>3</sup>, el espacio de direcciones IPv6 tendría el volumen de un cubo de 6,98 millones de km de arista, 547 veces el diámetro ecuatorial de la Tierra.

La población mundial actual (6000 millones de habitantes) tardaría 1798 trillones de años en contar esta

cantidad de direcciones a razón de una por segundo. A efectos de notación, los 128 bits de las direcciones IPv6 se agrupan en 8 palabras de 16 bits (o 4 dígitos hexadecimales) separadas por doble punto, por ejemplo: FEDC:BA98:7654:3210:0123:4567:89AB:CDEF.

- Calidad de servicio (Qos): IPv6 puede diferenciar los paquetes de datos como pertenecientes a un flujo particular, y así otorgar un ancho de banda en función de cada necesidad, ya sea para correo electrónico, comunicaciones de voz o videoconferencia.

- Capacidades de autenticación y privacidad: IPv6 emplea como parte integral el entorno de seguridad IPSec, que no está implementado en los hosts del IPv4 en forma nativa.

- Autoconfigurable (Neighbour Discovery): en IPv6 los nodos no necesitan ser configurados manualmente.

- End to end: IPv6 no usa NAT ya que tiene direcciones globales para todos los nodos. Así, éstos pueden reenviar cada paquete sin alterar su contenido.

- Simplificación del formato del encabezamiento: Es más sencillo y su tamaño es fijo. Se han suprimido campos como el checksum, ToS y fragmentación, y agregado uno para identificar flujos de datos. Las funciones de los campos eliminados se logran con encabezados de extensión, que permiten incorporar nuevas características al protocolo, como IPSec o movilidad.

Ahora es tiempo de admitir que se ha agotado el tiempo de IPv4 y que el despliegue de IPv6 se hace imprescindible. En el país son varias las instituciones que han adoptado IPv6 para sus redes, aunque por ahora en forma experimental. Cuando las organizaciones tomen la decisión de hacer la transición el proceso se deberá llevar a cabo tareas tales como: entrenamiento de personal, actualización de routers, sistemas operativos y herramientas de manejo, y el despliegue de aplicaciones compatibles con IPv6 (IPv6 ready).

La conjunción entre Redes Avanzadas e IPv6 forma una herramienta poderosa y a la vez simple para cumplir con el objetivo de dar impulso a la investigación colaborativa. Hay varios ejemplos en el mundo de Redes Avanzadas centradas en IPv6 desde la red de avanzada por excelencia como Internet 2 ([www.internet2.edu](http://www.internet2.edu)) en Estados Unidos, pasando por la red europea Geant2 ([www.geant2.net](http://www.geant2.net)) y la asiática APAN (Asia-Pacific Advanced Network Consortium-[apan.net](http://apan.net)). Como axial también en Latinoamérica tales como, Clara (Cooperación Latinoamericana de Redes Avanzadas - [www.clara.net](http://www.clara.net)), RNP de Brasil (RNP – Rede Nacional de Ensino e Pesquisa -[www.rnp.br](http://www.rnp.br)), CUDI de Méjico (Corporación Universitaria para el Desarrollo de Internet – [www.cudi.edu.mx](http://www.cudi.edu.mx)) y Reuna de Chile (Red Universitaria Nacional - [www.reuna.cl](http://www.reuna.cl)).

En nuestro país existe la red Inova ([www.innovared.net](http://www.innovared.net)) que da acceso a redes avanzadas a las instituciones nacionales por intermedio de Clara.

Dentro de la UTN existe la red RUT2 (Red Universitaria Tecnológica) que conecta a las Facultades Regionales.

Sin embargo en nuestro país no están desarrolladas las redes metropolitanas/provinciales como en Brasil. La red



Redecomep es una iniciativa del Ministerio de Ciencia y Tecnología de Brasil (MCT), coordinada por Rede Nacional de Ensino e Pesquisa (RNP), que tiene como objetivo implementar redes de alta velocidad en regiones metropolitanas del país servidas por los Puntos de Presencia de RNP.

En Mendoza existe la iniciativa ACyT Net (Red Académica Científica y Tecnológica de Mendoza) que une, al momento, el Cricyt Conicet, el Instituto Regional del INA y el GridTICS de la UTN FRM.

Esta red, que pretende ser la interconexión de avanzada de las instituciones metropolitanas de Mendoza, está en fase de desarrollo.

Este proyecto, SARA-Six, pretende impulsar esta red, instalando y evaluando aplicaciones competitivas y cooperativas, tales como IPv6, VoIP, Movilidad, calidad de Servicio Clustering, Virtualización entre otras.

### Líneas de investigación y desarrollo

El grupo de tecnología IPv6, perteneciente al grupo UTN GridTICS, se constituye en 2005 y ha tenido una vasta actividad y experiencia y es reconocido como uno de los grupos pioneros en IPv6 de la región. El grupo ha realizado publicaciones [12], presentación en congresos [13,14,15], cursos de grado y posgrado y asistencia a tesinas de grado [16] y tesis de posgrado. Además de participar activamente en las iniciativas de ISOC y de LANIC para la promoción y difusión de IPv6 [17,18,19], siendo también socio activo de la IPv6 Task Force Argentina [29].

### Objetivos y Resultados

#### Objetivo Principal

Analizar, implementar y evaluar servicios colaborativos competitivos aplicables a redes de avanzada en la región

#### Objetivos Secundarios

- Convergencia de multimedia y Qos sobre redes de avanzada con IPv6.

Organizar, implementar y evaluar metodologías para la convergencia de Voz Video y Datos en redes avanzadas con IPv6.

- Movilidad en Redes de Avanzado con IPv6.

Organizar, implementar y evaluar PMIP (Proxy Mobile IP), a través de un prototipo en red de avanzada con IPv6.

- GPU aplicado a seguridad en redes de avanzada e IPv6

Analizar y evaluar metodologías de mejora de la seguridad en redes de avanzado usando GPU

- Clustering y Virtualización en redes de avanzada con IPv6.

Estudiar, evaluar e implementar prototipos con mejoras para Clustering y Virtualización utilizando redes avanzadas e IPv6

- Metodología de producción de software

Investigar e implementar metodologías de calidad del desarrollo personal de software aplicado a la programación en networking.

- Difusión y Capacitación:

Participar en el fortalecimiento y difusión de las redes de avanzada y sus aplicaciones, incentivando su uso e implementación en el medio regional. Fortalecer la preparación de recursos humanos.

#### Avances y resultados preliminares

Se avanza sobre el objetivo de "Convergencia de multimedia y Qos sobre redes de avanzada con IPv6", realizando un análisis de convergencia de multimedia y Qos sobre IPv6 y diseñando e implementando funcionalidades de VoIP y sus sistemas más usados.

Se avanza sobre el objetivo de "Movilidad en Redes de Avanzado con IPv6", realizando el relevamiento de los sistemas de movilidad en IPv6. Se instala de sistema operativo con soporte de MIPv6. Se implementa stack de movilidad dentro del ámbito de la red de avanzada SARA Six.

Se avanza sobre el objetivo "GPU aplicado a seguridad en redes de avanzada e IPv6", realizando estudios teóricos y prácticos de la arquitectura de hardware de GPU y se parametrizan la GPU utilizando las métricas de contrastación entre teoría y práctica. Se realizan estudios técnicos de distribución de cómputo y se realizan mediciones de performance, tareas de optimización y estimaciones de máxima performance.

Se avanza sobre el objetivo "Clustering y Virtualización en redes de avanzada con IPv6" realizando estudios de métodos y tecnologías de clustring y virtualización y su implicancia en IPv6. Se comienza con la construcción de la extensión de la red AcyTNET desde su nodo principal hasta el laboratorio del gridTICS de la UTN mediante el tendido de fibra óptica.

Se avanza sobre el objetivo "Metodología de producción de software", realizando estudios y aplicaciones de los Modelos y Estándares de Evaluación y Mejora del Proceso de Software. Se estudia y se aplica la Perspectiva CTS (Ciencia, Tecnología y Sociedad) en el dictado de materias de grado y posgrado, y en el desarrollo de aplicativos (centrado en las personas y la naturaleza). Se estudia y aplicación de Herramientas de Software Libre (sistema operativo, base de datos, lenguaje de programación, versionado, pruebas, etc.) y se diseñó, desarrolló e implementó una Aplicativo de soporte para la carga y emisión de reportes del Instrumento COCTS en el marco del proyecto EANCYT (Enseñanza y Aprendizaje sobre la Naturaleza de la Ciencia y Tecnología).

Se avanza sobre el objetivo "Difusión y Capacitación", realizando múltiples presentaciones a congresos, cursos de grado y postgrado y seminarios

A continuación se agregan actividades de Difusión y Capacitación que no están incluidas en la sección de Publicaciones y Congresos:

- JORNADA MENDOZA WORLD IPV6 LAUNCH  
6 de Junio de 2012  
Facultad Regional Mendoza - Universidad Tecnológica Nacional  
Mendoza - Argentina  
Organizado por Mendoza IPv6 Group
- MENDOZA IPV6 SUMMER DAY  
10 de febrero de 2012

Aula Magna - Facultad de Ingeniería - Universidad de Mendoza

Mendoza - Argentina

Organizado por Mendoza IPv6 Group

- MENDOZA IPv6 DAY

8 de Junio de 2011

Facultad Regional Mendoza - Universidad

Tecnológica Nacional

Mendoza - Argentina

Organizado por Mendoza IPv6 Group

### Formación de Recursos Humanos

En el proyecto participan un doctorando con becas UTN, dos doctorandos docentes UTN, dos investigadores graduado con becas BINID UTN y 3 becarios alumnos con beca UTN.

### Referencias

- [1] José Silvio "Redes Académicas y Gestión del Conocimiento en América Latina: En Busca de la Calidad Educación Superior y Sociedad", Vol 3, No 2 1992 Copyright © 2010 UNESCO-IESALC.
- [2] Carmen Labbé, "Redes Avanzadas y Comunidades de Investigación para la Gestión del Conocimiento" Organización Universitaria Interamericana, Dec 2009
- [3] Pedro Munive S. y Mariana Rodríguez M, "Las redes académicas avanzadas, una oportunidad para compartir eventos interactivos de divulgación", X Reunión de la Red de Popularización de la Ciencia y la Tecnología en América Latina y el Caribe (RED POP - UNESCO) y IV Taller "Ciencia, Comunicación y Sociedad" San José, Costa Rica, 9 al 11 de mayo, 2007
- [4] Geoff Huston "Address Exhaustion" The Internet Protocol Journal, Volume 14, No.1, March 2011.
- [5] Niall Murphy, David Wilson, "The End of Eternity Part One: IPv4 Address Exhaustion and Consequences", The Internet Protocol Journal, Volume 11, No.4, Dec 2008
- [6] S. Deering, R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", RFC 2460, IETF, Dec 1998
- [11] G. Montenegro, J. Hui, D. Culler, "Transmission of IPv6 Packets over IEEE 802.15.4 Networks", RFC 4944, IETF, September 2007
- [12] Cristian Pérez Monte, María Inés Robles, Gustavo Mercado, Carlos Taffernaberry, Marcela Orbiscay, Sebastián Tobar, Raúl Moralejo y Santiago Pérez, "Implementation and Evaluation of Protocols Translating Methods for IPv4 to IPv6 Transition", Journal of Computer Science & Technology, ISSN 1666-6038 Vol. 12 No. 2 August 2012
- [13] Carlos Taffernaberry "6LoWPAN - IPv6 para WSN", SASE - Simposio Argentino de Sistemas Embebidos 2012, Facultad de Ingeniería de la Universidad de Buenos Aires, Buenos Aires, Agosto 2012.
- [14] Gustavo Mercado, Cristian Pérez Monte, Carlos Taffernaberry, María Inés Robles, Marcela Orbiscay, Sebastián Tobar, Raúl Moralejo, y Santiago Pérez, "Implementación y Evaluación de métodos de Traslación de Protocolos para la transición IPv4-IPv6". Anales del CACIC 2011, ISBN 978-950-34-0756-1, UNLP, La Plata, Octubre 2011.
- [15] C. Taffernaberry, G. Mercado, S. Tobar, C. Pérez Monte, P. Clérigo, I. Robles, M. Orbiscay, S. Pérez, R. Moralejo, "PMIP6: Análisis, Evaluación y Comparación de ambientes Proxy Mobile IP en versión 6, aplicado a Redes de Avanzada", Anuales del WICC 2011. XIII Workshop de Investigadores en Ciencias de la Computación, ISBN 978-950-673-892-1, UNR Editora. Editorial de la Universidad Nacional de Rosario, 2011
- [16] Carlos Tiviroli, Andres Gatti, Gustavo Mercado y Carlos Taffernaberry "QoSIP Meter: Sistema de determinación de condiciones de calidad de servicio en transmisiones de audio/video en tiempo real sobre Internet" CASE 2012 LIBRO DE TRABAJO pag. 263-264 ISBN 978-987-9374-82- Facultad de Ingeniería de la Universidad de Buenos Aires, Buenos Aires, Agosto 2012.
- [17] Robles I, Orbiscay M. "Educando en IPv6 a través de Conferencias y Talleres Abiertos", LANIC 18/ LACNOG 2012, Montevideo, Uruguay, Oct 2012
- [18] G. Mercado y C. Taffernaberry, "6LoWPAN IPv6 for Wireless Sensor Network", LACNIC XVI LACNOG 2012 ION, Buenos Aires, Oct 2011
- [19] G. Mercado "6LoWPAN IPv6 for Wireless Sensor Network", 9º Foro Latinoamericano de IPv6 – FLIP6, LACNIC XV 2011, Cancun - Mexico 15 al 20 de mayo de 2011.
- [20] Task Force IPv6 AR [www.ipv6.org.ar](http://www.ipv6.org.ar)

## TRANSACTION SIGNATURE (TSIG). UNA ALTERNATIVA DE SEGURIDAD PARA TRANSFERENCIAS DE ZONAS DNS.

Sánchez Ernesto, Di Mauro Juan, Silvera Jorge Arias Figueroa Daniel.  
C.I.D.I.A./Departamento de Informática/Facultad de Ciencias Exactas/Universidad Nacional de Salta  
Av. Bolivia 5150  
0387-4255547

[esanchez@cidia.unsa.edu.ar](mailto:esanchez@cidia.unsa.edu.ar); [juannombrepellido@gmail.com](mailto:juannombrepellido@gmail.com); [jsilvera@arnet.com.ar](mailto:jsilvera@arnet.com.ar);  
[daaf@cidia.unsa.edu.ar](mailto:daaf@cidia.unsa.edu.ar)

### RESUMEN

Desde su creación el Sistema de Nombres de Dominio, ha carecido de un diseño que asegure la comunicación entre las partes que intervienen en el proceso de resolución de nombres. Y dada su condición de sistema público y estructura jerárquica, es que se encuentra expuesto a posibles vectores de “ataque” a lo largo de todo el flujo de datos intercambiado por los componentes (Clientes y Servidores) que lo conforman.

Es así que, en el Centro de Investigación y Desarrollo en Informática Aplicada (C.I.D.I.A.) perteneciente a la Universidad Nacional de Salta se formó un equipo de trabajo que se encuentra realizando una investigación aplicada cuyo objetivo es presentar de manera práctica el conjunto de vulnerabilidades presentes en el Sistema de Nombres de Dominio y definir las medidas a implementar para mitigar las mismas.

Se detalla a continuación un caso práctico sobre la vulnerabilidad presente en una transferencia de zona y la alternativa *TSIG* (*Transaction Signature*) como parte de un conjunto de medidas para dotar de seguridad a los procesos involucrados en tales transferencias.

#### **Palabras clave:**

*Internet, Sistema de Nombres de Dominio, DNS Seguro, DNSSEC, TSIG. Transferencias de zonas DNS.*

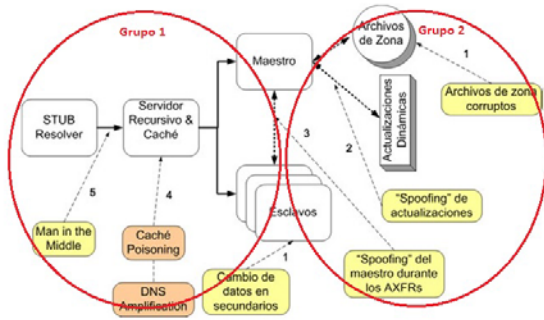
### CONTEXTO

Lo expuesto en el presente trabajo, surge del proyecto de investigación “Extensiones de Seguridad para el Sistema de Nombres de Dominio” (Consejo de Investigación de la Universidad Nacional de Salta), en conjunto con el Centro de Investigación y Desarrollo en Informática Aplicada (C.I.D.I.A. – UNSA).

#### 1. INTRODUCCION

La condición de sistema jerárquico del Sistema de Nombres de Dominio, lo expone a posibles vectores de “ataque” a lo largo de todo el flujo de datos intercambiado por los componentes (Clientes y Servidores) que lo conforman.

Las tareas iniciales en el marco del proyecto de investigación citado en el apartado anterior, fueron identificar y clasificar las amenazas a la seguridad de un Sistema DNS. El siguiente gráfico resume en términos generales los puntos críticos y se distinguen dos grandes grupos en el tráfico de red intercambiado [2]. El primer grupo se corresponde al tráfico intercambiado en procesos de consultas DNS (solicitud/respuesta) entre Clientes y Servidores, el segundo corresponde al tráfico correspondiente a la administración de archivos de zonas (actualización y transferencias de zonas) entre servidores maestro/esclavo.



En una etapa posterior, se seleccionó el segundo grupo de vulnerabilidades y se propuso el desarrollo de una aplicación en lenguaje Python, para enumeración y análisis de redes TCP/IP “Guesso” [4]. Mediante los módulos específicos para el protocolo DNS (*mydnsinterface* y *mydns*), basados en la librería *pydns* (bajo licencia OpenSource), se demostró la posibilidad de obtener información basada en transferencias de zonas DNS.

En resumen, la técnica consistió en hacer consultas DNS para cada uno de los tipos listados a continuación [2]:

Nombre	Descripción	código
A	dirección IP	1
NS	servidores DNS	2
CNAME	nombres canónicos	5
PTR	registro de puntero	12
HINFO	información de host	13
MX	Mail eXchange (servidores de correo)	15
AXFR	pedido de transferencia de zona	252

La tarea de obtención de información, se automatizó mediante el siguiente algoritmo, que volcaba los resultados obtenidos en una estructura de datos, lo que permitió enumerar los hosts pertenecientes a un dominio en particular:

```
def get_domain_hosts(dominio):
    tipos = ['NS','MX','A','PTR','CNAME','A']
    respuestas = [] //lista vacía
    for t in tipo:
        respuestas.add(consulta(dominio,t))
    return procesar_respuesta(respuestas)
```

Se puede observar que la función recibe un nombre de dominio como argumento y retorna un diccionario indizado por tipo de

consulta (o registro) cuyos valores son las respuestas de los servidores DNS para esa consulta.

Expuesta la vulnerabilidad para en caso de transferencia de zonas (tipo de consulta AXFR), y según se describe en [1], surge de manera natural como primera medida para mitigar este tipo de vulnerabilidad, denegar todo tipo de solicitud de transferencias de zonas y permitir las, en caso de ser necesario, solo para un grupo restringido de host (listas de accesos de IP origen).

Del mismo modo que se presenta, como una rápida y sencilla alternativa a implementar, así se descarta, ya que la misma está expuesta a ataques de “suplantación de identidad” (*IP Spoofing*), mediante la técnica de “hombre en el medio” [7].

Es así que el siguiente nivel en el proceso de asegurar la transferencia de zonas fue implementar el protocolo TSIG [3], lo que permite la autenticación entre las partes basada en el uso de técnicas criptográficas, que aseguran que Cliente y Servidor son quienes dicen ser y por otro lado, asegurar la integridad de los datos (los datos recibidos por el Cliente son los que envió el Servidor).

**Transaction Signature (TSIG):** Definido en el RFC 2845, es un mecanismo que hace uso de una clave única mediante la implementación de Códigos de Autenticación de Mensaje entre servidores maestros y esclavos. La distribución de la clave hacia los servidores esclavos debe hacerse de manera segura utilizando medios como email seguro, fax o correo postal, se recomienda para la misma, periodos de actualización de entre 30 y 60 días. En caso de presencia de más de un servidor esclavo, se deben utilizar pares de claves (maestro-esclavo) diferentes, ya que al verse comprometida alguna de las claves, solo de deshabilitarían temporalmente las transferencias entre las partes afectadas, sin involucrar al resto de las transferencias.

A los fines prácticos, se menciona que TSIG fue configurado para DNS BIND versión



9.9.2-P1, sobre plataforma Linux Fedora 17. La clave compartida se generó mediante la utilidad *dnssec-keygen* [5][6].

Los resultados obtenidos en cada una de las etapas descritas anteriormente, nos permiten concluir que no existe una solución integral que permita dotar de seguridad al protocolo DNS y a las partes intervinientes en los procesos y operaciones de consultas, sino que se deben adoptar un conjunto de medidas que permitirán minimizar la exposición a los posibles ataques a la información intercambiada.

## 2. LINEAS DE INVESTIGACION y DESARROLLO

Los principales ejes temáticos que se están investigando son los siguientes:

- Sistema de Nombres de Dominio.
- Ataques típicos al Sistema de Nombres de Dominio.
- Criptografía de clave privada.
- DNSSEC.
- Transaction Signature (TSIG).

## 3. RESULTADOS OBTENIDOS/ESPERADOS

Con los resultados obtenidos en cada una de las etapas realizadas hasta ahora, se logró documentar un conjunto de políticas de seguridad que se espera sea el marco de referencia para los responsables de administrar un sistema DNS dentro del ámbito educativo. Por otro lado, en términos generales, se pretende acompañar el despliegue global que vienen teniendo las extensiones de seguridad para el sistema DNS.

## 4. FORMACION DE RECURSOS HUMANOS

La estructura del equipo de investigación es de 5 (cinco) miembros incluidos el Director y Co-director.

Dos miembros están realizando el trabajo de Tesis de Posgrado en Redes de Datos, dependiente de la Universidad Nacional de La Plata.

Otros dos participantes se encuentran realizando el trabajo de Tesis de Grado (*DNS Curve*), de la Carrera de Licenciatura en Análisis de Sistemas de la Universidad Nacional de Salta.

## 5. BIBLIOGRAFIA

- [1] AITCHISON Ron, Pro DNS and BIND 10, 2da Edición, Apress, 2011, ISBN 978-1-4302-3049-6.
- [2] FALL Kevin R. - STEVENS W. Richard, TCP/IP Illustrated, Volume 1. The Protocols, 2da Edición, 2012, ISBN-13: 978-0-321-33631-6.
- [3] VIXIE P, GUDMUNDSSON O, EASTLAKE 3rd, D, WELLINGTON, B. RFC 2845: Secret Key Transaction Authentication for DNS (TSIG). Mayo 2000.
- [4] Di Mauro Juan, trabajo de Tesis de pregrado "Guesso, software para enumeración y análisis de redes TCP/IP". UNSa 2012.
- [5] Clarifications and Implementation Notes for DNSSECbis - S. Weiler.
- [6] RFC 4033: DNS Security Introduction and Requirements - R. Arends, R. Austein, M. Larson, D. Massey, S. Rose.
- [7] RFC 3833: Threat Analysis of the Domain Name System (DNS) - D. Atkins, R. Austein

# Generación de una biblioteca para control de tareas en tiempo real en un sistema operativo didáctico

Nicanor Casas, Graciela De Luca, Sergio Martín, Waldo Valiente,  
Gerardo Puyo, Federico Díaz

Universidad Nacional de la Matanza

Departamento de Ingeniería e Investigaciones Tecnológicas

Dirección: *Florencio Varela 1703* - Código Postal: 1754 - Teléfono: 4480-8900/8835

[ncasas@ing.unlam.edu.ar](mailto:ncasas@ing.unlam.edu.ar), {graciela.edl, eienburuu,gerardopuyo, fedediazceo}@gmail.com

## RESUMEN

### RESUMEN

En este trabajo de investigación se estudiaron los métodos de planificación de tiempo real y se buscó generar una biblioteca en busca de aportar soluciones a algunas de las cuestiones la política de planificación de tiempo real. Se dispusieron herramientas teóricas que permitieron conocer si el sistema podrá en todo momento, garantizar la correcta ejecución de todas las tareas críticas. Los planificadores basados en prioridades dinámicas, a pesar de ser capaces de garantizar un mayor número de tareas que los de prioridades estáticas, no disponen de un test de planificabilidad eficiente.

Los sistemas de tiempo real están compuestos tanto por tareas periódicas, que suelen asociarse a actividades críticas, como por tareas aperiódicas, sin ninguna urgencia en su ejecución, es deseable que estas se completen lo antes posible sin poner en peligro los plazos de las periódicas. Se han propuesto dos algoritmos para servir tareas aperiódicas basados en el concepto de holgura. El primero de ellos se apoya en una tabla pre calculada para aceptar las peticiones aperiódicas. El segundo, a cambio de reducir la complejidad espacial, tiene mayor coste temporal pues realiza todos los cálculos dinámicamente. Ambos algoritmos ofrecen, a las tareas aperiódicas, el menor tiempo de respuesta posible.

### PALABRAS CLAVE

Tiempo real, Granularidad, Baseline, Tecnologías Adaptativas

## CONTEXTO

Este trabajo forma parte del proyecto de investigación de la confección de un sistema operativo de características didácticas, el cual se encuentra en el marco de investigaciones que coordinada el departamento de Ingeniería e Investigaciones Tecnológicas que pertenece a la Facultad de Ingeniería de la Universidad Nacional de la Matanza

### INTRODUCCIÓN

El desarrollo de un sistema operativo didáctico de características generales y de estructura tradicional escrito por los alumnos del Departamento de Ingeniería e Investigaciones Tecnológicas de la carrera de Ingeniería en Informática que recibe el nombre de SODIUM (Sistema Operativo Departamento Ingeniería Universidad de La Matanza) llegó a un punto en que debía realizarse un análisis de código a fin de lograr una mejor respuesta, pero principalmente estandarizar los diversos programas.

Fue así que se comenzó por sacar del KERNEL del sistema operativo SODIUM, llamadas al sistema que efectuaban diversas funciones colocándolos en área de usuario., lo que mejoró la comprensión de su función y/o funciones y permitió una mayor estabilidad durante la ejecución del sistema operativo.

Se pensó entonces, en la posibilidad de pasar a un sistema operativo de tiempo real (TR), lo cual llevó a realizar un estudio para detectar diferencias entre lo existente y lo nuevo para saber cuáles son los pro y los contra de incluir o modificar el actual sistema operativo y llevarlo a un sistema operativo en tiempo real.

Para realizar la aplicación del pasaje de un sistema operativo de características tradicionales a otro sistema

operativo de tiempo real se debe basar en un estudio exhaustivo de las diferencias existentes en cada uno de ellos.

Esto nos llevó a la disyuntiva de tener que optar por tres opciones: la primera es generar otro sistema operativo de estudio; la segunda es establecer uno solo que permita la convivencia de los dos sistemas y la tercera transferir el sistema operativo actual a un sistema operativo en tiempo real exclusivamente. En cada una de las opciones se incluye la aplicación de tecnologías adaptativas.

La primera opción, generar otro sistema operativo, tiene las siguientes desventajas:

- a) Que al momento se están desarrollando nuevas funcionalidades definidas para el sistema operativo tradicional que está pasando por alguna de las siguientes etapas: de escritura, de test o de implementación que no podrían ser establecidas para el nuevo.
- b) Realizar una copia del sistema operativo tradicional para luego generar las estructuras necesarias para convertirlo en tiempo real no nos garantiza que las mismas funcionen al estar reestructuradas como servicios, para el primero.
- c) Que toda la programación es desarrollada por los alumnos que cursan la materia Sistemas Operativos, lo que impide solicitar modificaciones porque las mismas podrían comprometer la cursada.
- d) Generar dos sistemas operativos el mantenimiento separado de los mismos requeriría esfuerzos muy importantes que demorarían la estabilidad del sistema.
- e) Que se complica la posibilidad de realizar comparaciones de performance y funcionamiento entre ambos sistemas ya que deberíamos proveer de un tercero que tomara los datos de ambos para unificarlos y facilitar las conclusiones.

La segunda, establecer un solo sistema operativo que permita la convivencia de los demás, tiene los siguientes puntos a favor:

- a) Para que sea un sistema operativo didáctico se deben tener los dos sistemas ejecutándose de acuerdo a las necesidades de los alumnos a fin de que se puedan establecer las comparaciones necesarias.
- b) Permitiría realizar un mantenimiento más acorde a las fuerzas de desarrollo con que cuenta el equipo reduciendo los tiempos de test y de corrección que conlleva cada avance.

Por otro lado tiene como contrapartida los siguientes puntos:

- a) Complica la instalación del sistema operativo debido a que el volumen del mismo hace que tenga un peso importante
- b) Es necesario que partes importantes del Kernel que se desarrollaron y que son compiladas en forma monolítica deban ser compiladas en forma separadas para que los mismos se brinden en función de servicios.
- c) Que el punto anterior, en muchos de los casos sería como generar un nuevo sistema operativo.

La tercera opción, convertir el actual sistema operativo a uno de tiempo real tiene a favor:

- a) Que nos permitiría fijar la atención en un solo desarrollo.
- b) Que los nuevos sistemas operativos están ya prácticamente enrolados en esta característica.

Tienen como contrapartida que:

- a) En las cátedras actuales de la materia Sistemas Operativos se sigue trabajando sobre sistemas operativos tradicionales como punto de partida.

Como consecuencia de lo expuesto se llegó a la siguiente conclusión:

- 1) Generar una función mixta en la cual se mantendría al sistema operativo tradicional pero separando algunos servicios del Kernel actual para que se encuentren ubicadas a nivel de usuario, con el propósito de que las mismas cumplan la función especificada y no influyan en la performance de un modelo u otro.
- 2) Se decidió que se duplicarían las funciones de los diferentes algoritmos para evitar que los sistemas operativos deban ser compilados cada vez que se quiera probar una función en performance.
- 3) Aplicación de tecnologías adaptativas para permitir la transmutación de uno a otro de los algoritmos involucrados como también de la elección del tipo de sistema operativo que se quiere utilizar.

Se llegó a la conclusión de que era factible realizar, en primera instancia, una conjunción entre el Kernel del sistema operativo vigente y agregar los algoritmos de tiempo real utilizando para ello tecnologías adaptativas.

Se realizaron los análisis correspondientes a los diferentes algoritmos que se van a aplicar, teniendo en cuenta que serán en principio los mismos que se utilizan en los sistemas operativos tradicionales, como base de selección, para poder realizar comparaciones necesarias con el objetivo de marcar las diferencias entre los diferentes sistemas operativos.

Sin embargo es propio aclarar que no todos los algoritmos de planificación pueden verificar todos los eventos. Si tomamos como ejemplo los algoritmos FIFO, SJFS o PRORIDADES (non-preemptive) no pueden evaluar los cambios de estado ya que son algoritmos no expropiativos lo que implica que cambiar de algoritmo en medio de la ejecución de un proceso implicaría violar su integridad, porque al cambiarlo en medio de su ejecución se perderían los beneficios que intrínsecamente cada uno tiene.

Por otro lado no son sensibles a la creación de un proceso, ya que estos se encolarán a través de su propio criterio, pero no afectarán la ejecución actual.

Como los eventos a utilizar serán únicamente dependientes del Modo Actual y no al modo en que solicitarán las transiciones será posible codificar funciones en el SODIUM que automáticamente monitoreen los eventos necesarios según el algoritmo de ejecución. Así no hará falta incluir eventos como condiciones de la tabla de decisión final.

Una de las características más importantes del proyecto es que la codificación de los módulos principales que implementarán en el sistema operativo será realizada por los alumnos que cursan la materia Sistemas Operativos Avanzados.

Esto implica un consumo de tiempo elevado, debido a:

- Que los alumnos interactúan con otras materias y con sus responsabilidades laborales lo que determina una carga importante en el desarrollo del sistema operativo.
- Que los nuevos cursos deben tener capacitación sobre lo ya desarrollado para poder continuar con la generación del sistema operativo.
- Generar los trabajos prácticos necesarios para un normal desarrollo del proyecto.

Se presenta entonces la necesidad de acortar tiempos para que el producto pueda estar disponible en un lapso prudencial para que cumpla con los requerimientos para los que fue ideado, siendo la responsabilidad del equipo de investigación de generar los puentes necesarios para amalgamar los diferentes programas generados por los alumnos, como el de preparar la base necesaria para que los próximos cursos puedan continuar con la tarea emprendida.

Uno de esos puentes es la de construir una biblioteca que localice y circunscriba los desarrollos referentes a tiempo real

## PROBLEMAS A RESOLVER

La generación de una biblioteca de tareas que permita el control de planificaciones estáticas en tiempo real se constituyó en un trabajo de gran envergadura por los inconvenientes que se presentaron en la misma.

En un principio la generación de procesos de poca complejidad e independientes a los que se les asignó un prioridad externa, y sin posibilidades de modificación (prioridades internas) pareció ser un elemento simple pero posteriormente se fue complicando en forma escalonada.

La biblioteca está basada en un planificador de características expulsivas (preemptive) que provee una significativa garantía cuando ciertas condiciones de utilización del procesador son reunidas.

En principio las características de construcción de la biblioteca se establecieron en base a programas escritos en lenguaje C que interactuaban en área del Kernel del sistema operativo lo que para programas de usuario cortos no presentaban problemas teniendo un tiempo de respuesta interesante por ser una primera aproximación, aunque pasó a ser lento cuando se incorporaron procesos usuario más sofisticados. Esto hizo que una de las premisas establecidas anteriormente (cambio de tarea – context switch – nulo o despreciable) no se cumplieran, por lo que se está pasando a una colección de procesos livianos, en los cuales los hilos trabajen en área de usuario para obtener una mayor velocidad de respuesta y minimizar los tiempos de context switch, con lo que esperamos lograr una eficiencia máxima.

La biblioteca de tareas comprende un planificador rate monotonic que puede proporcionar una planificación eficiente cuando se cumplen ciertas condiciones en la utilización del procesador.

La diferencia entre un sistema de computación en tiempo real y un sistema de computación de propósitos generales se encuentra no en las especificaciones de performance, pero si en la importancia que cada sistema, en lo referente a tiempos, tiene como consideraciones.

En aplicaciones de tiempo real la exactitud de una computación depende no solo del resultado de la computación sino también del tiempo en que las salidas son generadas.

Pensamos que para poder medir la importancia del desarrollo de la biblioteca tenemos que tener en cuenta los siguientes puntos:

- a) Velocidad predecible para la respuesta a eventos urgentes.
- b) Alto grado de planificación. Definida planificación como el alto grado de utilización de recursos o el menor tiempo de requerimientos que puedan ser garantizados para todos los procesos intervinientes.
- c) Estabilidad bajo carga transitoria. Cuando el sistema es sobrecargado por eventos, el cumplimiento de todos los plazos se hace imposible. Es en este caso en que la biblioteca debe garantizar los plazos para tareas críticas.



Un sistema operativo de tiempo real debe satisfacer estas aplicaciones de computación con plazos implícitos, a través de la fuerza bruta implementada a través de hardware o por un golpe de suerte. Históricamente los más recientes sistemas operativos de tiempo real intentan ser “realmente rápidos” en lugar de ser de “tiempo real”. Esto es simple de explicar pero no tan simple de implementar. Lo último significa que la ejecución en tiempo de los servicios y las operaciones internas de un sistema operativo se ejecuten lo más rápidamente posible, para minimizar el tiempo de ejecución promedio y de esa manera tener un relativamente predecible límite superior para el peor caso de tiempo de ejecución.

También hay que tener en cuenta que esos supuestos a menudo no son explícitamente identificados o siquiera conocidos. Cada sistema puede operar satisfactoriamente en tiempo real y proveer soluciones específicas para ciertas aplicaciones.

Tradicionalmente, un sistema en tiempo real usa funciones cíclicas para planificar hilos de ejecución concurrentes. Bajo este acercamiento, un programador puede establecer una línea de tiempo de ejecución a mano para serializar la ejecución de secciones críticas y reunir plazos de tareas.

Todo esto viene a establecer que las funciones manejables para sistemas simples se torna terriblemente inmanejable para sistemas grandes.

Todo esto implica un doloroso proceso de desarrollo de código de modo que se ajuste a los intervalos de tiempo de un ciclo ejecutivo asegurando al mismo tiempo que la sección crítica de diferentes trabajos no se intercalen.

## Conclusiones

El trabajo debe ser continuado con un proyecto que abarque desarrollos para más de un procesador y con un incremento del grupo de investigación para poder unificar las diferentes investigaciones que están produciendo los alumnos, que divididos en distintos grupos, generan más información que la que el grupo puede concentrar, lo que hace lenta la aparición de nuevas versiones del sistema operativo SODIUM completas, ya que la prioridad de las incorporaciones se realiza de acuerdo a los próximos compromisos asumidos con los nuevos cursantes

## FORMACIÓN DE RECURSOS HUMANOS

Se realizó:

- la primera transferencia de los conocimientos obtenidos a los alumnos que cursan Sistemas Operativos, ya que realizaron el análisis de la arquitectura y las distintos formatos de ejecutables conjuntamente con el análisis del

SODIUM e intervinieron en el desarrollo de los administradores.

- Transferencia de conocimientos a los alumnos de Sistemas de Computación II de la Universidad de La Matanza y a los alumnos de Sistemas Operativos de la Universidad Tecnológica Nacional, Regional Buenos Aires.
- Publicación de los avances en la investigación en dos congresos internacionales.
- Se prevé continuar con las publicaciones en otros congresos internacionales

Se está estudiando:

- el realizar convenios de colaboración con otras universidades nacionales estatales y privadas de las cuales recibimos ofrecimientos de colaboración, con el objetivo de intercambiar conocimientos y ampliar los alcances del sistema.

En esta línea de investigación tenemos:

- dos trabajos de la Maestría en informática en curso.

## BIBLIOGRAFÍA

[ACI08] – Aciti, Claudio - Una Panorámica del Problema de Inversión de Prioridades en Sistemas Operativos de Tiempo Real - INTIA/INCA - Depto. de Computacion y Sistemas - Facultad de Ciencias Exactas Universidad Nacional del Centro de la Provincia de Buenos Aires (7000) - Tandil - Argentina  
[ALF12] – Alfenas Daniel Assis; Danilo Picagli Shibata; Marcos Ribeiro Pereira-Barretto; João José Neto – Sistemas de Markov Adaptativos: Formulação e Plataforma de Desenvolvimento – 6to. Workshop de Tecnología Adaptativa – San Pablo – Brasil 2012.

[ANG98] – Angulo José M y Funke Enrique M – Microprocesadores Avanzados 386 y 486 – 1998 – Editorial Paraninfo – Cuarta Edición.

[ATT08] Attiya Hagit, Leah Epstein, Hadas Shachnai, Tami Tamir. “Transactional Contention Management as a Non-Clairvoyant Scheduling Problem”. Springer Science+Business Media, LLC 2008

[BRE01] – Brey Barry B – Los microprocesadores INTEL – 2001 - Prentice Hall – Quinta Edición

[BRA96] Bradford Nichols, Dick Buttlar & Jacqueline Proulx Farrell, “Pthreads Programming A POSIX Standard for Better Multiprocessing”, Addison-Wesley, 1996.

[BUR03] Burns, A. “Sistemas de tiempo real y lenguajes de programación”, Addison Wesley 2003.

[CAS07] - Casas Nicanor, Martín Cortina, Graciela De Luca, “Desarrollo de un sistema operativo

didáctico". *XIII Congreso Argentino de Ciencia de la Computación*. Chaco, Argentina, 2007.

[CAS08] - Casas Nicanor, Graciela De Luca, Martín Cortina, Gerardo Puyo, Waldo Valiente, "Implementación de distintos tipos de memoria en un sistema operativo didáctico". *XIV Congreso Argentino de Ciencia de la Computación*. La Rioja, Argentina, 2008.

[CAS11] - Casas Nicanor, Graciela De Luca, Sergio Martín, Gerardo Puyo, Waldo Valiente, "Gestión de energía en el sistema operativo didáctico utilizando el modelo APM". *XIII Workshop de Investigadores en Ciencias de la Computación*. Santa Fe, Argentina, 2011.

[EDM08] Edmonds Feff. "How to think about Algorithms". Cambridge University Press. First Edition. 2008.

[GUY07] Guy Even, Magnús M. Halldórsson, Lotem Kaplan, Dana Ron, "Scheduling with conflicts: Online and Offline Algorithms"

[HOL97] Hollstein Thomas, Andreas Kirschbaum y Manfred Glesner. "A prototyping environment for fuzzy controllers", pp:482-490, Lecture Notes in Computer Science 1304, Ed. Springer, FPL, London, Sep 1997.

[KNU68] Knut Donald E. "The Art of computer programming". Standford University. Addison-Wesley Publishing Company. First Edition. 1968

[KOR09] Korhonen Vesa. "TIES246 Real-time systems". JAMK University of Applied Sciences, School of Business Administration, Mankola. Edition 2009 GSM 0400 451 752

[LIU00] Liu, Jane W. S. "Real-Time Systems", Prentice Hall, 2000.

[MAR12a] - Martín, Sergio - Aplicación de Tecnologías adaptativas en un sistema operativo reconfigurable - Tesis de Maestría en Informática - Universidad Nacional de la Matanza - Diciembre 2012

[MAR12b] - Sergio Martín, Casas Nicanor, Graciela De Luca, J. J. Neto, "Utilización de tecnologías adaptativas para la gestión de la energía de un sistema operativo didáctico". *XIV Workshop de Investigadores en Ciencias de la Computación*. Misiones, Argentina, 2012.

[PAT06] Patchrawat "Patch" Uthaisombut. "Generalization of EDF and LLF: Identifying All Optimal Online Algorithms for Minimizing Maximum Lateness". Springer Science+Business Media, LLC 2007

[RAM94] Ramamritham Krithi, and John a. Stankovic. "Scheduling Algorithms and Operating Systems Support for Real-Time Systems". Proceedings of the IEEE, vol. 82 N° I, 1994

[RIM93] Rimvall, C.M. and Jobling, C.P. "Computer Aided Control System Design". IEEE Control Systems Magazine, volume 13, pag 14, 1993, IEEE.

[SEN02] - Senart Aline, Olivier Charra, Jean-Bernard Stefani, "Developing dynamically reconfigurable operating system kernels with the THINK component architecture". *Workshop on Engineering Context-aware Object-Oriented Systems and Environments*. Washington, Estados Unidos, 2002.

[SIL12] - Silberschatz Abraham, Peter Galvin, Greg Gagne - Operating System Concepts - Jhon Wiley & Sons - New Jersey, Estados Unidos, 2012 - Octava Edición

[STO81] - Stonebraker Michael "Operating system support for database management". *Communications of the ACM*, 24(7) - ACM New York, Estados Unidos, 1981.

[TAN97] - Tanenbaum Andrew S y Woodhull Albert S - Operating systems Design and implementation - 1997 - Prentice Hall - Segunda Edición.

[TCH08] - Tchemra Angela Hum - "Aplicação da Tecnologia Adaptativa em Sistemas de Tomada de Decisão: Uma Abordagem Estratégica na Seleção de Fornecedores" - *Segundo Workshop de Tecnologia Adaptativa*. Escola Politécnica, USP, San Pablo, Brasil, 2008.

[TCH09] - Tchemra Angela Hum - "Tabela de decisão adaptativa na tomada de decisões multicritério". - *Tesis de doctorado*. Escola Politécnica, USP, San Pablo, Brasil, 2009.

[TCH10] - Tchemra, Angela Hum - Adaptatividade na tomada de decisão multicritério - 4to Workshop de Tecnología Adaptativa - Escola Politécnica, USP, Sao Pablo, Brasil 2010

[TIN98] - Tinetti Fernando G y De Giusti Armando - "Procesamiento Paralelo Conceptos de Arquitecturas y Algoritmos.

[VEI96] - Veitch A. C.; N.C. Hutchinson - "KEA" - A dynamically extensible and configurable Operating System Kernel - 3<sup>rd</sup>. International Conference on Configurable Distributed System - Vancouver - Canada, 1996

[VEI98] - Veitch A. C. - A dynamically reconfigurable and extensible Operating System - Tesis de doctorado, Department of Computer Science, The University of British Columbia, Canada, 1998

[WOJ07] Wojciech Jawor, Marek Chrobak, Christoph Dürr. "Competitive Analysis of Scheduling Algorithms for Aggregated Links". 2005. Springer Science+Business Media, LLC 2007

# Incorporar actividades virtuales en educación superior: Algoritmo de Segmentación de docentes según sus competencias

Lucia Rosario Malbernat

Departamento de Sistemas, Universidad CAECE, Subsede Mar del Plata  
Gascón 2464, Mar del Plata, Buenos Aires, República Argentina  
+54 233 499-3400

[lmalbernat@ucaecmdp.edu.ar](mailto:lmalbernat@ucaecmdp.edu.ar); [lmalbernat@gmail.com](mailto:lmalbernat@gmail.com)

## Resumen

Para incorporar actividades virtuales en las carreras de grado, los docentes de las universidades deben innovar en sus prácticas docentes y para ello deben desarrollar competencias vinculadas con su preparación y actitud para la virtualidad.

En este trabajo se propone un algoritmo de segmentación, basado en el método del centroide o *k-means*<sup>1</sup>, que agrupa a los docentes según su actitud innovadora tomando en consideración sus respectivas preparaciones y actitud para la virtualidad.

Se toman como variables de entrada la Preparación (índice P) y la Actitud (índice Q), -valores a los que se arriba mediante el cálculo de indicadores diseñados ad hoc- y se segmenta a los docentes identificando grupos o clústeres homogéneos con respecto a su vocación innovadora, clasificándolos en Innovadores, Indiferentes y Refractarios.

La información a la que se arribe con el análisis de los datos que surgen de la segmentación propuesta puede reducir la incertidumbre, por ejemplo, en relación a la toma de decisiones vinculadas con la selección de docentes, la incorporación de actividades online en las materias y la capacitación docente.

---

<sup>1</sup> Se ha tomado la adaptación de Hartigan y Wong (1979, pp. 100-108) del *k-means Clustering Algorithm*, publicado inicialmente por J.B. MacQueen en 1967.

**Palabras clave:** data mining; segmentación; innovación universitaria; TIC, educación virtual

## Contexto

Se toma como caso de estudio a la Universidad CAECE Mar del Plata, República Argentina, en una investigación llevada a cabo sobre carreras de grado en el marco del cursado de la Maestría en Gestión Universitaria en la Universidad Nacional de Mar del Plata, la que diera lugar a la presentación del Informe de Tesis aprobado en noviembre de 2012.

## Introducción

Se han tomado como variables de segmentación a los índices P y Q, cuantificados para cada docente en el marco de la investigación y se definieron las 3 categorías o clústeres (Innovadores, Indiferentes y Refractarios), identificadas a los fines del agrupamiento como A, B y C respectivamente, pues los profesores pueden clasificarse en, al menos, tres categorías [8], quienes generalmente tienen una actitud positiva hacia el uso de las TIC, alientan a sus estudiantes a adquirir conocimientos computacionales y por lo tanto aumentan los estándares de la enseñanza y el aprendizaje en todo el sistema, quienes asumen una posición neutral con relación al uso de las TIC en la educación y quienes tienen actitudes negativas explícitas hacia todas las nuevas tecnologías.

El modelo matemático diseñado ad hoc para cuantificar la preparación utiliza los indicadores nivel de uso de TIC, formación y experiencia en educación virtual y dominio de herramientas informáticas mientras que, para calcular la actitud para la virtualización, se entendieron necesarios los indicadores nivel de interés en el uso de TIC, interés en formación virtual, valoración del vínculo con las TIC y valoración a la educación virtual<sup>2</sup>.

La metodología de segmentación descrita en este trabajo está basada en el uso de heurísticas que proporcionan una solución aproximada que se pretende buena para esta situación, que puede encontrarse en tiempo y a costo razonables, que mejorará el proceso de toma de decisiones reduciendo el nivel de incertidumbre. Es una técnica estadística bivariada, propia del Data Mining, cuya finalidad es segmentar, dividir un conjunto de elementos en grupos de modo que las características de sus elementos sean muy similares entre sí, con fuerte cohesión interna y sean disímiles intragrupos.

Dado que cada segmento debía agrupar docentes con características similares fue necesario elegir una medida para evaluar diferencias y similitudes. Una forma de medir la similitud es calcular la distancia entre pares de docentes. Por eso se tomaron los indicadores P (Preparación) y Q (Actitud) calculados a partir las elecciones hechas por los docentes al responder el cuestionario diseñado ad hoc. Una distancia reducida implicará mayor similitud que una distancia más amplia.

A partir del análisis del contexto de segmentación y de las características del caso, se prefirió utilizar un método no jerárquico, cuyo algoritmo particiona a partir de un elemento central de cada clúster o segmento, capaz de conglomerar a los restantes elementos del grupo a partir de mínimas distancias, denominado método de centroide

o *k-means*, dónde *k* es un parámetro que define el número de elementos centrales o centroides (medias representativas de cada segmento) determinado por la cantidad de grupos o clústeres en que se desea segmentar (*k* coincide con el número de segmentos).

El objetivo de este método no es encontrar un grupo único y definitivo, sino ayudar a que el investigador obtenga una comprensión cualitativa y cuantitativa de los datos de modo de poder obtener grupos razonablemente similares [6].

En algunos contextos de segmentación se cuenta con datos de entrenamiento para diseñar el modelo, los cuales presentan un valor para la variable objetivo, es decir, los elementos a clasificar, ya están clasificados [9]. Para estos contextos, son apropiados los sistemas de clasificación supervisados que proponen el diseño de modelos a partir de los datos de entrenamiento. Para el caso de estudio, por el contrario, es apropiada una clasificación no supervisada.

Por otra parte, el número de clústeres incluidos en la segmentación puede ser o bien desconocido, o bien, conocido o dado por parámetro. Los métodos propuestos por muchos investigadores asumen esta última situación contextual [6], [2], [1], [3], [5], coincidente con el caso de estudio, en el que se han seleccionado 3 grupos.

Algunos autores [6], [1], [3], [10] proponen la elección al azar de los centros iniciales y otros proponen puntos iniciales depurados [2], [4], tal como ocurre en la presente propuesta en la que se conoce qué características se consideran buenas para los resultados pues se está en presencia de un agrupamiento con información externa [9].

Así, tratándose de un agrupamiento para el que se conocen de antemano las características de cada clúster y el rango de valores que pueden tomar las variables -datos calculados a partir del modelo matemático diseñado *ad hoc* que no presentarán valores extremos (*outliers*) que

---

<sup>2</sup> Ver Informe de Tesis “Innovación en educación universitaria: Factibilidad de incorporar actividades virtuales según las competencias docentes”, 2012.



podrían dispersar los objetos del clúster-, se seleccionaron de manera sistemática los centroides iniciales, tomando 3 puntos equidistantes entre sí y de los límite superior e inferior de valores válidos.

## Desarrollo

El algoritmo “*k-means*” encuentra una categorización que representa un valor óptimo según el criterio elegido [2], asignando a cada elemento el clúster del centroide más próximo siguiendo el procedimiento que se describe a continuación:

- Seleccionar  $k$  clúster iniciales  $\in \delta$ , conjunto de clústeres. En el caso de estudio,  $k = 3$  y  $\delta = \{A, B, C\}$
- Identificar casos (elementos) con valores centrales para definirlos como centroides iniciales de cada segmento. Los centroides iniciales, en el caso de estudio han sido definidos como  $A_{(0)}$ ,  $B_{(0)}$  y  $C_{(0)}$ .
- Repetir los siguientes pasos hasta que no se produzcan cambios significativos y no existan elementos equidistantes a 2 o más centroides.
  - Calcular las distancias  $Z$  de cada elemento a los 3 centroides iniciales.
  - Clasificar a cada elemento en el grupo del centroide más cercano (con menor valor de distancia).
  - Re-calcular los clúster iniciales promediando las variables de segmentación de cada clúster, es decir, obteniendo las medias de cada agrupación.
  - Volver a clasificar los elementos asignándolo al clúster del centroide más cercano.

Formalmente, se puede definir el modelo de segmentación que se propone de la siguiente manera para  $\omega$ , conjunto de docentes: Sea  $k$ , cantidad de segmentos en los que se ha decidido clasificar a los elementos  $w \in \omega$  y  $\delta$ , conjunto de clústeres,  $S_i \subseteq \omega$ , de la forma  $\delta = \{S_1, \dots, S_k\}$ , se cumplen las siguientes condiciones:

- a.  $S_i \cap S_i' = \emptyset$  con  $i \neq i'$ .
- b.  $\bigcup_{i=1}^k S_i = \omega$

La primera condición establece que, dados dos segmentos  $S_i, S_i'$ , no pueden tener elementos comunes (un docente no puede estar asignado a más de un grupo) y la segunda, que la unión de todos los segmentos  $S_i$  permite obtener al conjunto  $\omega$  de docentes y que, por lo tanto, todo elemento  $w \in \omega$  debe ser asignado a un grupo (todo docente debe ser situado en un segmento).

En consecuencia, se verifica que  $J$ , cantidad de elementos de  $\omega$  coincide con la sumatoria de los  $j_i$  cantidades de elementos de los segmentos  $S_i$ .

Ecuación 1 – Cálculo de  $J$ , cantidad de elementos de  $\omega$

$$\sum_{i=1}^k j_i = J;$$

con  $J$ , cantidad de elementos de  $\omega$ ,  $k$  cantidad de segmentos y  $j_i$  cantidad de elementos del segmento  $S_i$

El método debe encontrar una  $k$ -partición  $\delta = \{S_1, S_2, S_3\}$ , dónde se maximice la similitud de los  $j_i$  elementos de una partición con respecto a los índices  $P$  y  $Q$  calculados para cada docente.

La maximización de la similitud de los elementos se ha logrado obteniendo las mínimas distancias Euclídeas<sup>3</sup> al cuadrado, es decir, mediante la suma de los cuadrados de las diferencias de los índices de cada elemento a clasificar y de los centroides definidos. Dicha distancia, expresada como  $Z(x_w, y_l)$ , con  $x_w$  un par ordenado  $(P_x, Q_x)$  que representa al elemento  $w$  a clasificar, el cual describe la preparación y actitud del docente e  $y_l$  el par ordenado  $(P_y, Q_y)$  que representa al Centroide  $l$  de un

<sup>3</sup> La distancia Euclídea tradicional calcula la longitud de la recta que une puntos en el espacio euclídeo:  $d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$

segmento  $S_i$ , es calculada con la siguiente función de distancia de  $x_w$  a  $y_i$ :

$$\text{Ecuación 2 – Función de distancia } Z(x_w, y_i) \\ Z(x_w, y_i) = |P_x - P_y|^2 + |Q_x - Q_y|^2.$$

con  $P_x$  valor asignado al índice Preparación del elemento (docente) a clasificar,  $P_y$  el valor asignado al índice Preparación del centroide respecto del cual se va a calcular la distancia, y  $Q_x, Q_y$ , valores equivalentes correspondientes al índice Actitud.

Se cumplen para la función  $Z(x_w, y_i)$  las siguientes propiedades que generalizan en geometría la noción de distancia entre 2 puntos [1]:

- $Z(x_w, y_i) \geq 0$
- $\forall w, Z(x_w, x_w) = 0$ , la distancia entre un elemento y sí mismo es cero;
- $Z(x_w, y_i) = Z(y_i, x_w)$ , la distancia es simétrica;
- $Z(x_w, y_i) \leq Z(x_w, x_n) + Z(x_n, y_i)$ , la distancia verifica la propiedad triangular.

La complejidad computacional del algoritmo *K-means* propuesto es lineal y, por lo tanto, eficiente. Se puede definir como  $O(2Jki)$  con  $J$  cantidad de docentes,  $k$  cantidad de segmentos e  $i$ , número de iteraciones; el 2 representa la cantidad de variables sobre las que se calcula la distancia  $Z(x_w, y_i)$ .

Sea el centroide de un clúster un elemento de la forma  $Y_i = (P_y; Q_y)$ , en el caso de estudio se han tomado para los grupos A, B, y C respectivamente, los siguientes centroides iniciales:

$$A_{(0)} = (7,5; 7,5)$$

$$B_{(0)} = (5, 5)$$

$$C_{(0)} = (2,5; 2,5)$$

Aplicando la función  $Z$ , la distancia de un elemento a cada centroide se calculó de la siguiente manera, donde  $Q_x$  representa el valor  $Q$  (Actitud) del docente  $x$  y  $P_x$ , a su valor  $P$  (Preparación):

$$Z(x_w; A_{(0)}) = |P_x - 7,5|^2 + |(Q_x - 7,5)^2$$

$$Z(x_w; B_{(0)}) = |P_x - 5|^2 + |(Q_x - 5)^2$$

$$Z(x_w; C_{(0)}) = |P_x - 2,5|^2 + |(Q_x - 2,5)^2$$

Obtenidas las distancias de cada docente a cada centroide, -representado por su par ordenado  $(P_x; Q_x)$ -, se clasificó al docente asignándole la categoría más cercana (con menor valor de distancia).

El algoritmo básico *K-Means* propone [5], [4] calcular las medias de las distancias de los elementos del clúster y obtener así nuevos puntos centrales refinados.

Con los nuevos pares ordenados  $A_{(1)}$ ,  $B_{(1)}$  y  $C_{(1)}$  se debe calcular nuevamente la asignación de categoría de cada caso provisionalmente clasificado. En consecuencia, se redefine al centroide  $l$  del segmento  $S_i$  que contiene  $j_i$  elementos como el promedio de las distancias de cada elemento del segmento al centroide  $l$ :

Ecuación 3 – Re-cálculo del Centroide  $l$  de Segmento  $S_i$

$$l = \frac{\sum Z(x_w, y_i)}{j_i}.$$

Con  $x_w$  cada uno de los pares ordenados  $(P_x, Q_x)$  que representan elementos  $w \in S_i$ ,  $y_i$  par ordenado que representa al centroide  $l$  que se re-calcula.

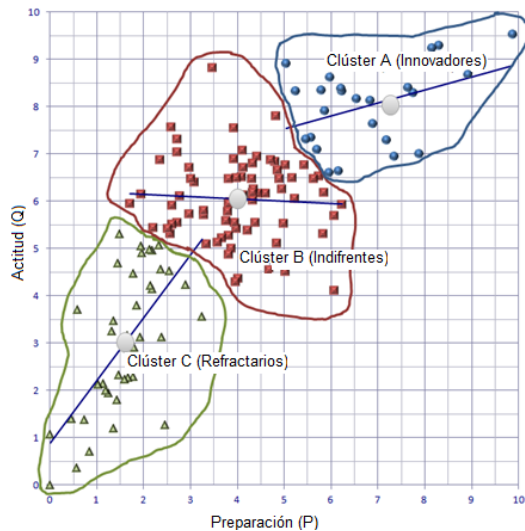
Este proceso de re-calcular los centroides tomando el promedio de las distancias de los puntos del segmento, re-calcular las distancias de los elementos y reasignar los elementos a un grupo según la distancia del elemento al centroide, se debe repetir hasta que no se produzcan clasificaciones dudosas y se puedan dar por clasificados a todos los docentes.

## Resultados y objetivos

Surge de la aplicación del algoritmo que el 17,39% del total de la muestra, -24 docentes-, fue incluido en el clúster de los innovadores, la amplia mayoría del 53,62 % cayó en el segmento de Indiferentes y el 28,99% en el de Refractarios.

El Gráfico 1 muestra la clasificación final de cada sujeto de la muestra según los valores del par ordenado (P; Q) que lo que califican según lo indicado precedentemente y lo ubican en uno de los 3 segmentos definidos.

Gráfico 1- Segmentación docente



Con la información generada a partir de la aplicación del algoritmo no sólo se podrá reducir la incertidumbre al momento de diseñar un plan de capacitación docente. También se podrá observar la situación de cada carrera en relación a la factibilidad de incorporar actividades virtuales por contar ya con docentes preparados y con actitud positiva para hacerlo, pudiéndose en consecuencia, mejorar el proceso de toma de decisiones.

## Formación de Recursos humanos

En noviembre de 2012, Lucía Rosario Malbernat obtuvo el título de Magister en Gestión Universitaria que expide la Universidad Nacional de Mar del Plata, presentando en el Informe de Tesis el trabajo desarrollado en esta línea de Investigación, bajo la dirección del Ph D.

Nicolás Dámaso Patetta.

## Referencias

- [1] M. Berry & G. Linoff, G Data Mining Techniques: for marketing, sales, and customer relationship management (2a ed.) USA: Wiley Publishing, Inc, 2004
- [2] P.S. Bradley & U.M. Fayyad Refining initial points for k-means clustering. In J. Shavlik, editor, Proceedings of the Fifteenth International Conference on Machine Learning (ICML '98), San Francisco, CA, 1998
- [3] C. Garcia Cambronero, I. Gomez Moreno Algoritmos de aprendizaje: KNN & Kmeans. Universidad Carlos III de Madrid, 2009. Recuperado de: [www.it.uc3m.es/jvillena/irc/practicas/08-09/06.pdf](http://www.it.uc3m.es/jvillena/irc/practicas/08-09/06.pdf)
- [4] J. Hartigan & A. Wong A k-means clustering algorithm. Journal of the Royal Statistical Society, Series C (Applied Statistics), Vol. 28, No. 1, 1979. Recuperado de: <http://www.jstor.org/stable/2346830>.
- [5] D. Huerta Muñoz Diseño de Planes eficientes para la segmentación de clientes con múltiples atributos. Tesis de Maestría de la Universidad Autónoma de Nuevo León. Facultad de Ingeniería Mecánica y Eléctrica. México, 2009.
- [6] J.B. MacQueen Some Methods for classification and Analysis of Multivariate Observations, Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability", Berkeley, USA: University of California Press, 1967.
- [7] L.R. Malbernat Innovación en educación universitaria: Factibilidad de incorporar actividades virtuales según las competencias docentes. Tesis de Maestría de Universidad Nacional de Mar del Plata. Facultad de Ciencias Económicas y Sociales. Argentina, 2012.
- [8] UNESCO Las tecnologías de la información y la comunicación en la formación docente. Guía de planificación. París: Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura, 2004 Recuperado de: <http://unesdoc.unesco.org/images/0012/001295/129533s.pdf>
- [9] S. Vega Pons Combinación de resultados de Clasificadores no supervisados. Tesis de doctorado. Rep. Téc. Reconocimiento de Patrones. Serie Azul. Cuba: Centro de Aplicaciones de Tecnologías de Avanzada, 2011.
- [10] E. Yolis, P. Britos, G. Perichisky & R. García-Martínez Algoritmos Genéticos Aplicados a la Categorización Automática de Documentos. Revista Electrónica de sistemas de Información. ISSN 1677-3071 Doi:10.5329/RESI, 2 (2), 2009. Recuperado de: <http://revistas.facecla.com.br/index.php/reinfo/articulo/view/133/27>.

# Ingeniería de Software para clasificar patrones cognitivo conductuales

Marcelo Fabio Roldán<sup>1</sup>, Germán Montejano<sup>2</sup>, Ana Funes<sup>2</sup>, Daniel Riesco<sup>2</sup>

<sup>1</sup>Departamento de Ciencias Exactas, Físicas y Naturales –  
Universidad Nacional de La Rioja  
La Rioja – Argentina

marcelofabio01@yahoo.com.ar

<sup>2</sup>Universidad Nacional de San Luis, Ejército de los Andes 950  
5700 San Luis, Argentina  
{gmonte, afunes, driesco}@unsl.edu.ar

## Resumen

En la presente línea de investigación nos proponemos utilizar el Adaptive Business Intelligence (ABI) como herramienta de extracción de conocimiento en bases de datos, para clasificar probabilísticamente y con un grado de error aceptable, aquellas causalidades del fracaso de los aspirantes a sub oficiales de policía que ingresan en la escuela de la fuerza. A partir de allí, generar aplicaciones informáticas que permitan evaluar al futuro aspirante y determinar predictivamente su factibilidad de éxito en la escuela.

Esto es posible, en concordancia con la puesta en práctica de una metodología de reciente creación [RFM12, RDF13], la que facilita el desarrollo de aplicaciones basadas en Business Intelligence, favorece el uso de métodos de predicción y técnicas de optimización.

Este análisis de datos, para encontrar los patrones que determinan las causalidades de las bajas, lo realizaremos optimizando aquello que se denomina Extracción de Conocimiento de Bases de Datos (KDD), mediante una tecnología actualizada a través de la metodología que la implementa de manera simplificada y práctica. Estas causalidades principales, por las cuales los aspirantes no culminan su formación, se encontrarán de una manera probabilística.

## Contexto

El presente trabajo de investigación se encuentra enmarcado en una colaboración entre investigadores del Proyecto de Incentivos “Ingeniería de Software para clasificar patrones cognitivo conductuales. Clasificación taxonómica predictiva y su impacto en la graduación de los aspirantes a la carrera de sub oficiales de policías”, de la Universidad Nacional de La Rioja e investigadores del Proyecto de Incentivos “Ingeniería de Software: Conceptos, Métodos y Herramientas en un contexto de Ingeniería de Software en Evolución”, de la Universidad Nacional de San Luis.

Este trabajo de investigación surge de la aplicación de la metodología presentada en su tesis de Maestría en “Ingeniería de Software” en la Universidad Nacional de San Luis por uno de los autores de este trabajo [RFM12, RDF13].

**Palabras clave:** Data Mining, Adaptive Business Intelligence, Cognitivo Conductual, Psicometría.

## Introducción

El Business Intelligence se nutre de grandes volúmenes de datos para permitir que los usuarios y aplicaciones accedan a ellos y los usen para realizar toma de decisiones



[WAT07] [AYA06] [REI00]. Es así que metodologías como CRISP-DM [CHA00] [GUT07] [AZE08], SEMMA [SAS11] [TAA11] y, en general, el KDD Process, nos introducen al proceso utilizado para alcanzar tal objetivo por medio del uso de técnicas de Minería de Datos. En particular, la metodología ABI a seguir [RFM12, RDF13], desde un punto de vista integrador y a partir de metodologías de data mining comunes, propone una serie de pasos a seguir, necesarios para alcanzar los resultados esperados. Su aplicación para el análisis prospectivo, avanza mucho más allá del razonamiento basado en eventos pasados que proveen las herramientas típicas de los sistemas de soporte a las decisiones. El conocimiento obtenido se potencia al aplicar esta metodología, pudiendo responder preguntas de difícil resolución o que demandan excesivo tiempo.

Con esta tecnología se pretende encontrar las causalidades para su aplicación predictiva probabilística en nuevos casos. Como lo indica Deshpande [DES10] el data mining predice tendencias futuras y eventos, ayuda a las organizaciones a la toma de decisiones proactivas basadas en conocimiento.

Teniendo en cuenta lo dicho anteriormente, creemos que los algoritmos de minería de datos se pueden usar para reducir los costos en la academia de policía de manera significativa.

Este uso de minería de datos en la determinación de perfiles psicológicos adecuados, es una aplicación con gran potencial, ya que la información psicológica es compleja y difícil de analizar.

## Resultados y Objetivos

Este trabajo de investigación tiene como objetivo principal aplicar una metodología, basada en las estructuras conceptuales de Adaptive Business Intelligence a un caso particular que sirva de referencia para su

aplicación posterior a otros ámbitos de las fuerzas de seguridad.

Esperamos que mediante el uso y las aplicaciones de Adaptive Business Intelligence en contextos predictivos, nos sea posible determinar probabilísticamente y con un grado de error aceptable, aquellas causalidades del fracaso de los aspirantes a policías que ingresan en la escuela de la fuerza tomando como premisa los patrones cognitivo conductuales y su clasificación taxonómica.

El trabajo lo llevaremos a cabo a partir de casos de estudio, y de bibliografía de referencia que permitan facilitar el uso de métodos de predicción.

Durante el proyecto se prevé utilizar las herramientas de Inteligencia Artificial avanzada siguiendo los lineamientos establecidos en la metodología mencionada, para encontrar respuestas a una problemática que se presenta en la instancia preliminar de formación de la policía en su escuela de sub-oficiales. Este problema se manifiesta como fracaso en la continuidad de los estudios que los policías deben aprobar para alcanzar la condición de tales. Los datos relacionados con el problema nos permitirán constatar que el fracaso de los estudios en las escuelas de policías, es un fenómeno con múltiples orígenes, tanto sociales, culturales como del perfil psicológico de los aspirantes o de sus aspectos académicos. El resultado de la clasificación resulta en dos clases: "Apto" y "No Apto", concordando con la evaluación que el profesional del área de Psicología Organizacional determina de manera analítica en cada ficha de aspirante.

Ante este escenario, la aplicación de la nueva metodología ABI, facilitará la producción de conocimiento a partir del aprendizaje y del descubrimiento.

Desarrollaremos durante el proyecto, las actividades necesarias para normalizar los datos de tests psicométricos, pertenecientes a historiales existentes de los policías aspirantes a la academia de sub oficiales de

policía de La Rioja. Posteriormente, sobre estos casos aplicaremos la Metodología Adaptive Business Intelligence en contextos Predictivos [RFM12, RDF13]. A través de la aplicación de las tecnológicas informáticas analizaremos el problema particular del fracaso de los ingresantes a la escuela de sub oficiales buscando encontrar patrones asociados a la misma.

Para el desarrollo de la aplicación predictiva, en este caso, trabajaremos con datos de aspirantes de la escuela de policía de La Rioja en el período que va del año 2005 al año 2010. Estos datos corresponden a personas reales, de las cuales se preserva su identidad, conservados en un archivo profesional cuyo acceso está restringido a estos estudios.

En una etapa posterior, correlacionaremos los resultados arrojados por los instrumentos de evaluación psicométrica con los perfiles de aptitud de los sub oficiales de policía. Estos resultados nos servirán asimismo para analizar la calidad de predicción de los modelos generados [PAL05].

Buscamos de esta forma facilitar el desarrollo informático y, mediante la aplicación de esta metodología en un trabajo interdisciplinario, demostrar su eficacia en un marco orientado a herramientas de extracción de conocimiento, usando problemas reales como eje ilustrativo de su aplicación.

En relación a los aspectos psicométricos sobre los cuales se han realizados las mediciones a los aspirantes – sujetos de las pruebas que son fuente de información para el proyecto- existe una amplia bibliografía al respecto, cuyo basamento científico está ampliamente demostrado [CUE93].

Se conoce que el índice de bajas figura como indicador de calidad y tiene una gran incidencia en el costo operativo del funcionamiento de las fuerzas policiales, ya que de la masa inicial que ingresa a las filas como aspirantes, un reducido porcentaje logra alcanzar la finalización exitosamente;

de allí la importancia de contar con indicadores para una detección precoz.

## **Líneas de Investigación y Desarrollo**

La línea de investigación aquí presentada, la cual está siendo llevada adelante entre investigadores del proyecto de incentivos de la Universidad Nacional de La Rioja, “Ingeniería de Software para clasificar patrones cognitivo conductuales. Clasificación taxonómica predictiva y su impacto en la graduación de los aspirantes a la carrera de sub oficiales de policías.” e investigadores del Proyecto de Incentivos “Ingeniería de Software: Conceptos, Métodos y Herramientas en un contexto de Ingeniería de Software en Evolución”, de la Universidad Nacional de San Luis, se trata de la continuación de una colaboración existente entre ambas instituciones. Dicha línea de investigación surge de la aplicación de la metodología ABI [RFM12, RDF13], la cual fue desarrollada como tesis de Maestría en Ingeniería de Software de la Universidad Nacional de San Luis por uno de los autores de este trabajo.

## **Formación de Recursos Humanos**

El presente trabajo surge como una de las tantas aplicaciones posibles de la metodología desarrollada en un trabajo de tesis de posgrado (Maestría en Ingeniería de Software de la Universidad, Nacional de San Luis). La misma ha sido recientemente defendida, y ha facilitado el desarrollo de proyectos de investigación entre investigadores de la Universidad Nacional de la Rioja y de la Universidad Nacional de San Luis, bajo un acuerdo entre ambas universidades.

## Referencias

[ARA11] J. G. Arancibia. Metodología para el Desarrollo de Proyectos en Minería de Datos CRISP-DM.

URL: <http://yoshibauco.wordpress.com/> - Último acceso: 2011

[ASE04] V. V. Asencios. Data Mining y el Descubrimiento del Conocimiento. Revista de la Facultad de Ingeniería Industrial. Vol (7). 2004

[AYA06] Alejandro Peña Ayala. Inteligencia de Negocios: Una Propuesta para su desarrollo en las organizaciones. Instituto Politécnico Nacional. Dirección de Publicaciones. 2006

[AZE08] A. Azevedo. M. F. Santos. KDD, SEMMA y CRISP-DM: A Parallel Overview. IADIS European Conference Data Mining 2008 (part of MCCSIS 2008)

[CHA00] Pete Chapman, Julian Clinton, Randy Kerber, Thomas Khabaza, Thomas Reinartz, Colin Shearer and Rüdiger Wirth. CRISP-DM 1.0 Step-by-step data mining guide, NCR Systems Engineering Copenhagen (USA and Denmark), DaimlerChrysler AG (Germany), SPSS Inc. (USA) and OHRA Verzekeringen en Bank Groep B.V (The Netherlands). Copyright © 1999, 2000

[CIO07] K.J.Cios,W.Pedrycz, R.W.Swiniarski, L.A.Kurgan. Data Mining. A Knowledge Discovery Approach. Springer. 2007

[CUE93] Cueto, E. G., Introducción a la psicometría. Editores: Madrid: España, 1993

[DES10] Data Mining System and Applications: A review. S.P.Deshpande. V.M. Thakare. International Journal of Distributed and Parallel systems. Vol 1. No 1. 2010

[GUT07] P-Chapman. J. Clinton. R. Kerber. T. Khabaza. T. Reinartz. C. Shearer. R. Wirth. Guía paso a paso de Minería de Datos.

[HAL09] Mark Hall, E. Frank, G. Holmes, B. Pfahringer. The WEKA Data Mining Software: An Update. Pentaho Corporation. 2009

[IEE07] Watson, H.J.; Wixom, B.H. The Current state of the Business Intelligence. Univ. of Georgia, Athens. 2007. Último acceso 2011 <http://ieeexplore.ieee.org/>

[MIC07] Z. Michalewicz, M. Schmidt, M. Michalewicz, C. Chiriac. Adaptive Business Intelligence. Springer 2007

[MOS03] Larissa T. Moss, Shaku Atre. Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications. Addison Wesley. 2003

[PAL05] Pal, N.R., Jain, L.C. Advanced Techniques in Knowledge Discovery and Data Mining, Springer Verlag. (Eds.) 2005.

[REI00] Joerg Reinschmidt, Allison Francoise. Business Intelligence Certification Guide IBM. International Technical Support Organization. 2000

[RMF12] Roldán, M. F., Montejano, G., Funes, A, Una Metodología para el Desarrollo de Aplicaciones Autoadaptativas basada en Business Intelligence. Aplicación en Medicina. Tesis de Maestría. 2012

[RDF13] Roldán, M. F., Debnath, N., Funes, A, Montejano, G., Riesco, D. A Methodology Based on Business Intelligence for the Development of Predictive Applications in Self-Adapting Environments, a aparecer en proceedings de ICCSEE 2013, Hangzhou,China.

[WAT07] Hugh J. Watson. Barbara H. Wixom. The Current State of Business Intelligence. University of Georgia. University of Virginia. 2007. IEEE Computer Society <http://www.computer.org>

[ZBI06] Adaptive Business Intelligence , Zbigniew Michalewicz, Springer 2006

# Redes Sociales y Motores de Búsqueda

Santiago Ricci<sup>1</sup>, Esteban Feuerstein<sup>2</sup>, Gabriel H. Tolosa<sup>1,2</sup>

sricci.soft@gmail.com; efeurest@dc.uba.ar; tolosoft@unlu.edu.ar

<sup>1</sup>Departamento de Ciencias Básicas, Universidad Nacional de Luján

<sup>2</sup>Departamento de Computación, FCEyN, Universidad de Buenos Aires

## Resumen

Internet puede verse desde hace algunos años como un flujo constante de información. Esto se debe en gran parte al contenido producido en las redes sociales. Dicha cuestión puede apreciarse, por ejemplo, en las plataformas de microblogging y dentro de ellas sobre todo en Twitter. Entre las características más importantes de esta plataforma se encuentran los *Trending Topics*, que son un listado de frases, términos y etiquetas relacionados a los temas emergentes más populares en la red social en determinado momento. Así, teniendo en cuenta las cuestiones previamente mencionadas, cabe preguntarse si el hecho de que cierto tema sea *Trending Topic*, influye de algún modo en el volumen de consultas que recibe un motor de búsqueda sobre dicho tema. Este proyecto propone explorar y analizar dicha cuestión y su aplicación en la mejora de los algoritmos de búsqueda.

**Palabras clave:** redes sociales, motores de búsqueda, trending topics.

## Contexto

Esta línea de investigación se encuentra en el marco del proyecto “Modelos y Algoritmos de Búsqueda + Redes Sociales para Aplicacio-

nes Verticales de Recuperación de Información”, presentado oportunamente [11]. El mismo pertenece al Departamento de Ciencias Básicas de la Universidad Nacional de Luján.

## Introducción

Desde hace ya algunos años, algunos servicios de Internet pueden verse como un flujo de información en tiempo real, lo cual implica agregar una componente temporal a la visión que se tiene de la red [1]. Esto puede apreciarse, por ejemplo, en los servicios de microblogging, los cuales apuntan a una necesidad de comunicación diferente de la de los blogs “tradicionales”: la comunicación rápida. Dado que la longitud de las publicaciones se encuentra fuertemente limitada, se requiere menos tiempo para la generación de contenido, lo cual conlleva a que la frecuencia de publicación sea mucho mayor respecto a la de los blogs tradicionales [4]. Un caso concreto de microblogging es Twitter. Este es un servicio basado en lo que se denominan *tweets*. Un tweet es un mensaje cuya longitud máxima es de 140 caracteres. De acuerdo a datos oficiales de Twitter<sup>1</sup>, sus usuarios generan más de 95 millones de *tweets* por día, lo cual sugiere la validez de la

<sup>1</sup><http://blog.twitter.com/2010/12/to-trend-or-not-to-trend.html>



visión mencionada de Internet.

Los *tweets* poseen una serie de características que son importantes. En primer lugar, pueden contener uno o más símbolos *hashtag* (“#”). Estos son utilizados para etiquetarlos con palabras claves o temas<sup>2</sup>. Por otro lado, en un *tweet* es posible mencionar a otros usuarios. También, es importante destacar a los *Trending Topics* (Tendencias). Esto es un listado de frases, términos y *hashtags* relacionados con los temas más populares en la red social en un determinado momento. En la entrada del blog oficial de Twitter mencionada anteriormente, se afirma que el algoritmo encargado de identificar las tendencias se enfoca principalmente en los temas sobre los que “más se habla” en la inmediatez más que en los temas sobre los que se habló. En otras palabras, intenta identificar las “publicaciones de último momento” y no solamente lo popular; prioriza que el tema sea novedoso sobre la popularidad. Para ello, tiene en cuenta dos factores: el volumen de los términos mencionados en Twitter y su velocidad en el crecimiento. Por defecto, las tendencias se determinan de forma personalizada para cada usuario en base a los usuarios que sigue y a su localización. Sin embargo, un usuario puede optar por obtener *Trendings Topics* de forma no personalizada, en base a su región geográfica.

Sobre el escenario de Internet también se encuentran los motores de búsqueda, herramientas indispensables para el acceso al contenido en la red. Teniendo en cuenta además, que los datos e información en la web son cada vez más ricos y complejos, se hace necesario disponer de los mismos en tiempo y forma. En consecuencia, la búsqueda se ha convertido en un proceso clave y central en la web. Esto se traduce en que los motores de búsqueda deban responder millones de consultas (*queries*) por día, lo cual implica

<sup>2</sup><https://support.twitter.com/articles/49309-what-are-hashtags-symbols#>

eficiencia y efectividad para poder otorgar a los usuarios respuestas relevantes lo más rápido posible [12].

Existen diversos estudios acerca de Twitter. En [4] se estudia al servicio desde el punto de vista estructural y del contenido. Una publicación posterior [6] amplía dicha caracterización. En [7] se exponen las características topológicas del servicio de microblogging y se sugiere que gran parte (85%) de los *Trending Topics* están relacionados con las noticias del momento o con aquellas que son persistentes en el tiempo y que una gran porción (31%) de los mismos dura aproximadamente un día. Además se concluye que los usuarios tienden a “hablar” sobre dichas noticias. En [2], también estudian los *Trending Topics* y se afirma que aquellos con grandes duraciones están caracterizados por la naturaleza “resonante” del contenido de sus *tweets* asociados, el cual proviene, generalmente, de los medios de comunicación tradicionales. De este modo, Twitter se comporta como un amplificador selectivo del contenido generado por los medios tradicionales mediante cadenas de *retweets*. En [10] se compara la tarea de búsqueda de los usuarios en Twitter respecto a la misma en los motores de búsqueda y se concluye que los usuarios utilizan Twitter para monitorear un tema y la Web para aprender más acerca del mismo.

También, es conocido el uso de esta red para expresar opiniones acerca de diferentes temas, lo cual se ha traducido en gran cantidad de artículos que plantean diferentes enfoques sobre cómo realizar Minería de Opinión sobre la red social, como por ejemplo, [14] o [9].

Sin embargo, aún no está clara la relación entre los motores de búsqueda y las redes sociales, por lo que en esta línea de investigación se trabaja sobre las siguientes cuestiones:

1. El hecho de que cierto tema sea *Trending Topic*, ¿Influye de algún modo en el volumen de consultas que recibe un motor de

búsqueda sobre dicho tema?

2. ¿Es posible emplear esta información de la red social para “optimizar” el algoritmo de búsqueda, por ejemplo, utilizándolo para refinar las políticas de reemplazo de una memoria caché o el *prefetching* [5] de resultados?
3. De igual manera, ¿Se podría aplicar dicha información en el proceso de desambiguación del sentido de una consulta?

## Líneas de investigación y desarrollo

Las hipótesis presentadas en la sección anterior sugieren las siguientes líneas de investigación:

1. **Caracterización de los Trending Topics:** en esta línea se pretende determinar propiedades de los *Trending Topics* que son útiles en relación con las temáticas propuestas. Una primer problemática es clasificar el contenido que intentan describir los *Trending Topics*, dado que es conocido que no todos los usuarios utilizan la red social con los mismos objetivos [4] y el espectro de temas abarcados por la misma es virtualmente ilimitado. En este sentido, también es importante estudiar los períodos en los que se encuentran activos, sus repeticiones en el tiempo, el porcentaje de usuarios que participan en los mismos, entre otras cuestiones. Aquí también puede enmarcarse el proceso de derivación de consultas desde los *Trending Topics*. Este proceso puede ser determinante en los resultados, dado que no siempre está claro el tema que representa. El *parsing* del mismo puede ser problemático e incluso generar consultas ambiguas.

2. **Caracterización del volumen de tráfico en un motor de búsqueda mediante métodos indirectos:** esta línea pretende intentar establecer el impacto en la carga de trabajo de un conjunto de consultas a un motor de búsqueda. Como esta información es propietaria de los proveedores de los servicios de búsqueda (se los caracteriza como “ambientes no cooperativos” [8]), se requieren métodos “externos” que permitan obtener aproximaciones válidas. Este problema ha sido abordado en el área de Recuperación de Información Distribuida para obtener descripciones de los recursos objetivo (por ejemplo, Query Based Sampling [3]). Esta idea ya ha sido explorada inicialmente en nuestro proyecto (se presenta en la próxima sección).
3. **Algoritmos que exploten los *features* de los *Trending Topics* para la desambiguación del sentido de una consulta:** múltiples elementos se tienen en cuenta a la hora de decidir el ranking final de una consulta para un usuario particular. Se sabe que su historial de búsqueda e información de perfil son indicadores de preferencias temáticas, utilizados por ejemplo, para el proceso de desambiguación. Se considera explorar si la información proveniente de una red social aporta a este proceso.
4. **Modelos de *caching* y *prefetching* que usan la información de los *Trending Topics* para tomar decisiones que mejoren la performance:** estas dos técnicas son ampliamente usadas en el ámbito de los motores de búsqueda y permanentemente se estudian nuevas estrategias. El uso propuesto puede permitir mejoras particulares bajo ciertas circunstancias, por ejemplo, búsquedas en tiempo real.

## Resultados y objetivos

Sobre estas líneas de investigación se ha comenzado a trabajar y se han obtenido algunos resultados prometedores. Se realizó una experiencia de pequeña escala sobre los *Trending Topics* de Twitter para Argentina durante una semana. Por otro lado, se observaron las tendencias de evolución de algunas consultas mediante Google Trends y se aplicaron tres métricas que apuntan a cuantificar el impacto de la red social en un motor de búsqueda:

1. **Variación porcentual del interés de la consulta derivada para un *Trending Topic* respecto al día anterior:** esta métrica intenta capturar el hecho de que si los *Trending Topics* influyen en el volumen de consultas, entonces debe existir una diferencia significativa en el interés en dicha consulta cuando el tema es *Trending Topic* y cuando no lo es.
2. **Cuantificación del cambio en la tendencia que experimenta cierta consulta derivada cuando el tema se convierte en *Trending Topic*, respecto a su tendencia en los  $n$  días anteriores.**
3. **Algoritmo de detección de picos (*burst detection*):** este algoritmo fue utilizado de acuerdo a [13]. Aquí se intenta demostrar que el día que un tema es *Trending Topic*, registra un pico en la aparición de consultas en el motor de búsqueda.

Los resultados preliminares muestran que aproximadamente el 70 % de las consultas aumentan su volumen de tráfico cuando aparecen como *Trending Topic* en Twitter y en aproximadamente el 60 % de los casos este crecimiento supera el 40 % respecto al día anterior.

## Formación de Recursos Humanos

Esta línea de investigación surgió de una Pasantía Interna Rentada (PIR) de la UNLu que el primer autor realizó en el proyecto en que está inserta (Ver "Contexto") bajo la supervisión de Prof. Gabriel Tolosa. Durante dicha pasantía surgieron varias direcciones futuras, por lo que el primer autor va a desarrollar su Trabajo Final de Licenciatura en Sistemas de Información en esta área y se prevé incorporar un nuevo pasante durante el año en curso.

## Referencias

- [1] Agarwal, S.; Agarwal, S. Social networks as Internet barometers for optimizing content delivery networks. In 3rd International Symposium on Advanced Networks and Telecommunication Systems (ANTS). 2009.
- [2] Asur, S.; Huberman, B. A.; Szabo, G.; Wang, C. Trends in social media: Persistence and decay. In 5th International AAAI Conference on Weblogs and Social Media. 2011.
- [3] Callan, J.; Connel, M. Query-based sampling of text databases. In ACM Transactions on Information Systems, v. 19, n. 2, pp. 97-130. 2001.
- [4] Java, A.; Song, X.; Finin, T.; Tseng B. Why we twitter: understanding microblogging usage and communities. In Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis (WebKDD/SNA-KDD '07), pp. 56-65. 2007
- [5] Jonassen, S.; Barla Cambazoglu B.; Silvestri F. Prefetching query results and its impact on search engines. In Proceedings of

- the 35th international ACM SIGIR conference on Research and development in information retrieval (SIGIR '12), pp. 631-640. 2012.
- [6] Krishnamurthy, B.; Gill, P.; Arlitt M. A few chirps about twitter. In Proceedings of the first workshop on Online social networks (WOSN '08), pp. 19-24. 2008.
- [7] Kwak, H.; Lee, C.; Park, H.; Moon, S. What is Twitter, a social network or a news media?. In Proceedings of the 19th international conference on World wide web (WWW '10), pp. 591-600. 2010.
- [8] Luo S. Federated search of text search engines in uncooperative environments. PhD Thesis. Carnegie Mellon University. 2006.
- [9] Meng X.; Wei, F.; Liu, X.; Zhou, M.; Li, S.; Wang, H. Entity-centric topic-oriented opinion summarization in twitter. In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '12), pp. 379-387. 2012.
- [10] Teevan, J.; Ramage, D.; Ringel Morris, M. #TwitterSearch: a comparison of microblog search and web search. In Proceedings of the fourth ACM international conference on Web search and data mining (WSDM '11), pp. 35-44. 2011.
- [11] Tolosa, G. H.; Bordignon F. Modelos y algoritmos de búsqueda + redes sociales para aplicaciones verticales de recuperación de información. En Proceedings del XIII Workshop de Investigadores en Ciencias de la Computación (WICC 2011), pp. 243-247. 978-950-673-892-1. 2011.
- [12] Tolosa G. H.; Feuerstein E. Mejoras algorítmicas y estructuras de datos para búsquedas altamente eficientes. En Proceedings del XIV Workshop de Investigadores en Ciencias de la Computación (WICC 2012), p. 740-744. 978-950-766-082-5. 2012.
- [13] Vlachos, M.; Meek, C.; Vagena, Z.; Gunopoulos, D. Identifying similarities, periodicities and bursts for online search queries. In Proceedings of the 2004 ACM SIGMOD international conference on Management of data (SIGMOD '04), pp. 131-142. 2004.
- [14] Wang, X.; Wei, F.; Liu, X.; Zhou, M.; Zhang, M. Topic sentiment analysis in twitter: a graph-based hashtag sentiment classification approach. In Proceedings of the 20th ACM international conference on Information and knowledge management (CIKM '11) pp. 1031-1040. 2011.



# Aplicación de técnicas de Minería de Datos al análisis de situación y comportamiento académico de alumnos de la UGD

Karina Eckert, Roberto Suénaga

Universidad Gastón Dachary

Salta 1968, Posadas, Misiones-Argentina. Tel: +54 (0376) - 4438677

karinaeck@gmail.com, rsuenaga@ugd.edu.ar

## Resumen

En el ámbito educativo es evidente la necesidad de disponer de sistemas de gestión que permitan tomar decisiones académicas y elaborar estrategias a partir del conocimiento oportuno, ya que esto no solo incide directamente sobre la funcionalidad de los departamentos académicos, u otras cuestiones internas, sino que también podrían incidir sobre actividades como las evaluaciones y acreditaciones de instituciones y carreras. Entre los problemas más complejos que enfrentan las instituciones de educación podemos mencionar: mejorar la calidad académica, disminuir la deserción y la reprobación, evitar el atraso estudiantil y los bajos índices de eficiencia relacionado con las tasas de graduación. Esto requiere gestionar estrategias y tomar medidas frente a estos acontecimientos; para ello es posible recurrir al proceso denominado Minería de Datos Educativo (MDE), es decir, la aplicación del proceso de Descubrimiento o Extracción de Conocimiento en Bases de Datos (KDD) en ámbito educativo.

En el presente trabajo se describe y expone la aplicación del proceso KDD (por su siglas en inglés), conocido como Minería de Datos (MD) en un entorno educativo, más precisamente a la información académica de la Universidad

Gastón Dachary (UGD). El proceso consiste en una serie de etapas que parten de la selección y captura de los datos, pasando por una serie de actividades relacionadas a la integración, recopilación y el filtrado de los mismos (pre-procesamiento), para luego ser procesados, analizados y evaluados hasta obtener conocimiento adicional. Para ello, es necesario llevar a cabo un proceso iterativo que incluye numerosas consultas de selección a la base de datos, depuración de los datos, utilización de diferentes criterios de representación; también se aplican diferentes técnicas y algoritmos de MD, tanto descriptivas como predictivas.

**Palabras clave:** Descubrimiento de Conocimiento en Bases de Datos (KDD), Minería de Datos Educativos (EDM), Rendimiento Académico, Herramientas de Minería de Datos.

## Contexto

El presente proyecto de investigación se encuentra dentro de una línea de investigación iniciada con la Tesis de Grado titulada *“Explotación de Datos Académicos a través de la Aplicación de Técnicas de Minería de Datos en Weka”*, de la carrera Ingeniería en Informática con Orientación a Sistemas de Información de la Universidad Gastón Dachary (UGD). La tesis derivó en una

postulación y adjudicación de beca de investigación financiado por el Comité Ejecutivo de Desarrollo e Innovación Tecnológica (CEDIT), bajo la denominación: *“Exploración de datos académicos para la determinación de causas de deserción universitaria a través de la aplicación de Técnicas de Minería de Datos”*.

## Introducción

Actualmente la sociedad se encuentra en la denominada era de la información, donde día tras día se incrementa significativamente la cantidad de datos almacenados en diferentes fuentes, estructuras y formatos. Empero, contrariamente a lo se espera, esta expansión de datos no siempre supone un aumento de conocimiento, puesto que procesarlos con los métodos clásicos resulta ser en muchos casos imposible o sumamente tedioso y con resultados superficiales e insatisfactorios. De modo que nos enfrentamos actualmente a la paradoja de que, cuantos más datos están disponibles, menos información se tiene, y algo peor que no tener información disponible es tener mucha y no saber qué hacer realmente con ella. La clave evidentemente está en tener la información adecuada, en el lugar y momento oportuno, y así incrementar la efectividad de toda organización. La idea clave es que los datos contienen más información oculta de la que se ve a simple vista, por lo que hay que “torturarlos hasta que ellos confiesen” [1], que es una explicación informal de la actividad que se realiza mediante la, denominada Minería de Datos (MD).

La Minería de Datos (MD), entre otras técnicas, utiliza Inteligencia Artificial para encontrar patrones y relaciones entre los datos, permitiendo la creación de modelos y representaciones abstractas de la realidad. La MD es una etapa dentro del proceso mayor llamado Descubrimiento de Conocimiento en Base de Datos (KDD), aunque en la

mayoría de los entornos, ambos términos se usan de manera indistinta. El proceso completo de KDD incluye entre otras cosas, la preparación de los datos y la interpretación de los resultados obtenidos, los cuales dan un significado a los patrones identificados por las técnicas y algoritmos de MD. Así el valor real de los datos reside en la información que se puede extraer de ellos, información que ayude a tomar decisiones o mejorar nuestra comprensión de los fenómenos que nos rodean [2]. KDD es el “proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y, en última instancia, comprensibles a partir de los datos”, relevantes y nuevos sobre un fenómeno o actividad mediante algoritmos eficientes, dadas las crecientes órdenes de magnitud en los datos [3]. Las metas son: procesar automáticamente grandes cantidades de datos crudos, identificar los patrones más significativos y relevantes, presentarlos como conocimiento apropiado para satisfacer las demandas del usuario. El conocimiento se obtiene para llevar a cabo acciones, ya sea incorporándolo dentro de un sistema de desempeño o para almacenarlo y reportarlo a los interesados, en este sentido, implica un proceso interactivo e iterativo.

Lo que en verdad hace la MD es reunir las ventajas de varias disciplinas como la estadística, la inteligencia artificial, la computación gráfica, los sistemas de bases de datos y el procesamiento masivo, principalmente usando como materia prima las bases de datos de las organizaciones. La MD como etapa en el proceso de KDD es el “paso consistente en el uso de algoritmos concretos que generan una enumeración de patrones a partir de los datos pre-procesados” [3]. Se coleccionan los datos y se espera que de ellos emerjan hipótesis. Se busca que los datos describan o indiquen por qué son como son. Luego entonces, se valida esa hipótesis inspirada por los datos en

los datos mismos, será numéricamente significativa, pero experimentalmente inválida. De ahí que la MD debe presentar un enfoque exploratorio, y no confirmador. Usar la MD para confirmar las hipótesis formuladas puede ser peligroso, pues se está haciendo una inferencia poco válida [4] [5].

La deserción, el rezago estudiantil y los bajos índices de eficiencia terminal se encuentran entre los problemas más complejos y frecuentes que enfrentan las Instituciones de Educación Superior.

La Minería de Datos Educativa (EDM) es una disciplina emergente, preocupada por el desarrollo de métodos para explorar las características singulares de los datos que provienen de entornos educativos, y utilizar esos métodos para comprender mejor el desempeño de los estudiantes, y las condiciones en las cuales ellos aprenden [6][7][8]. Es el proceso de transformar los datos en bruto, recopilados por los sistemas de enseñanza, en información útil que pueda utilizarse para tomar decisiones y responder preguntas de investigación [8][9][10][11]. La aplicación de la MD en el ámbito de la enseñanza, tiene como objetivo obtener una mejor comprensión del proceso de aprendizaje de los estudiantes y de su participación global en el proceso, orientado a la mejora de la calidad y la eficiencia del sistema educativo [10][12]. A partir de toda la información disponible, las diferentes técnicas de MD pueden ser aplicadas a fin de descubrir conocimiento útil que ayude a mejorar el proceso educativo, siendo este conocimiento muy diverso. Está dirigida a los alumnos, profesores o autoridades académicas, quienes a partir de ella pueden identificar tres tipos de objetivos: pedagógicos (ayuda en el diseño de contenidos didácticos, mejoras en el rendimiento académico de los alumnos), de gestión (optimizar la organización y mantenimiento de infraestructuras educativas, áreas de interés, cursos más solicitados) y

comerciales (permite realizar segmentación del mercado y facilita la captura de alumnos) [10].

## Líneas de investigación y desarrollo

El trabajo de investigación sigue la estructura del Proceso KDD, que consta de cinco fases (Figura 1).

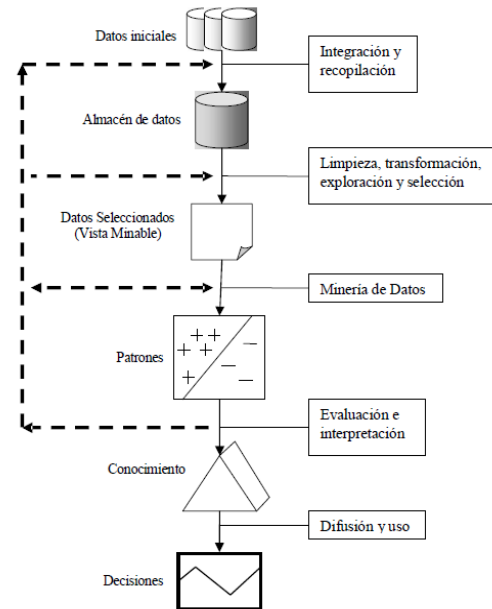


Figura 1: Fases del Proceso de KDD [13]

Durante el desarrollo del proceso de KDD, suele ser necesario interrumpir en algún punto de las fases del proceso y volver a comenzar en alguno de los pasos anteriores, siendo así un proceso *iterativo* e *interactivo* necesario para lograr una alta calidad del conocimiento a descubrir [3].

La fuente de datos proviene de la información académica de la UGD, datos proporcionados al ingreso (personales y antecedente de estudio a la institución) y durante el lapso de sus estudios en la institución; con la debida protección de datos personales, sin identificación de individuos, descartando cualquier información que pueda identificar directa o indirectamente a los alumnos (DNI, Apellidos, Matriculas, etc.), creando una

vista minable con las características (atributos) de las titulaciones seleccionadas, una colección de individuos, en este caso alumnos de la UGD, sobre los cuales se realizó el estudio, a la que se le aplica la fase de Minería de Datos para poder extraer conocimiento útil en lo que se refiere al rendimiento académico del alumnado. Dicha vista minable contiene información de las carreras tanto del departamento de administración como el de informática, correspondiente un periodo de 10 años, que va desde el año 1999 al 2009 respectivamente. La precisión de los datos a utilizar depende en gran medida de la completitud de los mismos.

## Objetivos y Resultados

### Objetivo General

Identificar características y patrones de comportamiento relacionadas con el desempeño académico de los alumnos, a través del proceso de Descubrimiento de Conocimiento en Bases de Datos (KDD) y herramientas de Minería de Datos. Los modelos identificados se propondrán como contribución a la toma de decisiones en el ámbito de la gestión académica.

### Objetivos Específicos

Utilizar herramientas de Minería de Datos para detectar patrones y relaciones entre los datos de la trayectoria académica de los estudiantes.

Identificar cuáles son las variables y áreas que inciden sobre el desempeño, permanencia y graduación de los alumnos universitarios.

Elaborar recomendaciones sobre los posibles usos de los resultados obtenidos y las características de las fuentes de información que sirvan como estrategias innovadoras y apropiadas para la gestión académica.

### Actividades realizadas para el alcance de los objetivos:

Búsqueda y análisis de información referente a los aspectos teóricos relacionados con el proceso de descubrimiento de conocimiento en base de datos (KDD), técnicas de minería de datos (MD), como ser: pre-procesado de los datos, clasificación, segmentación, y asociación; así como los algoritmos disponibles en ellas. Estas actividades estaban precisamente destinadas a identificar y describir el proceso, técnicas y algoritmos de MD dentro del ámbito educativo/académico.

Obtención de datos pertinentes a la investigación proporcionados por la UGD para su análisis. Integración y recopilación de datos: para esta etapa se utilizaron técnicas como la ejecución de instrucción en SQL en la base de datos utilizando los criterios correspondientes a los fines de la investigación, con el fin de generar el almacén de datos con el cual se trabajará en las siguientes etapas.

Conversión de los datos seleccionados: para poder procesar los datos en algunas de las herramientas de minería de datos, se requiere la conversión al formato CSV (archivos separados por comas) en cada una de las variantes y conjuntos de datos que se generen para tal fin.

Ejecución de pruebas con la herramienta de MD Weka: analizando los resultados preliminares obtenidos en el período anterior, se procedió a llevar a cabo una serie de pruebas con los datos seleccionados y acondicionados para observar el comportamiento de las técnicas y los algoritmos de MD de acuerdo a las nuevas series de datos seleccionadas, aplicando distintas variantes y ajustes a los parámetros de ejecución en los distintos algoritmos y adaptaciones de los datos con el fin de obtener variantes en los resultados.

Análisis y elaboración de resultados preliminares sobre las pruebas realizadas. Comprobación de resultados mediante confrontación de diferentes técnicas de MD.



Búsqueda y análisis de información referente a las prestaciones de una serie de herramientas de minería de datos, en busca de las más adecuadas para la investigación.

## Formación de Recursos Humanos

El trabajo corresponde a una tesis de grado de Ingeniería en Informática, que durante su desarrollo recibió una beca de iniciación a la investigación científica, otorgada por el Comité Ejecutivo de Desarrollo e Innovación Tecnológica de la Provincia de Misiones (CEDIT).

## Referencias

- [1] MOLINA, FÉLIX LUIS CARLOS. "Data mining: torturando a los datos hasta que confiesen". [En línea] 2002. <http://www.uoc.edu/web/esp/art/uoc/molina1102/molina1102.html#bibliografia>.
- [2] "Data Mining y el Descubrimiento del Conocimiento". VALCÁRCEL ASENCIOS, VIOLETA. 2, s.l. : Revista de la Facultad de Ingeniería Industrial, 2004, Vol. (7), págs. 83-86. ISSN: 1560-9146 (impreso) / ISSN: 1810-9993 (electrónico).
- [3] FAYYAD, U. M., PIATETSKY SHAPIRO, G., SMYTH, P., UHTURUDSAMY, R., "Advances in Knowledge Discovery & Data Mining", 1ª Edition, Editorial MIT Press, Cumberland, Rhode Island, EE.UU., 1996. ISBN: 9780262560979.
- [4] VERÓNICA S. BOGADO Y MARIANA C. ARRUZAZABALA, "Sistemas Operativos – Descubrimiento de Conocimiento en Base de Datos", Corrientes Argentina: Universidad Nacional del Nordeste - Facultad de Ciencias Exactas, Naturales y Agrimensura (UNNE), Trabajo Monográfico, 2003.
- [5] CARLOS MARIN, "Minería de Datos", <http://mineriadatos.blogspot.com/>.
- [6] RAMASWAMI, M. y BHASKARAN, R. "A Study on Feature Selection Techniques in Educational Data Mining". Journal of Computing, 2009, Vol. Vol.1 , Issue 1.
- [7] INTERNATIONAL WORKING GROUP ON EDUCATIONAL DATA MINING. [En línea] <http://www.educationaldatamining.org/>.
- [8] "Educational Data Mining 2008 - The 1st International Conference on Educational Data Mining". JOAZEIRO DE BAKER, RYAN S., BARNES, TIFFANY y BECK, JOSEPH. Montréal Québec Canada : s.n., 2008. URL: <http://www.educationaldatamining.org/EDM2008/uploads/proc/full%20proceedings.pdf>.
- [9] "Proceedings of Educational Data Mining workshop, held in conjunction with the 8th International Conference on Intelligent Tutoring Systems". HEINER, CECILY, BAKER, RYAN y YACEF, KALINA. Jhongli Taiwan : s.n., 2006.
- [10] INFORMATION, SCIENCE TODAY. "Educational data mining". [En línea] <http://www.infosciencetoday.org/document-management/educational-data-mining.html>.
- [11] "Supplementary Proceedings of the 13th International Conference of Artificial Intelligence in Education". HEINER, CECILY, HEFFERNAN, NEIL y BARNES, TIFFANY. Marina del Rey CA. USA : s.n., 2007.
- [12] VENTURA SOTO, SEBASTIÁN. "Minería de Datos en Sistemas Educativos". [En línea] <http://sci2s.ugr.es/docencia/doctoM6/EducationalDataMining.pdf>.
- [13] HERNÁNDEZ ORALLO, J., RAMÍREZ QUINTANA, M.J. y FERRI RAMÍREZ, C. "Introducción a la Minería de Datos". 1a Edición. Madrid España: Editorial Pearson, 2004. ISBN: 84-205-4091-9.

# GESTIÓN DE DATOS BIOMETRICOS EN BASES DE DATOS OBJETO-RELACIONALES

Graciela Etchart, Carlos Alvez, Marcelo Benedetto

Facultad de Ciencias de la Administración - Universidad Nacional de Entre Ríos

Av. Tavella 1424, Concordia, Entre Ríos - CP 3200

[getchart@fcad.uner.edu.ar](mailto:getchart@fcad.uner.edu.ar), [caralv@fcad.uner.edu.ar](mailto:caralv@fcad.uner.edu.ar), [marben@fcad.uner.edu.ar](mailto:marben@fcad.uner.edu.ar)

## Resumen

La aplicación de sistemas biométricos, particularmente en la administración pública, ha tomado importancia en los últimos años, tanto para controles de accesos físicos o virtuales, documentación, etc. El uso de esta tecnología no alcanza si no se cuenta con bases de datos sobre las cuales se pueda contrastar la información que brindan las aplicaciones basadas en biometría. Para esto, es importante que los organismos estatales que utilicen grandes volúmenes de datos biométricos, puedan consultar la información de manera eficiente e intercambiar la misma con otros organismos asociados. De allí la necesidad de contar con estándares que permitan la interoperabilidad entre sistemas y así facilitar la búsqueda de datos. Para esto, en este proyecto, se propone la gestión de datos biométricos en Bases de Datos Objeto-Relacionales, permitiendo la representación tanto de los datos crudos (imágenes/audios) como los metadatos requeridos por estándares utilizados en la comparación, e indexación para un acceso eficiente a los mismos. También, se crearán modelos de representación de la información necesaria para crear registros de transacciones ANSI-NIST-ITL-1 2011 [1], y de esta manera tener la posibilidad de intercambiar información entre organismos asociados, mediante procesos de exportación e importación.

**Palabras clave:** Bases de datos OR, biometría, organismos públicos.

## Contexto

Este trabajo da continuidad al Proyecto PID 07/G035 “*Identificación de personas mediante Sistemas Biométricos. Estudio de factibilidad y su implementación en organismos estatales*”, [2] [3] [4], cuyo objetivo fue analizar las dificultades en los procedimientos de autenticación de personas en organismos públicos e implementar posibles soluciones a través de la utilización de sistemas biométricos. En el proyecto mencionado, se realizó entre otras cosas, un sistema de reconocimiento de personas mediante el iris. También da continuidad al trabajo realizado en la tesis doctoral de C. Alvez [5] donde se abordó principalmente en modelos de representación de imágenes en Bases de Datos Objeto-Relacionales (BDOR). El principal objetivo de la tesis fue la recuperación de imágenes por similitud, por contenido físico y semántico.

## Introducción

La biometría es una tecnología que hoy en día se encuentra en pleno desarrollo, tanto en el ámbito de la vida cotidiana como en la investigación. Esta rápida evolución se debe a la creciente preocupación por la seguridad y por la vinculación que tiene esta tecnología para garantizar la misma. En cuanto a la vida cotidiana, desde una mirada macro del sector, la Argentina es un país con mucho potencial en la

región en implementar y desarrollar soluciones de biometría tanto para el sector público como privado [6] [7]. La mayor importancia en el uso de esta tecnología radica en su aplicación como mecanismo de control: control de acceso físico, acceso virtual (redes, sistemas, etc.), controles de horarios, entre otros.

Los organismos públicos que emplean esta tecnología necesitan consultar los datos biométricos de manera eficiente e intercambiar la misma con otros organismos asociados. De allí la necesidad de contar con modelos de datos y con estándares que permitan la interoperabilidad entre sistemas y así facilitar la búsqueda de datos.

Uno de los estándares que contempla este tipo de aspectos (entre otros), es el ANSI/NIST ITL 1-2011 [1], norma biométrica publicada en noviembre de 2011, que define cómo trabajar para garantizar la interoperabilidad de datos biométricos entre los distintos sistemas. Este estándar, define el contenido, el formato y las unidades de medida para el intercambio de huellas dactilares, palmares, plantares, faciales, el iris, el ácido desoxirribonucleico (ADN), y otras muestras biométricas y datos forenses que pueden ser utilizados en el proceso de identificación o verificación de una persona. Es el estándar más utilizado por entes estatales, y sobre el mismo se viene trabajando en el grupo de investigación del proyecto PID 07/G035, específicamente en el rasgo de iris (registro 17 del estándar). Existe un borrador para voz (registro 11) [8], que será concluido en 2013, lo cual es importante porque es uno de los rasgos sobre el que se trabajará en el presente proyecto.

Para que las muestras biométricas capturadas por un organismo determinado, puedan ser intercambiadas con otros organismos, el sistema que gestiona los datos biométricos debe soportar registros de transacciones conforme a esta norma, es decir, debe ser capaz de generar y/o utilizar las transacciones que sean morfológicas, sintácticas y semánticamente conformes a los requisitos del estándar. Para esto, el sistema debe contar con los módulos de software que se encarguen tanto de la recepción de estos registros, como también, de la

elaboración de los mismos para ser enviados a organismos asociados [4].

Los registros antes mencionados, se utilizan únicamente para el intercambio de información. Sin embargo, se debe también considerar la generación y el almacenamiento de metadatos necesarios para el reconocimiento, o sea para los procesos de identificación y/o verificación de personas. Como ejemplo de estos metadatos se pueden mencionar: las minucias de una huella dactilar, o el código del iris (*iriscode*) [9] [10], en un sistema de reconocimiento del iris, rasgos acústicos para el reconocimiento de voz, entre otros. O sea, que estos sistemas tienen que tratar, por un lado con: imágenes 2D (iris, huellas dactilares, etc.), imágenes 3D (reconocimiento facial, geometría de la mano, etc.), sonidos (reconocimiento de voz), videos (reconocimiento de gestos o movimientos corporales), y por otro lado, la gestión de metadatos extraídos de estos datos para ser utilizados en el reconocimiento (minucias de huellas, código de iris, etc.) que son utilizados para la comparación. Estos últimos son estructuras complejas, y pueden ocasionar algunos inconvenientes si se trabaja con el modelo de datos relacional, por las limitaciones impuestas por el mismo. Por esto, aquí se ha optado por utilizar la tecnología Objeto-Relacional (OR), como se detallará en la próxima sección [11].

## Líneas de Investigación y Desarrollo

En la actualidad, en el área específica de la biometría, se encuentran abiertas varias líneas de investigación con distintos niveles de desarrollo. Sin embargo, como se trató en la sección anterior, especialmente en organismos gubernamentales, es muy importante considerar el nivel de estandarización alcanzado, sobre todo, en lo referente al formato de intercambio de datos. Por lo tanto, se debe tener particularmente en cuenta esto, para el desarrollo de sistemas biométricos en organismos públicos [3].

Otra línea de investigación vigente está orientada a sistemas que combinan diferentes rasgos biométricos o multimodales. Estos

sistemas son más precisos y seguros dado que superan algunas limitaciones de los sistemas que utilizan un único rasgo biométrico. Las limitaciones de un rasgo pueden ser: no universalidad, ruidos en los datos, suplantación de identidad, entre otros [12].

Sin embargo, una de las líneas más importantes, y la que nos interesa aquí, está relacionada con la construcción de modelos para la representación de datos relacionados con los datos biométricos, de manera que los mismos puedan consultarse, recuperarse y compararse de manera simple y eficiente (el método de comparación “*matching*” se utiliza en los procesos de identificación/verificación). Esto además respetando las especificaciones de los estándares adoptados.

Estos modelos, tienen estructuras de datos complejas y pueden ocasionar algunos inconvenientes si se trabaja con el modelo de datos relacional:

1. Las estructuras de los metadatos para la búsqueda y comparación de datos biométricos (*iriscode*, minucias, etc.), por lo general son estructuras no atómicas, por lo que en las bases de datos relacionales (para respetar la primera forma normal) se deben separar en varias tablas. Esto hace que sea menos eficiente el tratamiento. Por ejemplo, en la comparación de dos conjuntos de metadatos, necesita realizar operaciones de *join*.
2. El procesamiento de los datos se debe hacer desde lenguajes particulares (C++, java, etc.), lo que presenta el problema de la compatibilidad (y dependencia) con los lenguajes de programación, ya que se deben transferir los datos generados por métodos implementados en distintos lenguajes a la base de datos para ser almacenados. Este tema no es trivial, y se debe tratar para cada lenguaje en particular.
3. Relacionado con lo anterior, para que un lenguaje pueda tratar con los datos, como por ejemplo con el método de comparación entre dos rasgos biométricos, se deben crear cursores que comúnmente deben ser transferidos a través de una red para que retornen un único valor. Esto puede

provocar problemas de tráfico de red innecesarios, dado que estos procesos retornan un único valor lógico, en el caso de la comparación, utilizado en el proceso de verificación, retorna si verifica o no verifica.

Frente a estos inconvenientes de las bases de datos relacionales, aquí se ha optado por utilizar el modelo Objeto-Relacional (OR) [12]. Este modelo permite dar soluciones a las limitaciones de las bases de datos relacionales ya que brindan las siguientes facilidades:

1. **Definir tipos de datos:** estos tipos pueden contener estructuras complejas como colecciones, objetos grandes, etc. sin la limitación de la primera forma normal.
2. **Definir e implementar el comportamiento de los datos:** se pueden crear los métodos que gestionen los datos de los tipos definidos, y así facilitar el acceso seguro desde las aplicaciones que los utilicen.
3. **Definir e implementar métodos de acceso de dominio:** esto permite mejorar los tiempos de acceso en la recuperación de datos de un dominio específico, como en el caso de los datos biométricos.

Las facilidades mencionadas, permiten crear infraestructuras que extienden los servicios de del Sistema de Gestión de Base de Datos [13]. Estas le permiten a este sistema tratar con datos de un dominio específico, en este caso, con datos biométricos.

En lo que concierne al proyecto, esta infraestructura debe contemplar:

- **La creación de tipos específicos para los datos biométricos:** esto incluye tipos para los datos crudos (imágenes, audios, videos), tipos para la codificación de los mismos (*iriscode*, minucias, etc.), y todos los metadatos necesarios para generar los registros de transacciones ANSI/NIST-ITL1-2011.
- **Definir e implementar los métodos que gestionen los tipos antes definidos:** aquí se incluyen, entre otros, la generación de las codificaciones específicas, métodos de comparación (*matching*), métodos de



generación e importación de registros ANSI/NIST-ITL1-2011.

- **Crear índices de dominio que permitan mejorar los tiempos de respuestas en los procesos de identificación:** en la identificación biométrica, se necesita acceder a varios registros para poder identificar a una persona. Mejorar la eficiencia en estos accesos, no es simple de realizar con índices tradicionales como árboles b, hash, etc. Esto se debe a que los códigos (plantillas) generadas para la comparación, son datos multidimensionales y necesitan métodos de acceso específicos para el dominio.

## Objetivos

El objetivo general de este trabajo es desarrollar una infraestructura genérica en BDOR para la identificación de personas mediante el reconocimiento de iris y voz que permita la interoperabilidad entre organismos.

Como objetivos específicos se tienen:

- Realizar los ajustes para adecuar a normas internacionales el software para reconocimiento mediante iris desarrollado en el proyecto anterior.
- Implementar un software para el reconocimiento de voz acorde a normas internacionales.
- Crear modelos en BDOR, que posibiliten almacenar tanto las plantillas como los metadatos necesarios para adecuar los formatos de intercambio de datos al estándar (ANSI/NIST-ITL 1-2011), de manera que permita compartir información con diferentes organismos.
- Proponer métodos de acceso en BDOR para la mejora de la eficiencia en los procesos de identificación de iris y/o voz.
- Crear prototipos utilizando dispositivos estándares y de bajo costo para la captura de la imagen del iris y voz.

## Formación de Recursos Humanos

La estructura del proyecto se formará por el director y co-director, cuatro integrantes docentes y dos alumnos. El director del proyecto se encuentra dirigiendo/co-dirigiendo tres tesis de Maestría en Sistemas de Información de la Facultad de Ciencias de la Administración de la UNER (dos son integrantes del proyecto).

Además, el proyecto contará con dos becarios de Iniciación en la Investigación. Las tareas realizadas por los becarios estarán relacionadas con el desarrollo y testeado de software.

En lo que hace a formación de doctorado la integrante Graciela Etchart se encuentra realizando cursos válidos para la obtención de créditos del Doctorado en Ciencias de la Computación de la UNSur. Estos cursos se dictan en la Facultad de Administración de la UNER por convenio con la UNSur.

## Referencias

1. Wing B. ANSI/NIST-ITL 1-2011. Information Technology: American National Standard for Information Systems Data Format for the Interchange of Fingerprint, Facial & Other Biometric Information. November, 2011.
2. Carlos E. Alvez, Marcelo G. Benedetto. "Los Sistemas Biométricos y su Factibilidad de Aplicación en los Organismos Estatales". XII Workshop de Investigadores en Ciencias de la Computación (WICC 2010), El Calafate, Santa Cruz Argentina, 5 y 6 de Mayo de 2010, Universidad Nacional de la Patagonia San Juan Bosco. Páginas 247-251.
3. Graciela Etchart, Lucas Luna, Carlos Leal, Marcelo Benedetto, Carlos Alvez. Sistemas de reconocimiento biométricos, importancia del uso de estándares en entes estatales. CGIV - XIII Workshop de Investigadores en Ciencias de la Computación (WICC 2011), 5 y 6 de Mayo de 2011. Universidad Nacional de Rosario. Rosario – Argentina. Páginas 339-343.

4. Graciela Etchart, Lucas Luna, Rafael Leal, Marcelo Benedetto, Carlos Alvez. "Sistema adecuado a estándares de reconocimiento de personas mediante el iris". CGIV - XIV Workshop de Investigadores en Ciencias de la Computación (WICC 2012), 25 y 26 de Abril de 2012. Universidad Nacional de Misiones. Posadas – Argentina. Páginas 321-325.
5. Alvez Carlos E. Modelos para la recuperación de imágenes por similitud en Bases de Datos Objeto-Relacionales. Tesis Doctoral. Santa Fe, Argentina, Abril de 2012. ISBN 978-987-33-2249-5.
6. Casal Gabriel, Rovolta Mercedes. Biometrías. Herramientas para la Identidad y la Seguridad Pública. Jefatura de Gabinete de Ministros. Presidencia de la Nación. Noviembre de 2010.
7. Julio Fuoco, Tendencias Biométricas, desafíos y oportunidades. En Biometrías 2. Jefatura de Gabinete de Ministros. Presidencia de la Nación. Octubre de 2011.
8. Development of Draft Type-11 Voice Signal Record. Investigatory Voice Biometrics Committee Report, 9 March, 2012. In [http://biometrics.nist.gov/cs\\_links/ibpc2012/presentations/Day4/IVBCReport\\_v1.8.pdf](http://biometrics.nist.gov/cs_links/ibpc2012/presentations/Day4/IVBCReport_v1.8.pdf).
9. Daugman, J.: High condence visual recognition of persons by a test of statistical independence. IEEE Transactions on Pattern Analysis and Machine Intelligence 15(11), 1148-1161 (November 1993).
10. J. Daugman and C. Downing, "Effect of severe image compression on iris recognition performance," IEEE Transactions on Information Forensics and Security, vol. 3, no. 1, p. 52–61, 2008.
11. Melton Jim, "(ISO-ANSI Working Draft) Foundation (SQL/Foundation)", ISO/IEC 9075-2:2003 (E), United States of America (ANSI), 2003.
12. Alvez C., Benedetto M., Berón G., Etchart G., Luna L. y Leal C. Desarrollo de un sistema multi-biométrico mediante reconocimiento de iris y voz, adecuado a estándares, para su aplicación en organismos públicos. SIE 2011 – Simposio de Informática en el Estado. Córdoba, 31 de Agosto, 01 y 02 de Septiembre de 2011. 40° JAIIO. Páginas: 206 - 220.
13. Carlos E. Alvez, Aldo R. Vecchietti. Combining Semantic and Content Based Image Retrieval in ORDBMS. KES'2010, Lecture Notes in Computer Science, 2010, Volume 6277/2010, 44-53. Editors Rossitza Setchi. Ivan Jordanov, Robert J. Howlett, Lakhmi C. Jain. Springer-Verlag Berlin Heidelberg (2010).

## Análisis de fuentes de información para el proceso de diseño de un datawarehouse sobre pacientes diabéticos

M. E. Llorente<sup>1</sup>, A. Sigura<sup>1,2</sup>, J. Besso<sup>1</sup>, E. Mangia<sup>1</sup>, A. J. Hadad<sup>1,2</sup>, B. Drozdowicz<sup>1,2</sup>

<sup>1</sup>Facultad Ciencia y Tecnología, Universidad Autónoma de Entre Ríos

<sup>2</sup>Facultad Ingeniería, Universidad Nacional de Entre Ríos

Ruta 11, Oro Verde, Entre Ríos, Argentina

[mellorente@arnet.com.ar](mailto:mellorente@arnet.com.ar), [bdrozdo@santafe-conicet.gov.ar](mailto:bdrozdo@santafe-conicet.gov.ar)

### Resumen

Para el desarrollo de un modelo de datos para un Datawarehouse (DW) se tuvo en cuenta la evolución en las metodologías de diseño de los Modelos Multidimensionales, en consecuencia es posible definir que las mismas están basadas en dos aspectos fundamentales: Requerimientos y Fuentes de Información existentes.

En este trabajo se describen las diversas fuentes de información y sus estructuras de datos a tener en cuenta para el diseño y desarrollo de un DW para el apoyo a las decisiones de profesionales médicos que atienden pacientes diabéticos.

Analizando las fuentes de información existentes, en una primera clasificación, se distinguen datos estructurados y no estructurados. Se denomina estructurada a toda aquella fuente de información que tenga un diseño lógico y conceptual, ejemplo una base de datos de historia clínica. Es decir que exista información sobre el dato en sí mismo.

Dentro del conjunto de las fuentes no estructuradas se incluyen aquellas que necesitan un procesamiento previo para contextualizar los datos contenidos y convertirlos en información. En este grupo se encuentran las señales fisiológicas, imágenes y el texto libre, los cuales indefectiblemente necesitan un preprocesamiento, para formar parte de una estructura orientada al análisis, como es un DW.

**Palabras clave:** Datawarehouse, Pacientes

Diabéticos, Estructuras de Datos, Metodologías ETL, Herramienta de Desarrollo

### Contexto

El presente trabajo se inserta en un Proyecto de Investigación Plurianual (PIDP) denominado “*Sistema de Soporte a la Toma de Decisiones basado en datawarehouse para pacientes diabéticos*”. Dicho proyecto es desarrollado en la Facultad de Ciencia y Tecnología de la Universidad Autónoma de Entre Ríos (FCYT - UADER).

### Introducción

Este trabajo está orientado a describir parte del proceso para un primer modelo multidimensional del DW a ser utilizado en el PIDP mencionado.

El diseño del mismo está basado en un enfoque de tres niveles de análisis, a través de un proceso iterativo [1].

Para este desarrollo el más bajo es el nivel del paciente individual, donde los datos sobre el mismo se pueden visualizar y analizar, por ejemplo, para encontrar un patrón en el desarrollo de una enfermedad vinculado al mismo. Este nivel de análisis se centra en dar al paciente en particular el mejor tratamiento

posible, y por tanto es importante para la práctica de la atención médica. El siguiente es el nivel de grupo de pacientes, donde los datos sobre el mismo son analizados, por ejemplo, cuando tengan una enfermedad particular asociada. El tercer nivel es el relacionado con una empresa/institución de salud, donde profesionales clínicos, administradores y especialistas en epidemiología, combinan datos para investigar la calidad, efectividad y eficiencia global de los servicios proporcionados.

En esta etapa del modelado se considerarán solamente los datos relacionado con el primer nivel, es decir el del paciente individual.

Para el desarrollo del modelo se tuvo en cuenta la evolución en las metodologías de diseño de los Modelos Multidimensionales desde el año 1998 a la actualidad, y analizando las propuestas de diferentes autores de referencia [2], es posible definir que las metodologías están basadas en dos aspectos fundamentales:

- Requerimientos
- Fuentes de Información existentes

En el trabajo presentado en [1] se realizó una primera propuesta de modelado, considerando solamente los Requerimientos basados en Casos de Uso, dentro de un proceso iterativo.

Esto fue necesario teniendo en cuenta la diversidad de las fuentes de información y la variabilidad de las características y condiciones de los pacientes diabéticos. Por este motivo generalmente los usuarios médicos de un DW clínico, como el propuesto en este trabajo, no tienen perfectamente definidos como van a analizar los datos y por lo tanto resulta casi imposible comenzar su diseño conociendo todos los requerimientos a priori [3]. En consecuencia en un proyecto de DW clínico puede resultar necesario implementar para su diseño un análisis del tipo iterativo. Un objetivo

de estos DW es ser lo suficientemente flexibles para tratar con estos cambios, esto conlleva a considerar un proceso de diseño iterativo diferente al proceso incremental, aún cuando este último sea el más convencional.

Por su parte el presente trabajo se enfoca en el segundo de los aspectos indicados anteriormente, las fuentes de información existentes [1,4], en lo que refiere a su estructura de origen y necesidades de transformación en cada caso. Estos aspectos forman parte de las líneas de investigación descriptas en [4] para el PIDP.

En un DW clínico se deben involucrar procesos para el procesamiento de imágenes, señales y datos, como ser: registración, extracción y cuantificación de características. También se requiere el modelado de datos de multimedia en variadas formas- texto libre, reportes estructurados, imágenes en 2 o 3 D, capacidad de hacer zoom a las imágenes, señales, datos espectrales, gráficos, video, publicaciones escaneadas, en lugar de datos en formato texto como la mayor parte de las DW empresariales. Deben tener capacidad para buscar cualquier formato de información en forma cualitativa o cuantitativa y no solamente texto estructurado o números en registros.

Los DW clínicos enfatizan la preparación y adquisición de datos con protocolos predefinidos. Además proveen herramientas analíticas y estadísticas para soportar procesos de verificación o de Minería de Datos.

## **Líneas de investigación y desarrollo**

1. Estructuras de datos representativas del dominio de análisis.
2. Métodos ETL para fuentes de información de referencia.



## Resultados y Objetivos

Analizando las fuentes de información existentes, en una primera clasificación, es posible distinguir datos estructurados y no estructurados. En este contexto se denomina estructurado a toda aquella fuente de información que tenga un diseño lógico y conceptual, como por ejemplo una base de datos de Historias Clínicas. Es decir que exista información sobre el dato en sí mismo.

Dentro del conjunto de las fuentes no estructuradas se incluyen aquellas que necesitan un procesamiento previo para contextualizar los datos contenidos y convertirlos en información. En este grupo se encuentran las señales fisiológicas, imágenes y el texto libre, los cuales indefectiblemente necesitan un preprocesamiento, para formar parte de una estructura orientada al análisis, como lo es un DW. Un resumen de esta clasificación se presenta en la Figura 1.

Esta diferenciación entre fuentes de información tiene consecuencias en ciertos aspectos de la implementación, como ser la

formulación de las ETL que proveerán de datos al DW.

Los requerimientos de información identificados anteriormente proporcionarán las bases para realizar el diseño y la modelización del Modelo Multidimensional del DW. En esta fase se identificarán las fuentes de los datos (sistema operacional, fuentes externas) y las transformaciones necesarias para, a partir de dichas fuentes, obtener el modelo lógico de datos del DW. Este modelo estará formado por entidades y relaciones que permitirán resolver las necesidades del proceso de atención del paciente. El modelo lógico se traducirá posteriormente en el modelo físico de datos que se almacenará en el DW y que definirá la arquitectura de almacenamiento del mismo adaptándose al tipo de uso que se realice.

Teniendo en cuenta este análisis se realizará el desarrollo de un primer prototipo utilizando la herramienta provista por IBM (IBM InfoSphere Platform), para la cual la institución tiene un convenio de uso académico de sus desarrollos [5].

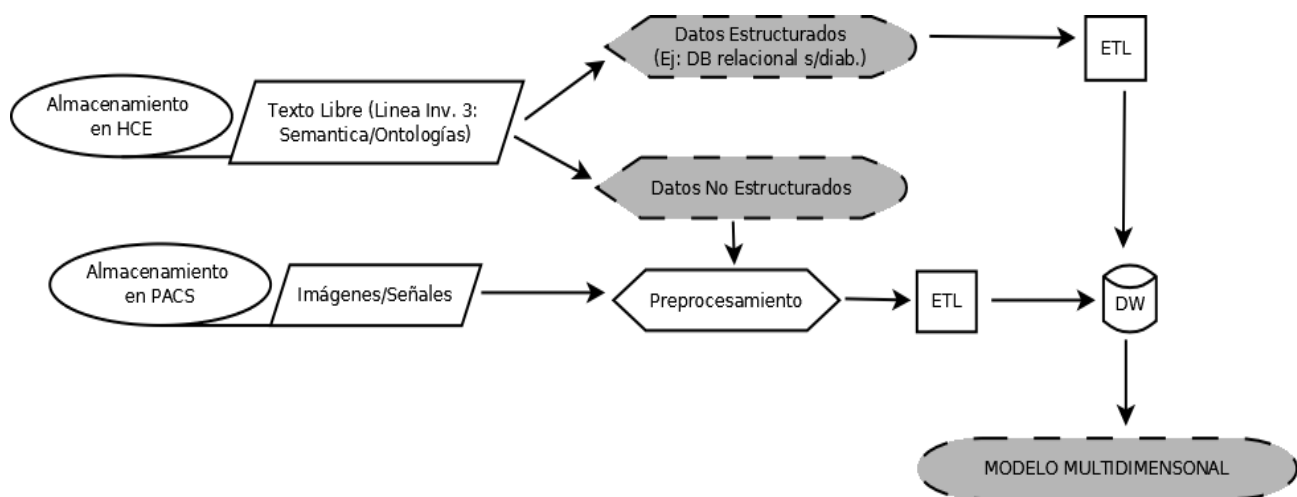


Figura 1 - Clasificación de las fuentes de información

Considerando las diferentes estructuras de datos que generan información se procederá a la definición de las ETL correspondientes y los preprocesamientos previos necesarios.

La herramienta indicada anteriormente tiene capacidad para extraer datos automáticamente de diferentes fuentes u orígenes, como las mencionadas a continuación:

- Base de datos relacionales (MySQL, MS SQLServer, Progress, DB2, Informix, Oracle, Teradata).
- Archivos de texto.
- Extracción desde planillas de cálculo, archivos separados por punto y coma (.csv).
- Servicios web
- Archivos XML

Los orígenes de datos que puede manipular la herramienta incluyen archivos indexados, archivos secuenciales, bases de datos relacionales, orígenes de datos externos, aplicaciones y colas de mensajes. Ello puede implicar algunas de las transformaciones siguientes:

- Conversiones de tipos de datos y de formato de serie y numérico.
- Derivaciones y cálculos que apliquen algoritmos y normas a los datos.
- Comprobaciones y aplicación de datos de referencia para validar identificadores. Este proceso se utiliza para crear un depósito de datos normalizado.
- Conversión de datos de referencia de orígenes dispares a un conjunto de referencia común, creando coherencia entre estos sistemas. Esta técnica se utiliza para crear un conjunto de datos maestro (o dimensiones conformadas).
- Agregaciones para generación de informes y análisis.
- Creación de bases de datos analíticas o de informes, como por ejemplo cubos o depósitos de datos. Este proceso implica desnormalizar datos en estructuras tales como esquemas de

estrella o de copo de nieve para mejorar el rendimiento y la facilidad de uso para los usuarios.

Las fuentes de información que se consideran en el dominio de atención de pacientes diabéticos son:

- Historias Clínicas (Consultorio, Eventos, Medicamentos). Para este trabajo se considera que las historias clínicas están almacenadas en una Base de Datos Relacional (BDR).
- Datos de Laboratorio o Tipos de análisis y/o estudios. Contenido. Tablas y Relaciones. **BDR**
- Datos Recogidos por el paciente en su casa (Ej: Toma de datos de niveles de glucosa en sangre a través de dispositivos portátiles). **Archivos de Texto Estructurado**
- Base de Datos de Imágenes en formato DICOM. La imagen contenida no tiene ningún procesamiento lo cual se considera un dato no estructurado. Para que pueda ser interpretado por la herramienta, se debe generar un módulo de preprocesamiento que brinde la información asociada en un modelo de datos. Ejemplos: Imágenes de Fondo de Ojo con presencia de Retinopatías Diabéticas. Análisis Evolutivo de la patología, registración de imágenes, etc.[6-9]
- Base de Datos de señales fisiológicas. Caso similar al de las imágenes. Ejemplos: señales de electrocardiograma, presión arterial, saturación de oxígeno, etc. Análisis de estabilidad hemodinámica, patrones temporales de arritmias, índices de severidad, etc. [10-12]

- Base de Datos de Farmacia. Medicamentos. **BDR + Planillas de cálculo**
- Información complementaria accesible vía **servicios web**, como ser aspectos regulatorios (ANMAT) o información de otras instituciones (Colegios Médicos, Obras Sociales, etc.).

Este relevamiento ha permitido considerar los aspectos más relevantes del dominio, sin embargo con el avance del proyecto pueden surgir otras consideraciones sobre esta temática.

Como trabajo futuro, lo descrito en este trabajo relacionado con las diferentes estructuras de datos con las que va a tener interacción el DW, se complementará con el proceso iterativo de diseño de la DW basado en Casos de Uso [1].

## Formación de Recursos Humanos

El equipo de trabajo está conformado por especialistas del área informática y de bioingeniería. Integrantes del equipo tienen formación de postgrado tanto en el área de sistemas de información como en el área biomédica, así como también experiencia en el ámbito profesional en lo que refiere al desarrollo de sistemas.

## Referencias

- [1] Proceso de Diseño basado en Casos de Uso para un Datawarehouse Clínico. M. E. Llorente, Aldo Daniel Sigura, Javier Besso, Alejandro Hadad, Bartolomé Drozdowicz. CACIC 2012
- [2] A survey of Multidimensional Modeling Methodologies. Oscar Romero, Alberto Abelló. International Journal of Data Warehousing & Mining, 5(2), 1-23, April-June 2009
- [3] Wong S, Hoo K, Knowlton R., et al. Design an

applications of a multimodality image datawarehouse framework. J Am Med Inform Assoc. 2002; 9: 239-254

[4] Sistema de soporte a la toma de decisiones basado en datawarehouse para pacientes diabéticos. M. E. Llorente, Aldo Daniel Sigura, Alejandro Hadad, Bartolomé Drozdowicz. WICC 2012

[5] Sitio Oficial de IBM Infosphere DataStage: <http://www-01.ibm.com/software/data/infosphere/datastage/>

[6] "Implementación y aplicación de algoritmos Retinex al preprocesamiento de imágenes de retinografía color", N. Londoño, G. Bizai, B. Drozdowicz, Revista Ingeniería Biomédica, ISSN 1909-9762, volumen 3, número 6, julio-diciembre 2009, págs. 36-46.

[7] "Algorithms evaluation for fundus images enhancement", Braem V., Marcos M., Bizai G, Drozdowicz B, Salvatelli A., Journal of Physics: ConferenceSeries 332 (2011) 012035.

[8] "Analysis and Implementation of Methodologies for the Monitoring of Changes in Eye Fundus Images", A. Gelroth, D. Rodríguez, A. Salvatelli, B. Drozdowicz, G. Bizai, Journal of Physics: Conference Series (JPCS), 23 December 2011, Vol. 332 (2011) 012036, con referato doi:10.1088/1742-6596/332/1/012036

[9] "Monitoring of Changes in Fundus Image", Authors: R. M. Torres, A. Gelroth, D. Rodriguez, A. Salvatelli, B. Drozdowicz, G. Bizai. ARVO/ISIE Imaging Conference Saturday, May 5, 2012, Grand B Ballroom, Greater Fort Lauderdale/Broward-Poster, <http://www.arvo.org/eweb/startpage.aspx?site=isie>

[10] A. Hadad, D. Evin, B. Drozdowicz, O. Chiotti. Temporal Abstraction for the Analysis of Intensive Care Information. Journal of Physics: Conference Series. Volume 90, 2007. ISSN: 1742-6596

[11] Hypotension States' Prediction by using the Hidden Markov Models. Diego Evin, Alejandro Hadad, Mauro Martina, Bartolomé Drozdowicz Revista Facultad de Ingeniería, UPTC, 2011, vol. 20, No. 30, pp 55-63, ISSN 0121-1129.

[12] Hadad, Alejandro Javier; Solano, Agustin Ezequiel y Drozdowicz, Bartolomé (2012). "Prototipo para la comparación de patrones temporales secuenciales de arritmias cardíacas". En: Ventana Informática No. 26 (ene.-jun., 2012). Manizales (Colombia): Facultad de Ciencias e Ingeniería, Universidad de Manizales. pp 29-43 ISSN: 0123-9678

## Análisis de Incidentes Informáticos usando Modelos de Asociación y Métodos del Análisis de Datos Multivariante.

García, Alejandro (1), Corso, Cynthia Lorena (2), Gibellini, Fabián (3), Rapallini, Marcos (4)

Laboratorio de Sistemas de Información. / Dpto. Ingeniería en Sistemas de Información / Facultad Regional Córdoba / Universidad Tecnológica Nacional  
Maestro Marcelo López s/n. Ciudad Universitaria.

Teléfono: 5986011

[malejandrogarcia@hotmail.com](mailto:malejandrogarcia@hotmail.com) / [cynthia@bbs.frc.utn.edu.ar](mailto:cynthia@bbs.frc.utn.edu.ar) / [speaker@bbs.frc.utn.edu.ar](mailto:speaker@bbs.frc.utn.edu.ar) / [marco@bbs.frc.utn.edu.ar](mailto:marco@bbs.frc.utn.edu.ar)

### Resumen

El objetivo de este proyecto es la caracterización y búsqueda de relaciones y asociaciones significativas entre variables relacionadas con la ocurrencia de incidentes en equipos informáticos, en el contexto de un laboratorio académico y de Investigación y Desarrollo. Para ello se implementa el uso de técnicas pertenecientes a la rama de Minería de Datos y métodos incluidos en el Análisis de Datos Multivariante.

Mediante la aplicación de técnicas y métodos apropiados, se pretende la elaboración de un modelo de conocimiento de carácter predictivo.

El mismo nos debe aportar conocimiento de las variables o factores de mayor incidencia en la presentación de incidentes, como así también el establecimiento de relaciones y modelos subyacentes en las mismas.

**Palabras Clave:** Minería de Datos, Análisis de Datos Multivariante, Técnicas de Asociación, Gestión de Incidentes Informáticos.

### Contexto

Este proyecto surge de necesidades concretas en el contexto de un Laboratorio Informático y está relacionado con la gestión de mantenimientos y prevención de incidentes en equipos informáticos.

El mismo forma parte de un conjunto de proyectos pertenecientes al Laboratorio de Sistemas, cuyo propósito es el estudio y análisis de temáticas relacionadas con problemáticas y acciones de mejoras a implementar en este contexto.

### Introducción

La gestión de las actividades relacionadas con el área de mantenimiento es clave, para que cualquier unidad de negocio pueda desarrollar sus actividades de manera eficiente. Por ello es fundamental el establecimiento de un sistema, que permita la detección de acciones de carácter preventivo con el objetivo de minimizar la presencia de fallos o incidentes.

La presentación frecuentes de fallos o incidentes, en cualquier unidad de negocio u organizacional, tienen un impacto negativo desde el punto de vista operacional, como así también económico. Las estadísticas reflejan que un alto porcentaje de las horas hombre son dedicadas a la ejecución de tareas relacionadas a la solución de fallos o incidentes y que no han sido detectados por el área de mantenimiento.

Por todo lo expuesto, se considera necesario y significativo el estudio que permita el desarrollo de una metodología que permita el descubrimiento de factores influyentes que propician la presentación de incidentes o fallos en equipos, como así también detectar si existen relaciones entre los mismos. Para ello es fundamental el análisis y evaluación de datos históricos de los informes de mantenimientos correctivos reportados por el personal del Área Técnica, perteneciente al laboratorio.

Las aplicaciones de minería de datos posibilitan la identificación de tendencias y comportamientos en los datos que no son evidentes.

Para esta investigación se ha seleccionado una de las técnicas de Minería de Datos, que en función a



la problemática que se desea resolver, la más adecuada es la técnica de Asociación. Esta nos permitirá obtener un modelo de conocimiento en forma de reglas de asociación que permiten develar hechos que ocurren dentro de un conjunto de datos determinado. La selección y futura aplicación de este tipo de técnica puede resultar interesante para el descubrimiento de relaciones entre variables o atributos de un conjunto de datos.

El importante avance que ha tenido el campo de la tecnología y el abaratamiento de costos ha traído como consecuencia un aumento significativo en la cantidad de datos que son almacenados en muchas ocasiones en diferentes formatos.

La Minería de Datos es un mecanismo que nos permite facilitar precisamente, la búsqueda de información valiosa en grandes volúmenes de datos.

La implementación y construcción de un modelo de conocimiento que permita conocer el comportamiento de los incidentes en el periodo considerado, facilita la elaboración de un plan de prevención que permita la disminución de los reportes de incidentes.

Esto permite lograr mayor disponibilidad de los equipos informáticos para las diversas actividades académicas que se desarrollan en el Laboratorio de Sistemas. Desde el punto de vista económico se logra disminuir los costos relacionados con la adquisición de determinados insumos o componentes/piezas que son utilizados en las tareas de mantenimiento.

Teniendo en cuenta los resultados obtenidos, esto podría impactar principalmente en el ámbito de mantenimiento y fiabilidad de equipos informáticos en el contexto de laboratorios informáticos perteneciente a entidades públicas municipales, provinciales y nacionales.

El descubrimiento de reglas, factores y relaciones entre variables que intervienen en la presentación de incidentes en los equipos informáticos, permitirá planificar diferentes aspectos como:

- Identificación del él/los posible/s origen/es de incidentes, como por ejemplo conocer si la presentación de fallos está relacionado mayormente por inconvenientes de un determinado tipo de componente o pieza.

- Elaboración de procedimientos técnicos y tareas de mantenimiento preventivo a efectuar periódicamente.
- Capacitación al personal de Área Técnica.
- Disminución de los tiempos muertos o de parada de los equipos.
- Aprovechamiento y uniformidad en la carga de trabajo del personal de Área Técnica, debido a una planificación de actividades.
- Diseño de un proceso de compras de componentes o piezas de los equipos, que facilite la disponibilidad de un stock aceptable.

Las herramientas seleccionadas, para el procesamiento de datos y obtención de los modelos en esta investigación son Weka (Software Libre) e InfoStat (Software Comercial).

[1] Weka: esta herramienta fue concebida en la universidad de Waikato (Nueva Zelanda) implementado en lenguaje Java y que dispone de un conjunto de librerías que facilitan la integración con otras herramientas.

Además Weka contiene las herramientas necesarias para realizar transformaciones con los datos, tareas de clasificación, regresión, clustering y asociación.

La licencia de Weka es GPL, lo que significa que este programa es de libre distribución y difusión.

[2] InfoStat: es un software para análisis estadístico de aplicación general. Dispone de una amplia gama de herramientas para el tratamiento de estadísticas descriptivas y gráficos para el análisis exploratorio, como así también métodos avanzados de modelación estadística y análisis de datos multivariado.

El objetivo general de esta investigación es la implementación de herramientas relacionadas con la Minería de Datos y la Estadística; facilitando la generación de un modelo de conocimiento que permita describir y caracterizar el comportamiento de los incidentes reportados de los equipos informáticos.

### **Líneas de Investigación y Desarrollo**

- Herramientas de Inteligencia de Negocios
- Minería de Datos
- Algoritmos de Asociación
- Mantenimiento y prevención de incidentes.

- Técnicas de asociación, aplicada al ámbito de laboratorio informático.
- Software de Minería de Datos.
- Métodos de Pre-procesamiento de datos.
- Elaboración de metodología relacionada con la implementación de proyecto de Minería de Datos, aplicada al ámbito ingenieril.
- Parámetros o métricas de calidad para el modelo de conocimiento obtenido.

### Resultados y Objetivos

Entre los resultados del avance de este proyecto se pueden mencionar los siguientes:

- Se ha investigado sobre diversas herramientas estadísticas y de aprendizaje automático, con el propósito de facilitar la implementación del proyecto de minería de datos.
- Se ha investigado sobre métodos que facilitan la selección de atributos relevantes
- Se ha investigado sobre diversos algoritmos dentro de la rama de las técnicas predictivas, con el objetivo de seleccionar la más adecuada, en relación a la problemática a resolver.
- Integrante del grupo han participado en experiencias en el rol de coordinación en Panel: “Aplicación de técnicas de Minería de Datos usando software Weka”, en el Congreso Argentino de Estudiantes de Ingeniería Industrial y carreras afines (CAEII 2009).
- Se ha participado en Congresos Nacionales (JAIOO 2012/CAIM y WICC (2011))

La diagramación de actividades planificadas para este proyecto son las que se detallan a continuación:

- Búsqueda de alternativas respecto a herramientas estadísticas y de aprendizaje automático para facilitar la implementación del proyecto de minería de datos.
- Identificación de criterios para la selección de la herramienta que se adapte mejor a la problemática que se desea resolver.
- Análisis de los datos o reportes existentes relacionado con la gestión de incidentes.
- Selección de muestras de datos considerados para este estudio.
- Evaluación de los atributos o variables que serán consideradas para el estudio de gestión de incidentes de los equipos informáticos.

- Carga y migración de los datos en un formato adecuado para ser interpretada por la herramienta de minería de datos seleccionada.
- Selección y Evaluación de algoritmos de aprendizaje supervisado para la implementación de técnica de asociación.
- Implementación y validación.

Dentro de los objetivos específicos.

a) Determinar dentro de la Técnica de Asociación, la variedad de algoritmos existentes y evaluar cuál de ellos se adapta mejor a la situación problemática planteada. Considerando criterios o indicadores de confiabilidad que determinen el nivel de calidad de las reglas de asociación resultantes.

b) Selección y Evaluación de herramientas para la implementación de proyectos de Minería de Datos que dispongan una variedad significativa de algoritmos pertenecientes a las Técnicas de Asociación.

c) Difundir el uso herramientas, como una alternativa de instrumento de inteligencia de negocios para la ejecución de proyectos de Minería de Datos, a través de charlas, conferencias y publicaciones en el ámbito universitario.

c) Generación de un modelo de conocimiento (Reglas de asociación) que nos permita la detección de los factores que tienen alta incidencia en el reporte de incidentes de los equipos informáticos pertenecientes al laboratorio.

d) Transferir al seno de la cátedra Inteligencia Artificial (5to. Nivel de la carrera Ingeniería en Sistemas de la U.T.N. - F.R.C.) los resultados obtenidos, con la finalidad de mejorar el diseño curricular de las asignaturas y enriquecer las mismas con el intercambio interdisciplinario entre investigación, aplicación y análisis de la evidencias resultantes.

f) Evaluar la posibilidad de extender este estudio a otros laboratorios informáticos con esta problemática, principalmente en el contexto académico para los distintos niveles educativos

Los resultados obtenidos:

- Definición de la arquitectura tecnológica en torno al proyecto de Minería de Datos, que generalmente suele tener una arquitectura cliente-servidor.

[3] Para lograr una aplicación óptima de estas técnicas avanzadas, las mismas deben estar totalmente integradas con el data warehouse así como con herramientas flexibles e interactivas para el análisis de negocios.

Varias herramientas de Data Mining actualmente operan fuera del warehouse, requiriendo pasos extra para extraer, importar y analizar los datos. Además, cuando nuevos conceptos requieren implementación operacional, la integración con el warehouse simplifica la aplicación de los resultados desde Data Mining.

Este warehouse puede ser implementado en una variedad de sistemas de bases relacionales y debe ser optimizado para un acceso a los datos flexible y rápido.

- Selección de variables: Aún después de haber sido preprocesados, en la mayoría de los casos se tiene una cantidad significativa de datos. La selección de características reduce el tamaño de los datos eligiendo las variables más influyentes en el problema, sin apenas sacrificar la calidad del modelo de conocimiento obtenido del proceso de minería.

Los métodos para la selección de características son básicamente dos:

- Aquellos basados en la elección de los mejores atributos del problema,
- Y aquellos que buscan variables independientes mediante tests de sensibilidad, algoritmos de distancia o heurísticos.

Se ha investigado que la herramienta Weka dispone de un conjunto de algoritmos que facilitan esta tarea. En general estos algoritmos pueden ser clasificados por diversos criterios.

[4] Una categorización popular es aquella en la que los algoritmos se distinguen por su forma de evaluar atributos y se clasifican en: Filtros, donde se seleccionan y evalúan los atributos en forma independiente del algoritmo de aprendizaje y Wrappers (envoltorios), los cuales usan el desempeño de algún clasificador (algoritmo de

aprendizaje) para determinar lo deseable de un subconjunto. Los algoritmos evaluadores en un subconjunto de atributos disponibles en Weka testeados son:

CfsSubsetEval, ConsistencySubsetEval (Filtros), ClassifierSubsetEval, WrapperSubsetEval (Wrappers)

Se complementará esta etapa con las pruebas de los algoritmos que permiten la evaluación de atributos individuales como:

ChiSquaredAttributeEval, GainRatioAttributeEval, InfoGainAttributeEval y OneRAttributeEval.

- Implementación de herramienta para la carga de datos.

- Algoritmos de Extracción de Conocimiento:

Mediante una técnica de minería de datos, se obtiene un modelo de conocimiento, que representa patrones de comportamiento observados en los valores de las variables del problema o relaciones de asociación entre dichas variables.

También pueden usarse varias técnicas a la vez para generar distintos modelos, aunque generalmente cada técnica obliga a un preprocesado diferente de los datos.

[5] La Minería de reglas de asociación es una técnica importante en la Minería de Datos y consiste en encontrar relaciones de implicación entre los valores de los atributos de los objetos de un conjunto de datos.

Actualmente en esta fase del proyecto se está investigando acerca de las características y ventajas de los algoritmos de asociación que dispone la herramienta Weka son: A Priori, Filtro Asociado, HotSpot, A priori-Predictivo y el Tertius. Una línea futura de investigación es el análisis de algoritmos de asociación implementados en otras herramientas de aprendizaje automático.

## Formación de Recursos Humanos

Los integrantes de este proyecto, en su gran mayoría están conformados por docentes pertenecientes al plantel académico de la carrera de Ingeniería en Sistemas de Información.

Los mismos poseen formación académica de postgrado como especialización y magister en curso.

Este trabajo de investigación pretende integrar y articular contenidos relacionados con el campo de la Minería de Datos en un contexto del campo de la Ingeniería, con los docentes de la asignatura electiva "Inteligencia de Negocios" perteneciente al quinto nivel de la carrera de Ingeniería de Sistemas de Información.

Una de las integrantes está elaborando el plan de tesis, para la carrera de Magister en Sistemas de Información, relacionada con línea de investigación de este proyecto, complementando la investigación con técnicas avanzadas pertenecientes al análisis de datos multivariante.

Todos los integrantes docentes del PID han participado del proceso de categorizaciones en investigación dentro del Programa de Incentivos del MECyT; así como en la categorización interna que posee la U.T.N.

### Bibliografía

- [1] “Manual de Weka”, Diego García Morate  
<http://www.metaemotion.com/diego.garcia.morate/download/weka.pdf>
- [2] <http://www.infostat.com.ar/>
- [3] “Minería de Datos”, Vallejos Sofía, (2006)  
[http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/Mineria\\_Datos\\_Vallejos.pdf](http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/Mineria_Datos_Vallejos.pdf)
- [4] “Aplicación de métodos de selección de atributos para determinar factores relevantes en la evaluación nutricional”, Roxana Ramos, Rosa Ramos Palmero, Ricardo Avalos, María Matilde García Lorenzo,(2007)  
[http://bvs.sld.cu/revistas/gme/pub/vol.9.%281%29\\_01/p1.html](http://bvs.sld.cu/revistas/gme/pub/vol.9.%281%29_01/p1.html)
- [5] “Minería de Reglas de Asociación sobre datos Mezclados”, Ansel Rodríguez González, José Martínez Trinidad, Jesús Ochoa, José Ruiz Shulcloper, (2009)  
<http://ccc.inaoep.mx/portalfiles/file/CCC-09-001.pdf>
- [6] Aprendizaje Automático: conceptos básicos y avanzados. Aspectos prácticos utilizando el software Weka, Basilio Serra Araujo, Pearson Educación, Madrid 2006.

[7] “Introducción a la Minería de Datos”, José Hernández Orallo, Ramírez Quintana, M<sup>a</sup> José, Ferri Ramírez, César, Pearson Prentice Hall, 2005.

[8] “Minería de Datos. Técnicas y Herramientas”, Pérez López, César, González Daniel Santin, Thompson. Madrid. (2007)

[9] “Extracción Automática de Conocimiento en Base de Datos e Ingeniería del Software”, Quintana Ramírez María José, Orallo José Hernández, España (2003)

[10] “Data Mining: Concepts and Techniques”, Jiawei Han & Micheline Kamber, Vipin Kumar, Addison-Wesley (2006)



## Investigación básica y aplicada en Bases de Datos Deductivas

H. Liberatori, J. Paganini, C. González, S. Figueroa, M. Tejerina, A. Vega, C. Castillo, V. Battezzati

Facultad de Ingeniería/ Universidad Nacional de Jujuy

Ítalo Palanca N° 10 San Salvador de Jujuy (4600) Tel 0388 4221582

[hliberatori@hotmail.com](mailto:hliberatori@hotmail.com); [jhpaganini@fi.unju.edu.ar](mailto:jhpaganini@fi.unju.edu.ar); [scgonzalez@fi.unju.edu.ar](mailto:scgonzalez@fi.unju.edu.ar); [tkdfigueroa@hotmail.com](mailto:tkdfigueroa@hotmail.com);  
[mariotejerina@gmail.com](mailto:mariotejerina@gmail.com); [arielalejandrovega@gmail.com](mailto:arielalejandrovega@gmail.com); [ce\\_al\\_castillo@yahoo.com.ar](mailto:ce_al_castillo@yahoo.com.ar),  
[vbattezzati@fi.unju.edu.ar](mailto:vbattezzati@fi.unju.edu.ar)

### Resumen

Con el proyecto presente: “investigación básica y aplicada en Bases de Datos Deductivas”; se pretende avanzar en el estudio, aplicaciones y desarrollo de técnicas para el procesamiento de consultas recursivas para bases de datos deductivas, tendientes a ofrecer alternativas de nuevas prestaciones en Bases de datos relacionales. Constituyendo, por sus características en un encuentro entre conceptos teóricos provenientes de la Lógica Computacional, y del Álgebra Relacional para bases de datos; con técnicas de programación y desarrollo para la implementación de algoritmos concretos de aplicación.

Cabe destacar también que toda la investigación y desarrollo se formaliza y documenta con técnicas de especificación formal del método RAISE

El presente proyecto está categorizado con categoría “A”, la máxima, que lo sitúa en la línea de prioridades de la Facultad de Ingeniería de la UNJu; y se enmarca dentro de un proyecto general de la línea PICTO tendiente a implementar una base de datos de composición de alimentos para Argentina y Latinoamérica.

Otro aspecto que contempla el presente proyecto es el de contribuir a la formación de recursos humanos; ya que como parte del mismo se desarrolla una tesis de maestría en Ingeniería del Software y un proyecto final para el grado de Ingeniero Informático.

### Palabras clave:

Bases de datos deductivas; Algoritmos Lógicos; RAISE, Datalog,

### Contexto

El presente proyecto se inscribe dentro de las líneas prioritarias de investigación de la Facultad de Ingeniería. En el marco de la Universidad Nacional de Jujuy, está acreditado con categoría “A” y financiado por SECTER (Secretaría de Ciencia y Técnica y Estudios Regionales).

Los resultados a obtener, contribuirán al desarrollo de una Base de composición de Alimentos desarrollada mediante el auspicio de un

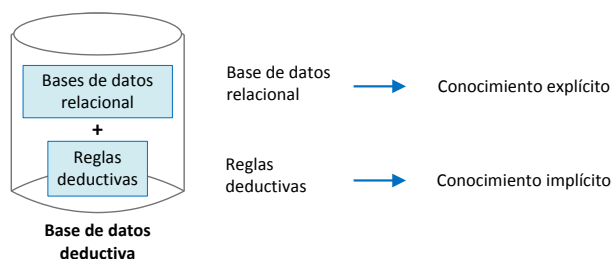
proyecto PICTO, de la Agencia Nacional de Promoción Científica y Tecnológica (FONCyT)

Los investigadores afectados a este proyecto provienen de las cátedras de Lógica Computacional y de la de Bases de Datos, a su vez integran el grupo del proyecto PICTO mencionado.

### Introducción

Las bases de datos deductivas nacen de la necesidad de almacenar y utilizar “conocimiento” de una manera eficiente. Consisten de un conjunto de aserciones (hechos) referidos a la base de datos extensional y de un conjunto de reglas (axiomas) referidos a la base de datos intensional. La base de datos extensional se encuentra generalmente almacenada en una base de datos relacional y constituye el conocimiento básico de las bases de datos deductivas. Las reglas permiten obtener o deducir un nuevo conocimiento a partir de la base de datos extensional, el conocimiento no se encuentra almacenado directamente en la base de datos extensional. De esta manera, las bases de datos deductivas permiten inferir un nuevo conocimiento que puede ser utilizado para la toma de decisiones importantes de una organización.

Es conveniente situar las bases de datos deductivas en un contexto apropiado respecto de las Base de Datos Orientadas a Objetos. Estas últimas proporcionan un mecanismo de modelado natural para los objetos del mundo real, encapsulando su estructura junto con su comportamiento. En cambio las bases de datos deductivas extienden la capacidad expresiva de las bases de datos relacionales, proporcionando interrelaciones con el mundo real en forma de un conjunto de reglas, que permiten definir un conocimiento implícito a partir de datos existentes, como se muestra en la siguiente figura:



Una base de datos deductiva se compone de 3 conjuntos finitos: un conjunto de hechos, un

conjunto de reglas deductivas y un conjunto de restricciones de integridad:

- **Hechos:** se especifican de manera similar a las relaciones, excepto que no es necesario incluir los nombres de los atributos. Una tupla en una relación describe algún hecho del mundo real cuyo significado queda determinado en parte por los nombres de los atributos. En una base de datos deductiva, el significado del valor de un atributo en una tupla queda determinado exclusivamente por su posición dentro de la tupla.
- **Reglas:** presentan un parecido a las vistas relacionales, especifican relaciones virtuales que no están almacenadas, pero que se pueden formar a partir de los hechos, aplicando mecanismos de inferencia basados en las especificaciones de las reglas. La principal diferencia entre las reglas y las vistas, es que en las primeras puede haber recursión y por lo tanto, pueden producir relaciones virtuales que no es posible definir en términos de las vistas relacionales.

### Mecanismos de Inferencia

Otra forma para interpretar el significado de las reglas implica definir un mecanismo de inferencia que el sistema utilice para deducir hechos a partir de reglas. Este mecanismo define una interpretación computacional del significado de las reglas.

El lenguaje de programación lógica Prolog se vale de su mecanismo de inferencia para definir el significado de las reglas y hechos en un programa. Sin embargo, en muchos programas simples de Prolog, el mecanismo de inferencia, infiere los hechos ya sea interpretando por la teoría de la demostración, o bien con un modelo mínimo en la interpretación por la teoría de modelos.

Los mecanismos de inferencia se basan en principios de la inteligencia artificial, basándose en mecanismos de inferencia tales como: inferencia ascendente (encadenamiento hacia adelante), inferencia descendente (encadenamiento hacia atrás). En el primer caso se parten de los hechos contenidos en la base de datos y se generan nuevos hechos utilizando las reglas de inferencia buscando llegar al hecho objetivo. En la inferencia descendente, por el contrario, se parte del hecho objetivo y se buscan hechos que lo satisfagan mediante el uso de reglas.

Los dos tipos principales de mecanismos de inferencia computacional se basan en la interpretación de reglas por la teoría de la demostración. Estos son los mecanismos de inferencia ascendente y descendente.

### Evaluación Consultas Recursivas

Una de las características más notables de los sistemas de bases de datos deductivas es su soporte

para la recursión, mediante definiciones de reglas recursivas y por lo tanto, también de consultas recursivas. Una regla es recursiva cuando el predicado de la cabeza aparece también en el cuerpo. Existen diferentes tipos de reglas recursivas:

La mayor parte de las reglas de la vida real se pueden describir como reglas recursivas lineales. Se han definido algoritmos para ejecutar de manera eficiente conjuntos lineales de reglas.

El enfoque del procesamiento de consultas recursivas se clasifican de acuerdo a:

- Su Objetivo en: Evaluación y Reescritura.
- Técnicas de búsqueda en Bottom up y Top Down.

### Técnicas para el procesamiento de consultas recursivas

Las consultas a bases de datos deductivas se pueden resolver mediante reglas no recursivas o reglas recursivas. Este último tipo de reglas presenta el problema de caer en bucles infinitos. Se han propuesto muchos métodos para resolver este problema, los cuales dieron como resultado una gran variedad de algoritmos que agregan la dificultad de elegir el adecuado que se deba aplicar.

En la clasificación de las técnicas de consulta se tienen en cuenta dos aspectos:

- Procesamiento de la consulta, que se puede descomponer en dos métodos:
  - Evaluación: se crea un esquema de evaluación de consultas que dé respuesta a la consulta y queda definida por: el dominio de aplicación y por un algoritmo que resuelva las consultas dadas por un conjunto de reglas.
  - Reescritura: se optimiza el esquema en una estrategia más eficiente, reescribiendo las reglas en función del argumento instanciado que proporciona la consulta.
- Mecanismo de inferencia que se utiliza para deducir hechos a partir de reglas. Los dos tipos principales de mecanismos de inferencia computacional se basan en la interpretación de reglas:
  - Encadenamiento hacia adelante (Bottom Up): el motor de inferencia parte de los hechos y aplica las reglas para generar hechos nuevos, éstos se comparan con el predicado que es el objetivo de la consulta para comprobar si coinciden.
  - Encadenamiento hacia atrás (Top Down): parte del predicado que es el objetivo de la consulta e intenta encontrar coincidencia con las variables que conduzcan a los hechos válidos en la base de datos.

La siguiente tabla muestra los aspectos definidos con anterioridad para la clasificación de las técnicas de consulta con ejemplos de los algoritmos:

### Método de filtrado estático (static filtering)

Actualmente se está investigando, en el marco de este proyecto, el método de Filtrado estático, que introduce una optimización a los Métodos de Evaluación.

El método construye un grafo de axioma\_relación a partir de un programa y una meta.

La optimización se logra imponiendo restricciones, las cuales deben ser cumplidas por las tuplas de la solución, es decir, los filtros se propagan desde el puerto de salida del nodo meta hacia todos los nodos relación, quedando las variables ligadas.

Las tuplas que no cumplan la condición son cortadas, tan pronto como sea posible, en la primera etapa de su flujo hacia el nodo meta. El método finaliza cuando un nodo de axioma no genera nuevas tuplas.

La estrategia asume un proceso de evaluación "Bottom\_Up" a partir de un programa y una meta, construye un grafo llamado grafo axioma\_relación, para ello se efectúa una reescritura de los argumentos para poder diferenciar el mismo argumento, en diferentes ocurrencias del mismo predicado.

### Líneas de investigación y desarrollo

Este proyecto es nuevo, por lo tanto no posee una línea anterior a la cual pertenece. Sin embargo, puede establecerse que su línea de investigación deviene de dos fuentes; por un lado de un proyecto anterior sobre especificación formal y aplicaciones de la Lógica. ; y por otro de la necesidad de dotar de mecanismos de inferencia más eficientes a las bases de datos relacionales.

### Resultados y Objetivos

Como resultados obtenidos se cita el estudio de PROLOG y DATALOG, como lenguajes lógicos de aplicación.

A la fecha, el equipo de investigación se encuentra abocado a la adaptación del método de filtrado estático para realizar consultas recursivas a la Base de Composición de Alimentos (ya implementada).

Como resultados futuros adaptar nuevos métodos de reescritura como por ejemplo reducción de variables y reducción de constantes.

Como objetivo final este proyecto propone desarrollar un sistema organizador y recuperador de la información, dotado de regla eurísticas sobre una base de datos relacional.

### Formación de Recursos Humanos

Este proyecto se inició en Octubre del 2012, a la fecha no hay ninguna tesina de grado ni tesis presentada.

Sin embargo, se estima que para mayo se defienda una tesina (Proyecto Final) de grado; y en el curso de 2013 se culmine con una Tesis de Maestría acerca de Bases de Datos Deductivas.

### Referencias

- E. F. Codd**, Relational Database: a practical foundation for productivity, Commun. ACM 25 (2) (1982) 109\_117.
- M. P. Atkinson, F. Bancilhon, D. J. DeWitt, K. R. Dittrich, D. Maier, S. B. Zdonik**, The Object-Oriented Database System Manifesto., in: the First International Conference on Deductive and Object-Oriented Databases, 1989.
- M. Stonebraker, D. Moore, P. Brown**, Object-Relational DBMSs: Tracking the Next Great Wave, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1998.
- Kuhns, J.** "Logical aspects of question answering by computer" Technical report, Rand Corporation, RM-5428-Pr, 1967.
- Robinson, J. A.** "Logic: Form and Function: The Mechanization of Deductive Reasoning", University Press, Edinburgh, 1978.
- Kowalski, R.** "Logic for Problem Solving" North-Holland, Amsterdam, 1979.
- Codd, E.F.** "A Data Base sublanguage Founded on the Relational Calculus" Proc. of ACM SIGFIDET Workshop on Data Description, Access and Control, San Diego, California. 1971.
- Das, K.** "Deductive Databases and Logic Programming" Addison Wesley, 1992.
- Ramakrishnan, R.; Ullman, J.** "A Survey of Deductive Database Systems" Journal of Logia Programming 1995.
- Liu, M.** "Deductive Database Languages: Problems and solutions" ACM Computing Surveys, 31(1): 27-62, 1999.
- Wagner, G.** "Foundations of Knowledge Systems - with Applications to Databases and Agents", Kluwer Academic Publishers, 1998.
- R. Barbuti, P. Mancarella, D. Pedreschi and F. Turini.** Intensional negation of logic programs: Examples and implementation techniques. In Proc. International Joint Conference on Theory and Practice of Software Development (TAPSOFT'87), pages 96-110. Springer LNCS 250, 1987.
- R. Barbuti, P. Mancarella, D. Pedreschi and F. Turini.** A transformational approach to negation

in logic programming. *Journal of Logic Programming*, 8:201–228, 1990.

- P. Bruscoli, F. Levi, and M.C. Meo.** Compilative constructive negation in constraint logic programs. In *Proc. Colloquium on Trees in Algebra and Programming, (CAAP'94)*, pages 52–67. Springer LNCS 787, 1994.
- Gallaire, H.; Minker, J.** “Logic and Databases” Plenum Press 1978.
- Levesque, H.J.; Lakemeyer, G.** “The Logic of Knowledge Bases” MIT Press, 2001.
- Leondes, C.T. (ed).** “Knowledge-Based Systems. Vols. I-IV” Harcourt International, Academic Press 2000.
- Tansel, A.; Clifford, J.; Gadia, S. et al. (eds.).** “Temporal Databases” Benjamin Cummings, 1993.
- Bouguettaya, Athman (Editor).** “Ontologies and Databases”, Kluwer Academic Publishers, 1999.
- Reiter, R.** 1984. “Towards a logical reconstruction of Relational Database Theory”. *On Conceptual Modelling*. Springer-Verlag.
- Lloyd, J.W.** 1987. “Foundations of logic Programming”. Springer-Verlag.
- Bjørner, Dines.** “The SE Book”, Principle and Techniques of Software Engineering, Volume 3. Technical University of Denmark. February 2, 2004. Springer, Berlin, Heildenberg, New York, London, Paris, Milan Tokyo (free download DB Home Page) versión febrero / marzo 2004



## PERTINENCIAS DE PLANES DE ESTUDIO DE CARRERAS DE INFORMATICA CON NORMATIVAS ESTABLECIDAS POR CONEAU

Lic. Laura Gutiérrez Mag. Raúl Klenzi,  
Departamento Informática (DI) / Facultad de Ciencias Exactas Físicas y Naturales  
(FCEFN) / Universidad Nacional de San Juan (UNSJ)  
Av. Ignacio de la Roza 590 (O), Complejo Universitario "Islas Malvinas", San Juan  
gutierrez.laura, rauloscarklenzi@gmail.com

### Resumen

Este trabajo analiza la similitud entre documentos de texto con los contenidos mínimos de los planes de estudio de las carreras Licenciatura en Ciencias de la Computación (LCC) y Licenciatura en Sistemas de Información (LSI) del Departamento de Informática (DI) de la FCEFN de UNSJ, con las pautas establecidas por CONEAU en la Resolución Ministerial N° 786/2009 con el objetivo de evaluar la pertinencia de los contenidos mínimos de los planes de estudio de las carreras con las áreas mencionadas en dicha resolución.

Se trata de determinar la afinidad o pertinencia de los planes de estudio con respecto a las áreas determinadas en la Resolución 786/2009, mediante la aplicación de la herramienta de software libre RapidMiner (RM) [7] utilizando sus módulos de modelado y minería de texto (TextMining-TM). Esta herramienta, como medida de similitud sintáctica entre documentos, permite la utilización de métricas y tareas de segmentación que serán comparadas desde el punto de vista de la calidad del resultado, con la finalidad de mejorar los planes de estudio de las carreras LCC y LSI del DI de la FCEFN de la UNSJ.

**Palabras clave:** TextMining, pertinencias, planes de estudio.

### Contexto

Esta línea de investigación se enmarca en el proyecto bianual 2011-2012 “**MINERÍA DE DATOS EN LA DETERMINACIÓN DE PATRONES DE USO Y PERFILES DE USUARIO**” código 21/E889 que se desarrolla en el ámbito de la FCEFN-UNSJ, aprobado por el Consejo de Investigaciones Científicas Técnicas y de Creación Artística (CICITCA), financiado por la propia Universidad y ajustado a evaluación externa.

Los datos sobre los que se trabaja en el proyecto, son relativos a las áreas de salud y farmacia como así también al área educación. En esta última área se trabajará con datos generados en el marco de la acreditación de las carreras del DI, del análisis de rendimiento académico de alumnos, planes de estudios de las carreras del citado departamento y datos inherentes a la Biblioteca de la FCEFN.

En este contexto, toda posible mejora en los planes de estudio de las carreras LCC y LSI del DI, será considerada positivamente. La presente propuesta de trabajo y línea de investigación se centra en la determinación de pertinencias de planes de estudio de las carreras LCC y LSI del DI que se dictan en la FCEFN de la UNSJ respecto de las pautas establecidas por CONEAU en la Resolución Ministerial N° 786/2009.

## Introducción

Conforme lo manifestado por la resolución 786/09 del Ministerio de Educación de la Nación quedan definidos los contenidos curriculares básicos correspondientes a las diferentes titulaciones del área de conocimiento informática y computación como se detalla brevemente en las primeras hojas de la mencionada resolución.

La citada resolución aprueba los contenidos curriculares básicos, carga horaria mínima, criterios de intensidad de formación práctica, estándares y nómina de actividades profesionales reservadas para las carreras correspondientes a los títulos de LICENCIADO EN CIENCIAS DE LA COMPUTACION, LICENCIADO EN SISTEMAS /SISTEMAS DE INFORMACION /ANALISIS DE SISTEMAS, LICENCIADO EN INFORMATICA, INGENIERO EN COMPUTACION e INGENIERO EN SISTEMAS DE INFORMACION /INFORMATICA.

Los contenidos curriculares básicos de las diferentes carreras abarcadas por la presente resolución, son evaluados según las definiciones explicitadas en los diferentes anexos (I, II, y III) que conforman dicha resolución.

En el presente trabajo se realiza una comparación entre **los contenidos mínimos de los planes de estudio** de las carreras LCC y LSI del DI de la FCFN de la UNSJ., vigentes hasta el año 2011, con los contenidos de las distintas **áreas** establecidas en la Resolución 786/2009. Se hace uso de técnicas de DM, TM, IR y de medidas de similitud, con el fin de encontrar **pertinencias de contenidos mínimos de planes de estudio con**

**respecto a las áreas**, para ello se utiliza el módulo Text Processing de la herramienta de software RM que permite llevar a cabo las tareas de preprocesamiento y determinación de pertinencias.

## Líneas de investigación y desarrollo

El presente trabajo se encuentra enmarcado en el proyecto “Minería de datos en la determinación de Patrones de uso y perfiles de usuarios”, las líneas de investigación que se detallan a continuación y siempre con el afán de que la investigación aplicada ayude a la toma de decisiones de la autoridad competente. Por un lado y desde el relevamiento de una encuesta asociada a un constructo que permite aproximar la capacidad de resiliencia, se ha encuestado a alumnos avanzados e ingresantes de las carreras del departamento informática con el objetivo que, mediante tareas de modelación se pueda aproximar el perfil de un alumno rezagado o posible desertor y que en función de ello las autoridades de la FCFN-UNSJ arbitren las medidas necesarias a nivel de tutorías o apoyaturas docentes y de pares que atenúen la deserción y el rezago.

Por otro y desde la aplicación de diferentes técnicas de minería de datos DM y text mining procesar títulos bibliográfico de la biblioteca midiendo la similitud sintáctica de los mismos con los contenidos de las diferentes carreras que se dictan en la FCFN-UNSJ y proponiendo una primera aproximación a la determinación de código Dewey para aquella bibliografía que no lo posee o es edición, en español, de la propia universidad.

El uso de métricas de minería de texto es también un camino seguido en la presente propuesta de trabajo y es un eslabón que pretende ir cerrando la cadena de aplicaciones que en el tema de satisfacción de usuarios de bibliotecas universitarias y desde el procesamiento de la encuesta a usuarios alumnos y docentes de la biblioteca encaró el grupo de investigadores pertenecientes al proyecto.

En la mayoría de los casos los trabajos se han llevado adelante mediante la utilización de herramientas de software libre del área de minería de datos.

En este caso los resultados que se presentan se han alcanzado mediante el uso de algoritmos de DM que posee la herramienta RapidMiner cuya última versión es la 5.3.005.

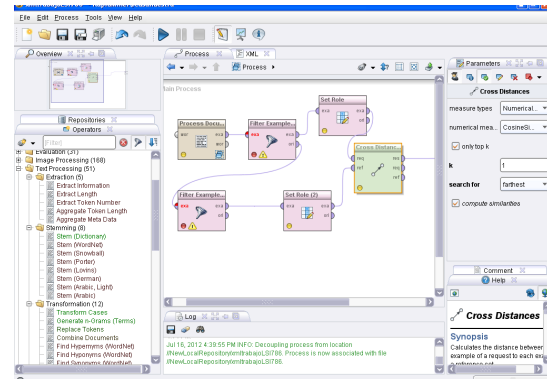
Para este trabajo se propone para contrastar documentos la técnica de la similaridad del coseno, que no es otra cosa que el coseno del ángulo que forman un vector consulta  $q$  (contenido mínimo) y un vector documento  $d_j$  (áreas de la 786/2009). [4]

A partir de una consulta se obtiene una lista de documentos ordenados por distancia (los más relevantes primeros). Luego, se procede a realizar los cálculos algebraicos para determinar la semejanza entre el vector consulta y cada uno de los vectores que representan documentos de la colección. [5]

### Preprocesamiento de datos

Al comenzar este trabajo se contaba con los contenidos mínimos de los planes de Estudio de las carreras LCC y LSI del DI en archivos .doc y la Resolución 786/2009 en un archivo .pdf. Lo primero que se realizó fue separar cada contenido mínimo de los distintos planes de estudio en archivos .txt, como así también se

llevaron a archivos .txt cada una de las áreas de la Resolución 786/2009.



Figural: Entorno de RapidMiner utilizando entre otros el módulo de Textmining.

En la Figural se observa las diferentes áreas en que se divide el entorno visual de la herramienta Rapidminer, y desde donde se pueden elegir los operadores a aplicar y definir los parámetros de los mismos.

Los documentos, que contienen los contenidos mínimos de planes de estudio de las carreras LCC y LSI (Consulta o request **req**) y los que contienen las áreas establecidas por CONEAU en la Resolución 786/2009 (Base de Datos de Referencia **ref**) son preprocesados por un módulo de RM [3]

Tras la instancia de preprocesamiento los documentos se separan en contenidos mínimos por un lado (Consulta o Request, **req**) y áreas establecidas en la Resolución 786/2009 (Referencia, **ref**) por otro. Esta separación o filtrado permite, desde el módulo (Cross Distances), la aplicación de métricas de similitud (la del coseno) entre documentos.

La métrica de similitud considerada, posee un rango de valores posibles que oscila en forma continua entre 0 o 1, cuando los documentos comparados son

sintácticamente diferentes, y 1 cuando reflejan una similitud total. [3]

En la Figura2, se observa lo dicho anteriormente.

Row No.	request	document	distanc
1	24- BASE DE DATOS AVANZADA	Ingeniería de Software Base de Datos y Sistemas de Información	0.129
2	4- INFORMACIÓN Y SISTEMA	Ingeniería de Software Base de Datos y Sistemas de Información	0.125
3	7- ASPECTOS PROFESIONALES Y SOCIALE	Aspectos profesionales y Sociales	0.122
4	21- PRINCIPIOS DE INGENIERIA DE SOFTWARE	Ingeniería de Software Base de Datos y Sistemas de Información	0.112
6	19- BASE DE DATOS	Ingeniería de Software Base de Datos y Sistemas de Información	0.112
5	25- DISEÑO DE SOFTWARE	Ingeniería de Software Base de Datos y Sistemas de Información	0.070
7	14- ESTRUCTURA Y FUNCIONAMIENTO DE LAS COMPUTADORAS I	Arquitectura, Sistemas Operativos y Red	0.061
8	27- ARQUITECTURAS AVANZADAS DE COMPUTADORA	Arquitectura, Sistemas Operativos y Red	0.056
9	29- REDES	Arquitectura, Sistemas Operativos y Red	0.047
10	3- ESTRUCTURA Y FUNCIONAMIENTO DE LAS COMPUTADORAS	Arquitectura, Sistemas Operativos y Red	0.044
11	19- INTRODUCCIÓN A LOS SISTEMAS DIGITALES I	Arquitectura, Sistemas Operativos y Red	0.031
12	8- INTRODUCCIÓN A LOS SISTEMAS DIGITALES	Arquitectura, Sistemas Operativos y Red	0.023
13	13- ESTRUCTURAS DE DATOS Y ALGORITMOS	Ingeniería de Software Base de Datos y Sistemas de Información	0.021
14	26- INTELIGENCIA ARTIFICIAL	Arquitectura, Sistemas Operativos y Red	0.019
15	9- PROGRAMACION ORIENTADA A OBJETOS	Arquitectura, Sistemas Operativos y Red	0.015
16	5- PROGRAMACION PROCEDURAL	Ingeniería de Software Base de Datos y Sistemas de Información	0.015
17	29- COMPILADORE	Arquitectura, Sistemas Operativos y Red	0.013
18	23- TEORIA DE LA COMPUTACION	Aspectos profesionales y Sociales	0.011
19	16- MATEMATICA DISCRET	Arquitectura, Sistemas Operativos y Red	0.010
20	1- ALGORITMOS Y RESOLUCIÓN DE PROBLEMA	Ingeniería de Software Base de Datos y Sistemas de Información	0.010
21	22- INOLESI	Ingeniería de Software Base de Datos y Sistemas de Información	0.010
22	16- PARADIGMAS DE LENGUAJE	Arquitectura, Sistemas Operativos y Red	0.009
23	3- MATEMATICA BASIC	Aspectos profesionales y Sociales	0.009

Figura2 LCCcon-min LCC786

## Resultados y Objetivos

Con los objetivos iniciales de comparar los planes de estudio de las carreras LCC y LSI del DI con los contenidos de las áreas establecidas por CONEAU en la Resolución Ministerial N° 786/2009, utilizando el módulo de Text Processing de RM, y encontrar pertinencias se han logrado los siguientes resultados preliminares:

- De la utilización de la medida de similitud del Coseno se pudo determinar cuán pertinente es el contenido mínimo de los planes de estudios de las carreras LCC y LSI con las áreas establecidas en la Resolución Ministerial N° 786/2009, ya que ante documentos cuya pertinencia se aproxima a 1 indica lo cercano de los contenidos mínimos con las áreas consideradas en la Resolución Ministerial N° 786/2009, sus valores fluctúan entre 0 (sin coincidencia) y 1 (coincidencia total).
- Se logró una buena aproximación entre la tarea de segmentación que encuentra los grados de pertinencia y las áreas

consideradas por los docentes en la realización de los planes de estudio.

Se pretende desde este proyecto profundizar este análisis de pertinencias semánticamente, utilizando herramientas de textmining proporcionadas por RM, haciendo uso del módulo del Diccionario de Wordnet para sinónimos, el cual se encuentra en inglés y como nuestro dominio de trabajo es en español, se está trabajando en los pasos a seguir

## Formación de Recursos Humanos

La formación de recursos humanos es un tema de vital importancia en marco del presente proyecto de investigación. En este marco y en temas afines a las temáticas abordadas en el mismo, se están dirigiendo un conjunto de trabajos finales de grado de alumnos becarios de becas de finalización de carrera de la Agencia Nacional de Promoción Científica y Tecnológica.

En particular trabajos finales de grado y tesis de posgrado, se está trabajando con datos de títulos bibliográficos pertenecientes a la biblioteca de la FCFN-UNSJ a los que en forma automática se intenta asignar la numeración dewey asociada a la temática abordada por el material bibliográfico bajo estudio. Así mismo y desde una tesis de posgrado, se intenta encontrar automáticamente también, y mediante similitud sintáctica qué contenidos mínimos establecidos en la resolución 786/09 son contenidos y brindados por los planes de estudios de las carreras de informática de la FCFN-UNSJ



**Licenciatura en Sistemas de Información**

- *Estrategias de marketing web y minería de datos para promover tráfico de calidad hacia un sitio web.* Rafa
- *Minería web a los datos de acceso a sitios pertenecientes a centros de información académica.* Vero
- *Minería de Datos en la segmentación y clasificación de un banco de germoplasma.* Kari

**LCC**

- *Minería de Texto en la determinación automática de código Dewey.* J.Araya
- *Analítica web en centros de información.* S. Lobo
- *Desarrollo de un sistema de información Intranet para el Inst. de Ing. Química-UNSJ.* A. Sepúlveda

**LCI**

- *Técnicas de Minería de Texto en la determinación de la correlación entre valoración numérica y comentarios asociados en encuestas.*

A su vez se están dirigiendo dos tesis de posgrado en el área de Minería de Texto y el área de Inteligencia de Negocios respectivamente.

**Referencias**

- [1] Kantardzic, M (2003) “Data Mining: Concepts, Models, Methods, and Algorithms” ISBN:0471228524 John Wiley & Sons © (343 pages)
- [2] Klenzi, R. Tesis de posgrado de maestría “Aplicación de minería de datos a la gestión bibliotecaria”. Biblioteca FCFN-UNSJ. 2008.
- [3] Klenzi, R, Gutiérrez L., Villafañe V. “Técnicas de recuperación de información en la determinación de pertinencias bibliográficas” 2012.
- [4] Liu B., “Web DataMining. Exploring Hyperlinks, Contents, and Usage Data” Springer-Verlag Berlin Heidelberg 2007
- [5] Manning C, Prabhakar R. Hinrich & Hinrich Schütze, “An Introduction to Information Retrieval ”, Cambridge University Press. 2009.
- [6] Min, S; Yi-Fang B. Handbook of Research on Text and Web Mining Technologies -Information science reference- Editorial Advisory Board 2009.
- [7] Rapid-I. <http://rapid-i.com> ver.5.3.005 de 2013.
- [8] Tolosa G y Bordignon F., “Introducción a la Recuperación de Información. Conceptos, modelos y algoritmos básicos” UNLu, Arg. 2007.

## RELACIONANDO COMENTARIOS TEXTUALES Y VALORES NUMÉRICOS EN ENCUESTA DE SATISFACCIÓN DE USUARIOS

Mag. Raúl Klenzi, Lic. L. Gutierrez, Alum. Tamara Pinto  
Instituto de Informática (IdeI) / Departamento Informática (DI) / Facultad de Ciencias Exactas Físicas y Naturales (FCEFN) / Universidad Nacional de San Juan (UNSJ)  
Av. Ignacio de la Roza 590 (O), Complejo Universitario "Islas Malvinas", San Juan  
rauloscarklenzi, gutierrez.laura, tamara932 @gmail.com

### Resumen

En este trabajo se procesa información extraída de encuesta de satisfacción de usuarios alumnos de bibliotecas universitarias. En este contexto la encuesta consiste en valorar 17 atributos según una escala Likert de 5 estados asociados a conceptos que van desde insatisfecho (1) a totalmente satisfecho (5) respectivamente. Además, y como parte de la encuesta, cada encuestado puede expresar en formato de texto comentarios adicionales. El objetivo central de este trabajo consiste en relacionar los atributos inductores que describen la encuesta desde las respuestas numéricas, con aquellos que se obtienen desde un análisis de Text Mining (TM) aplicado a los comentarios. La encuesta procesada, como caso de estudio, se realizó a alumnos de la biblioteca de la Facultad de Ingeniería de la Universidad Nacional de San Juan (FI-UNSJ)

**Palabras clave:** Data Mining, Text Mining, Satisfacción de Usuarios.

### Contexto

En el marco de proyectos anteriores y del actual "Minería de datos (DM) en la Determinación de Patrones de Uso y Perfiles de usuarios" 21/E889 se realizaron encuestas a los usuarios de Bibliotecas tratando de encontrar su grado de satisfacción respecto de los servicios y funcionamiento de las mismas.

El constructo que permitió realizar la encuesta, se conforma de 17 atributos, asociados a diferentes servicios que ofrece la biblioteca, que deben ser respondidos, con valoraciones numéricas entre 1 y 5 según sus percepciones. A la vez se permite, a los usuarios, escribir todo aquello que considere válido a analizar por parte de los encuestadores en formato de texto libre a modo de comentario.

Como tareas inherentes al proyecto se han procesado las encuestas encontrando los inductores (atributos más relevantes) que conforman la imagen que los usuarios tienen de su biblioteca. La instancia a considerar en la presente propuesta, es procesar los comentarios que se redactaron por parte de los usuarios mediante técnicas de TM.

El propósito de este procesamiento será contrastar, para aquellas encuestas que contienen comentarios, la coincidencia o complementación del conocimiento extraído de la valoración numérica realizada por los usuarios a los diferentes atributos.

### Introducción

Muchas encuestas en las que los atributos por los que se consulta deben ser respondidos en forma discreta y numérica, permiten en un apartado de la misma, que el encuestado brinde una opinión en formato texto que amplíe o aclare lo expresado en sus respuestas numéricas.

Este es el caso del constructo validado en [1], el cual se relevó a usuarios alumnos de la biblioteca de la FI-UNSJ. Estos atributos cubren entre otros, aspectos edilicios, de personal, de material bibliográfico y tecnológico.

La hipótesis, que mueve la presente propuesta, es que todo comentario expresado en formato de texto en principio se considera una “queja” o “reclamo” sobre alguna condición de la biblioteca que, a criterio del encuestado, puede mejorarse y que por lo tanto deberá estar relacionada con aquellos atributos menos valorados numéricamente por el mismo.

Mediante tareas de segmentación aplicadas a las respuestas numéricas de las encuestas se dividen las respuestas en dos grupos. Por un lado el grupo de los encuestados satisfechos con las prestaciones de la biblioteca y por otro lado el grupo de aquellos usuarios que consideran que la biblioteca puede mejorar sus servicios. Tras esta agrupación y desde una tarea de clasificación se obtienen los atributos inductores que describen la encuesta detectando además, cuáles de éstos definen la pertenencia a uno u otro segmento.

Por otro lado, mediante tareas de TM y métricas de similitud sintáctica entre el nombre de los atributos y los comentarios expresados en cada encuesta, se encontrará información complementaria sobre los atributos evaluados.

Con el objetivo de verificar la hipótesis manifestada anteriormente, es de esperar que aquellos atributos inductores obtenidos desde el procesamiento numérico y que definen al segmento de “posibles mejoras” coincidan con los de mayor similitud sintáctica, encontrados desde el procesamiento de los comentarios.

## Líneas de investigación y desarrollo

El proyecto en el que está enmarcado el presente trabajo, ha llevado adelante durante el último año, diferentes líneas de investigación con el afán de que la investigación aplicada ayude a la toma de decisiones de la autoridad competente.

Así, desde la aplicación de diferentes técnicas de DM y TM se están procesando títulos bibliográficos de la biblioteca midiendo la similitud sintáctica de los mismos con los contenidos de las diferentes carreras que se dictan en la FCFN-UNSJ tratando de proponer una primera aproximación a la determinación de código Dewey para bibliografía que no lo posee o es edición, en español, de la propia universidad.

El uso de métricas de TM es también un camino seguido en la presente propuesta de trabajo y es un eslabón que pretende ir cerrando la cadena de aplicaciones que, en el tema de satisfacción de usuarios de bibliotecas universitarias y desde el procesamiento de la encuesta a usuarios alumnos y docentes de la biblioteca, encaró el grupo de investigadores pertenecientes al proyecto.

En la mayoría de los casos los trabajos se han llevado adelante mediante la utilización de herramientas de software libre del área de DM. Los resultados se han alcanzado mediante el uso de algoritmos de DM que posee la herramienta RapidMiner cuya última versión es la 5.3.005 [5].

### Desarrollo

La aplicación consta de tres pasos:

- 1) El procesamiento de datos correspondientes a la fracción numérica de la encuesta, consiste en la aplicación sucesiva del algoritmo de segmentación W-SimpleKmeans a un grupo de 46

encuestas de alumnos, de las 150 relevadas, que poseen comentarios y que permite segmentar y etiquetar en dos grupos, las opiniones de los encuestados. Seguidamente el algoritmo clasificador WJ48 permite describir la encuesta con sus etiquetas asociadas, con la menor cantidad de atributos, denominados inductores reconociendo además, mediante aplicaciones sucesivas del operador Single Rule Induction, cuáles de ellos definen cada segmento.

2) Esta parte del desarrollo consiste en encontrar una medida de similitud sintáctica entre el comentario de una encuesta denominada consulta o **request** y cada uno de los nombres asociados a los atributos de la encuesta denominado **referencia**, realizando además, un mínimo análisis de sinonimias.

3) El tercer y último paso consiste en analizar comparativamente los resultados obtenidos en 1) y 2) tratando de verificar la hipótesis.

La elección de los algoritmos se fundamenta en la necesidad de procesar registros incompletos permitido por W-SimpleKmeans y W-J48 como así también la rápida interpretación de resultados que posee una estrategia de árboles basada en Ganancia de Información Relativa [4].

La tarea de segmentación llevada adelante en 1) asignó 28 registros al cluster0 y 18 al cluster1. A los efectos de asignar un significado a cada cluster, la Tabla 1 presenta el valor de sus respectivos centroides. Allí se aprecia, según el valor de los centroides, que los usuarios tienen una buena percepción de los servicios ofrecidos por la biblioteca, dado que ambos valores están por encima de la media. De todas maneras se puede considerar al segmento con menor valor de centroide, como aquel en el que los

encuestados consideran que la biblioteca es factible de mejorar en sus prestaciones.

Row No.	SEGMENTO	CLUSTER	average(CENTROIDE)
1	POSIBLES MEJORAS	cluster1	3.0893518518518515
2	SATISFACTORIO	cluster0	3.8625550220088027

Tabla 1

Tras la tarea de segmentación y desde la aplicación del algoritmo clasificador se obtuvo el siguiente árbol

### **W-J48**

#### **J48 unpruned tree**

```

CANTIDADPERSONAL <= 2: POSIBLES MEJORAS (8.0)
CANTIDADPERSONAL > 2
| CANTIDADMATERIAL <= 2: POSIBLES MEJORAS (6.16/0.16)
| CANTIDADMATERIAL > 2
| | IDONPERSONAL <= 3
| | | SERVPRESTAMODOMICILIO <=3:SATISFACTORIO (2.0)
| | | SERVPRESTAMODOMICILIO >3:POSIBLESMEJORAS (4.0)
| | IDONPERSONAL > 3: SATISFACTORIO (25.84)

```

Number of Leaves : 5  
Size of the tree : 9

#### Árbol de clasificación descriptor de la encuesta

En el clasificador anterior se observa que solamente 4 atributos, de los 17 iniciales, denominados inductores, describen la totalidad de la encuesta. Con estos cuatro atributos inductores como universo, se aplican sucesivamente algoritmos del tipo Single Rule Induction (Single Attribute), cuyo objetivo es describir la encuesta, con el menor error posible, mediante la consulta de un único atributo y paso seguido se elimina el atributo encontrado.

De los atributos inductores el que mejor describe la encuesta es Idoneidad de personal (IDONPERSONAL) que asigna correctamente 39 de los 46 registros. Eliminado IDOPPERSONAL, de los restantes, es Cantidad de material (CANTIDADMATERIAL) el que describe mejor la encuesta con 37 registros asignados correctamente de los 46. La secuencia de pasos antes descriptos y hasta agotar los atributos inductores permite obtener la Tabla 2.



ATRIBUTO	SEGMENTO	
	SATISFAC-TORIO	POSIBLES MEJORAS
IDONPERSONAL 39/46 (84,78%)	26/31 (83,87%)	13/15 (86,66%)
CANTIDADMATERIAL 37/46 (80,43%)	28/37 (75,67%)	9/9 (100%)
CANTIDADPERSONAL 36/46 (78,26%)	28/38 (73,68%)	8/8 (100%)
SERVPRESTAMO- DOMICILIO 28/46 (60,86%)	28/46 (60,86%)	0

Tabla 2

En la Tabla 2 se observa que los atributos CANTIDADMATERIAL y CANTIDADPERSONAL describen con mayor exactitud el segmento “POSIBLES MEJORAS”

Para la realización de 2), en trabajos previos presentados en WICC 2012 [2] [3], se constató que entre las diferentes métricas de similitud utilizables y que posee la herramienta de software disponible, la del coseno es la que mejores resultados brinda. En este caso los valores de similitud (distance) varían entre 0 (**req** y **ref** sin similitud) y 1 en que ambos son iguales. Así mismo y dado que el comentario escrito por el encuestado está en formato libre, es necesario generar una base de datos de sinónimos que permita buscar aproximaciones entre el **request** y la **referencia**. Si bien la herramienta RM posee un módulo de wordnet que facilita las tareas de procesamiento de sinónimos y significado de palabras, el mismo por el momento funciona correctamente para el idioma inglés. Por ello la tarea de procesamiento de sinónimos se embebió en el módulo (Process document from file Replace Token) de la propia herramienta como se muestra en la Figura 1.

En la Figura 1 se aprecia que palabras como *avisador* o *anunciador*, que pudieran aparecer en los comentarios, automáticamente se reemplazan, en tareas de preprocesamiento, por *cartelera* que es una palabra contenida en algún nombre de atributo de la encuesta.

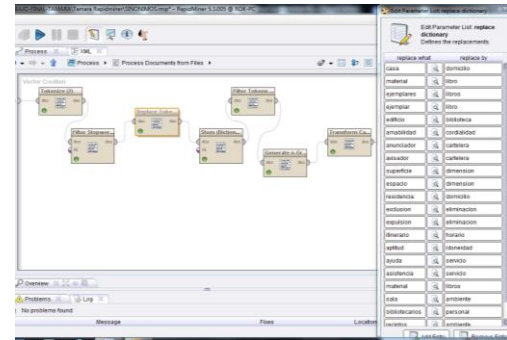


Figura 1

La aplicación de la secuencia de módulos que se presenta en Figura 2 permitió obtener las métricas, entre nombres de atributos y comentarios, como se observa en la Tabla 3.

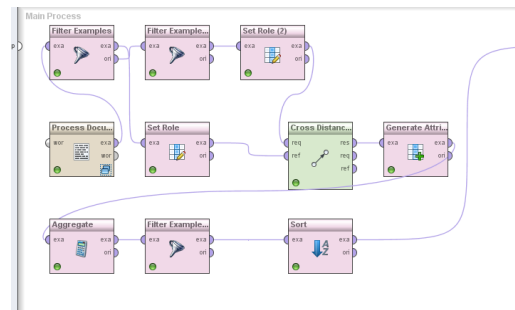


Figura 2

En la Tabla 3 se aprecia que para el comentario “Con respecto a los días sábado...” la mayor similitud (distance) se da con el atributo Horario de la Biblioteca y solamente 4 de los 17 atributos presentan un valor de similitud (distance) mayor que cero.

request	document	distance
23) con respecto al día sábado debería estar abierto más tiempo si no fueran personal podían bajar alumnos avanzados	horario en que se encuentra abierta la biblioteca	0,124219716688
48) con respecto al día sábado debería estar abierto más tiempo si no fueran personal podían bajar alumnos avanzados	cantidad del personal	0,024238196827
10) con respecto al día sábado debería estar abierto más tiempo si no fueran personal podían bajar alumnos avanzados	identidad del personal	0,020230353115
102) con respecto al día sábado debería estar abierto más tiempo si no fueran personal podían bajar alumnos avanzados	información en cartelera	0,018346467991
20) con respecto al día sábado debería estar abierto más tiempo si no fueran personal podían bajar alumnos avanzados	cartelera del material	0
24) con respecto al día sábado debería estar abierto más tiempo si no fueran personal podían bajar alumnos avanzados	cantidad de material	0
25) con respecto al día sábado debería estar abierto más tiempo si no fueran personal podían bajar alumnos avanzados	servicio de préstamo en casa	0
26) con respecto al día sábado debería estar abierto más tiempo si no fueran personal podían bajar alumnos avanzados	equipo tecnológico	0
27) con respecto al día sábado debería estar abierto más tiempo si no fueran personal podían bajar alumnos avanzados	servicio de reserva de material	0
28) con respecto al día sábado debería estar abierto más tiempo si no fueran personal podían bajar alumnos avanzados	servicio de consulta de búsqueda de material	0
29) con respecto al día sábado debería estar abierto más tiempo si no fueran personal podían bajar alumnos avanzados	tiempo de completación de reportes/métricas	0
30) con respecto al día sábado debería estar abierto más tiempo si no fueran personal podían bajar alumnos avanzados	servicio de préstamo a domicilio	0
31) con respecto al día sábado debería estar abierto más tiempo si no fueran personal podían bajar alumnos avanzados	dimensión de la biblioteca en general	0
32) con respecto al día sábado debería estar abierto más tiempo si no fueran personal podían bajar alumnos avanzados	distribución de espacios	0
33) con respecto al día sábado debería estar abierto más tiempo si no fueran personal podían bajar alumnos avanzados	ubicación de la biblioteca	0
34) con respecto al día sábado debería estar abierto más tiempo si no fueran personal podían bajar alumnos avanzados	ambiente de estudio dentro de la biblioteca	0

Tabla 3

Finalmente la aplicación desarrollada en RM presenta los resultados, que a modo de resumen figuran en la Tabla 4

document	ATRIBUTOQUEJA	count(CUENTA)
cantidad de material	SI	23
cantidad de personal	SI	23
servicio de ayuda de busqueda de material	SI	23
calidad del material	SI	22
horario en que se encuentra abierta la biblioteca	SI	22
servicio de reserva de material	SI	22
ambiente de estudio dentro de la biblioteca	SI	18
cordialidad del personal	SI	16
idoneidad del personal	SI	16
dimension de la biblioteca en general	SI	14
ubicacion de la biblioteca	SI	14
servicio de prestamo en sala	SI	6
servicio de prestamo a domicilio	SI	5
distribucion de espacios	SI	1
informacion en cartelera	SI	1

Tabla 4

La Tabla 4 muestra que la mitad de los comentarios (23) tienen una medida de similitud distinta de cero con atributos referidos a: cantidad de material, cantidad de personal y servicio de ayuda de búsqueda de material, respectivamente. Esto verifica la hipótesis que justamente los atributos inductores del segmento “posibles mejoras” coinciden con los encontrados desde las métricas de similitud entre nombre de atributo y comentario.

## Resultados y Objetivos

La aplicación permitió comprobar la hipótesis de que los comentarios, mayormente, expresan disconformidades de los usuarios para con un servicio bajo evaluación.

En el marco de trabajos futuros y en el contexto de la aplicación, se pretende extender el análisis a encuestas relevadas en otras bibliotecas de la UNSJ y ampliarla utilizando el módulo de RM de wordnet adaptado a lenguaje español que acrecentaría la posibilidad de mejorar métricas y eliminaría el trabajo cuasi manual de reemplazos implementado, comenzando a transitar el camino desde el análisis sintáctico al análisis semántico.

## Formación de Recursos Humanos

La formación de recursos humanos es un tema de vital importancia. En este marco, se están dirigiendo un conjunto de trabajos finales de grado de becas de finalización de carrera de la Agencia Nacional de Promoción Científica y Tecnológica.

En particular, se está trabajando con datos de títulos bibliográficos pertenecientes a la biblioteca de la FCFN-UNSJ a los que en forma automática se intenta asignar la codificación Dewey correspondiente. Así mismo y desde una tesis de posgrado, se intenta encontrar automáticamente también, y mediante similitud sintáctica qué contenidos mínimos establecidos en la resolución 786/09 son contenidos brindados por los planes de estudios de las carreras de informática de la FCFN-UNSJ

Al momento, integrantes del proyecto dirigen 7 trabajos de grado y 2 de posgrado en temáticas afines a las tratadas en el proyecto.

## Referencias

- [1] Beguerí, G. “Logística como garantía de satisfacción del usuario”. Tesis de maestría- Universidad Nacional de Cuyo. (2007)
- [2] Klenzi, R., Gutierrez, L., Villafañe V. “Técnicas de Recuperación de Información en la Determinación de Pertinencias Bibliográficas”. WICC-2012.
- [3] Liu B., “Web DataMining. Exploring Hyperlinks, Contents, and Usage Data” Springer-Verlag Berlin Heidelberg 2007
- [4] North M. “Data Mining for the Masses” ISBN: 0615684378. A Global Text Project Book. Ed 2012.
- [5] <https://rapid-i.com/content/>

## MINERIA DE DATOS EN REDES EDUCATIVAS

Marcelo Omar Sosa, Sosa Bruchmann Eugenia Cecilia, Raúl Marcelo Vega

Departamento Computación/Facultad de Ciencias Exactas y Naturales/Universidad Nacional de Catamarca

Av. Belgrano N° 300 - Planta alta - C.P: 4700 - San Fernando del Valle de Catamarca

Telefono: 0383- 4425610

[sosamod1@hotmail.com](mailto:sosamod1@hotmail.com), [sosab\\_ec@hotmail.com](mailto:sosab_ec@hotmail.com), [vega\\_rm@hotmail.com](mailto:vega_rm@hotmail.com)

### Resumen

La minería de datos (*data mining*) viene ampliando sus áreas de aplicación, demostrando la validez de sus descripciones y predicciones en el análisis de datos. Las nuevas áreas en las que se vienen desarrollando estudios se relacionan directamente con el fenómeno de los últimos años como son las redes sociales. Siguiendo las características de estas últimas se crearon redes como lo son las redes educativas, las que se conforman en estratos académicos, que pueden ser de tamaño muy grandes si se trata a nivel institucional o considerable como las que se desarrollan en una asignatura. Sin tener en cuenta este aspecto, dichas redes, generan gran cantidad de datos debido a la interacción de sus miembros que pueden ser analizados aplicando algoritmos y técnicas de minería de datos con el objetivo de encontrar información relevante para mejorar el proceso de enseñanza y aprendizaje. La utilización de las Nuevas tecnologías de la información y la comunicación (NTIC) en la educación trae aparejado la generación de bases de datos educativas que al aplicarse los algoritmos de minería de datos conforman una nueva área de estudio la que se la denominada minería en redes educativas (*educational networks mining*).

**Palabras clave:** Minería de datos; Redes educativas.

### Contexto

El presente trabajo sienta las bases para el estudio y desarrollos en el marco de tesis doctoral y de maestría en Ciencias de la

Computación. Se desarrolla en la Facultad de Ciencias Exactas y Naturales (*Fa.C.E.N.*) de la Universidad Nacional de Catamarca (*U.N.Ca.*) en convenio con la Universidad Nacional del Sur (*U.N.S.*). Los autores se desempeñan como docentes de las carreras Profesorado en Computación y Tecnicatura en informática en sus distintas orientaciones en la *U.N.Ca.*.

### Introducción

La utilización generalizada de las NTIC en los distintos niveles de la educación y las redes sociales como un ámbito para el desarrollo de relaciones personales conlleva a la aparición de redes particulares con fines específicos como son las redes educativas. Las instituciones que promueven la incorporación sistemática de las NTIC también desarrollan las denominadas redes de educación cuyas bases se asientan en las de las redes sociales. Estas generan grandes volúmenes de datos creado por las interacciones de los miembros los que se pueden agrupar por los niveles en los cuales se producen. Pueden clasificarse según el tipo de relación (uno a uno- uno a varios y varios a varios) y además de la jerarquía establecida por el contacto con el docente (docente-alumno, docente-alumnos, alumnos-alumnos)

La creciente disponibilidad de información organizada en bases de datos sobre todo en áreas como la educación, son un campo propicio para la aplicación de algoritmos y técnicas de minería de datos. La conjunción de áreas como la educación y la Ciencia de la Información conforman un área de estudio denominada minería en redes educativas (*educational networks mining*) [1,2].

Esto grandes volúmenes requieren un tratamiento especial para ser procesados, inicialmente deben seleccionarse los datos relevantes para el estudio [2]. A continuación determinar las actividades realizadas por cada integrante de la red. En el caso particular de las redes educativas, el usuario debe ser identificado (es decir no puede ser anónimo) [4] para relacionar el comportamiento y las actividades desarrolladas por cada integrante de la red con las paginas e informaciones disponibles. La registración de estos comportamientos en las estadísticas aporta otros tipos de datos como lo son el tiempo de permanencia en la red, cantidad de páginas visitadas, software utilizado en las consultas, participación en foros, aportes a la WIKI entre otros que pueden también ser analizadas utilizando técnicas de minería de datos[3].

Finalmente se debe realizar dos procesos que se consideran los más importantes, estos son: la integración y la reducción de los datos [2]. Una vez realizados estos procesos, se puede aplicar diferentes algoritmos de minería de datos a fin de obtener información relevante para la determinación de patrones, perfiles psicológicos y de evolución cognitiva de cada integrante de la red educativa.

Además dentro de la investigación a realizar está la de comparar e identificar diversas técnicas de minería de datos a ser aplicadas en el análisis de los datos obtenidos, ajustándolas a los objetivos propios del estudio [5].

## Líneas de investigación y desarrollo

El trabajo se enmarca en la investigación de aplicación de técnicas y algoritmos de minería de datos en bases de datos educativas. Con el principal objetivo de cuantificar la efectividad del proceso de enseñanza y aprendizaje, organizar adecuadamente el contenido de la asignatura, aplicar el agrupamiento de alumnos de acuerdo a los perfiles encontrados, entre otros.

El presente trabajo está íntimamente relacionado con trabajos desarrollados en el área y próximos a ser presentados en encuentros de investigadores en el tema.

A partir de estos trabajos se analiza la posibilidad de desarrollar tesis de posgrado.

## Resultados y Objetivos

Los resultados esperados que guían a este trabajo son los de determinar las técnicas y algoritmos de data mining más adecuados para analizar el gran volumen de información generada en este tipo de redes, la que sería imposible de realizar en forma manual. Además de la obtención de información relevante para la toma de decisiones.

Tiene como objetivos principales:

- Determinar la efectividad del proceso de enseñanza y aprendizaje.
- Establecer los patrones de comportamiento de los alumnos.
- Evaluar las actividades particulares realizadas por los integrantes de un determinado curso.
- Encontrar errores frecuentes en los desarrollos de actividades previstas en el curso.
- Establecer patrones de comportamientos de los participantes del curso.
- Ajustar los contenidos en forma adecuada para lograr un mejor entendimiento de lo conceptos.
- Encontrar características similares en los integrantes del curso.
- Establecer preferencias en los métodos desarrollados para la transmisión de contenidos.
- Evaluar cuantitativamente las metodologías pedagógicas utilizadas.
- Favorecer la motivación en los alumnos.
- Mejorar el diseño de las actividades previstas.
- Diseñar líneas de investigación que relaciones el área de minería de datos con la Ciencias de la Computación.



## Formación de Recursos Humanos

Los autores del trabajo se encuentran en la etapa de desarrollo de sus tesis de posgrado en diferentes carreras relacionadas con el temas de investigación, tales como la Maestría en Docencia Universitaria en disciplinas tecnológicas dictada en la Facultad de Ciencias Agrarias – U.N.Ca, el cursado para la actualización de contenidos de la carreras Especialización en educación tecnológica dictada por la Facultad de Ciencias Exactas y Naturales – U.N.Ca.

La Docente Investigadora Lic. Eugenia Cecilia Sosa Bruchmann desarrolla sus actividades docentes en la cátedra Técnicas Digitales de las carreras de Profesorado en Computación y Tecnicatura en Informática en sus distintas orientaciones.

El Ing. Marcelo Omar Diógenes Sosa se encuentra en la etapa de planificación de su tesis doctoral en el área de minería de datos en el Doctorado en Ciencias dictado en la Facultad de Ciencias Exactas y Naturales en convenio con la Universidad Nacional del Sur.

Durante el año 2012 se desarrollo la tesis del alumno Jorge Paseto en la carrera de Licenciatura en Tecnología Educativa denominada: "Los recursos comunicacionales como recursos en la educación a distancia"

El trabajo en el área de minería de datos establece en su diseño la formación de recursos humanos en diferentes niveles como el de tesinas de grado y tesis de posgrado para los integrantes del equipo de investigación. Dentro de las actividades podemos destacar:

- La creación de un grupo de alumnos colaboradores, que desarrollan investigación con acompañamiento de docentes.
- La selección de temas posibles para tesis en el área relacionadas con la investigación.
- La producción científica para su presentación en congresos locales, nacionales e internacionales.

- Participación de los integrantes en cursos de actualización y posgrado en el área de estudio.
- La actualización y capacitación permanente de los investigadores en talleres o workshop relacionadas con el tema del trabajo.
- La participación de los investigadores como consultores en proyecto afines que se desarrollan en la Facultad de Ciencias exactas y Naturales en distintas áreas.
- La planificación de seminarios para docentes en temas relacionados con la investigación y resultados obtenidos en la investigación.

## REFERENCIAS

- 1- C. ROMERO & S. VENTURA. *Data Mining E-Learning*. Editorial WIT Press, 2006.
- 2- C. ROMERO, S. VENTURA & E. GARCÍA. *Data Mining course management systems: Moodle case study and tutorial*. *Computers and Education*, 2007.
- 3- SCHEUER, O., & MCLAREN, B. M. (2012). *Educational data mining*. *Encyclopedia of the sciences of learning*, Springer, New York, 1075-1079.
- 4- TALAVERA, L., AND GAUDIOSO, E. 2004. *La minería de datos estudiantiles para caracterizar los grupos de comportamiento similares en espacios de colaboración*. Trabajo presentado en el Workshop de inteligencia artificial CSCL, 16<sup>th</sup> Conferencia Europea sobre Inteligencia Artificial, Valencia España.
- 5- LAROSE, D. T. (2005). *Discovering Knowledge in Data, an introduction to Data Mining*. John Wiley & Sons. Estados Unidos.

# GENERACIÓN DE UN PROCEDIMIENTO DE BUSQUEDA DE OUTLIERS SOBRE CAMPOS ALFANUMERICOS EN LOGS DE AUDITORIA

M. Rey<sup>1</sup>, H. Kuna<sup>1</sup>, M. D. Rolón<sup>1</sup>

1. Depto. de Informática, Facultad de Ciencias Exactas Quím. y Naturales Universidad Nacional de Misiones.

hdkuna@gmail.com , m.rey00@gmail.com

## RESUMEN

El término "outlier" se puede definir como un dato que difiere de forma significativa de otros presentes en un conjunto de datos. Dentro de la auditoría de sistemas existen herramientas informáticas que un auditor puede utilizar para realizar algunas de sus tareas, como es el análisis de datos. Se reconocen varios trabajos que utilizan técnicas de minería de datos para dar soporte a las tareas de un auditor de sistemas que se relacionan con el análisis de bases de datos, no abundando aquellas que trabajan sobre datos de tipo alfanumérico. En este contexto, se presenta la generación de un procedimiento de búsqueda de outliers sobre datos alfanuméricos en logs de auditoría de un sistema, con el objetivo de constituir una herramienta para un auditor de sistemas. El procedimiento generado se valida a través de la experimentación realizada con bases de datos artificiales y reales, obteniendo resultados satisfactorios.

**Palabras clave:** minería de datos, auditoría de sistemas, detección de outliers, datos alfanuméricos.

## CONTEXTO

Esta línea de investigación articula el "Programa de Investigación en Computación" de la Facultad de Ciencias Exactas Químicas y Naturales de la Universidad Nacional de Misiones; el "Proyecto 33A081: Sistemas de Información e Inteligencia de Negocio" del Departamento de Desarrollo Productivo y Tecnológico de la Universidad Nacional de

Lanús; y el "Programa de Doctorado en Ingeniería de Sistemas y Computación del Departamento de Lenguajes y Ciencias de la Computación" de la Universidad de Málaga-España.

## 1 INTRODUCCION

### 1.1 Minería de Datos para la Detección de Outliers en Bases de Datos

Un outlier se puede definir como un dato que por ser muy diferente a los demás pertenecientes a un mismo conjunto de datos, por ejemplo una base de datos (BD), puede considerarse que fue creado por un mecanismo diferente (Hawkins, 1980).

En la actualidad la minería de datos (MD) tiene un rol fundamental en la detección de outliers con una amplia diversidad de técnicas que buscan detectar outliers a través de diferentes clases de algoritmos, estableciendo diferentes definiciones de outliers en base a sus características específicas (Zhang, Meratnia, & Havinga, 2007), debe destacarse que, con el paso del tiempo, las técnicas han evolucionado en términos de efectividad y eficiencia (Hodge & Austin, 2004), llegando a niveles óptimos en sus respectivas clases.

### 1.2 Clasificación de Técnicas para Detección de Outliers

En la literatura del área de ciencias de la computación se puede encontrar diversos métodos para la detección de outliers, entre los que se pueden mencionar (Hodge & Austin, 2004), (Zhang et al., 2007):

- *Métodos basados en la distancia:* en este caso las técnicas identifican a los outliers en base a una medida de distancia, calculada utilizando todas las dimensiones disponibles, entre

un punto y su vecindario dentro del conjunto de datos (Knorr & Ng, 1998), (Knorr, Ng, & Tucakov, 2000).

- *Métodos basados en la densidad:* aquellos que toman en consideración la densidad de los datos al momento de calcular las distancias entre los puntos del conjunto de datos, para determinar la presencia de outliers “locales” (Breunig, Kriegel, Ng, & Sander, 2000).
- *Métodos basados en agrupamientos:* se trata de técnicas que a través de procesos de agrupamiento aíslan a los outliers en alguno de los clusters generados, variando según la técnica la caracteriza a tal cluster (Ester, Kriegel, Sander, & Xu, 1996).
- *Métodos basados en sub-espacios:* en este caso los outliers se detectan a partir de una observación de la distribución de densidad de clusters en un sub-espacio de pocas dimensiones, siendo detectados aquellos que tienen menor densidad a la media (C. Aggarwal & Yu, 2005).
- *Métodos basados en redes neuronales:* aquellos que a partir del uso de redes neuronales identifican a los outliers en tareas de clasificación o regresión (Sykacek, 1997).

### 1.3 Minería de Datos para la Auditoría de Sistemas

Dentro de la auditoría de sistemas existen las CAATs (Técnicas de Auditoría Asistidas por Computadora, por sus siglas en inglés) que son herramientas informáticas al servicio de un auditor. La MD provee de diversas técnicas que pueden ser usadas por un auditor para facilitar su trabajo, en particular al trabajar sobre logs de auditoría se encuentran trabajos para la detección de intrusos en sistemas (Lee, Stolfo, & Mok, 1998), identificar patrones de uso de sitios web (Mamčenko & Kulvietienė, 2005), entre otros. También se

encuentran aplicaciones de detección de outliers para auditoría de sistemas como ser: (Wu, Shi, Jiang, & Weng, 2007), (Yoon, Kwon, & Bae, 2007), entre otros.

## 2 LINEAS DE INVESTIGACION Y DESARROLLO

Existen aplicaciones de métodos de detección de outliers sobre logs de auditoría actuando como CAATs, pero constituyen implementaciones aisladas que no representan procedimientos formalmente definidos para tal efecto. En este sentido existen iniciativas tendientes al establecimiento de procesos formales de MD para la detección de datos anómalos en BD, éstos tienen por objetivo constituir una alternativa útil para la tarea de auditoría de sistemas a partir de la automatización de tareas de detección de outliers (Kuna et al., 2012).

Sin embargo no existen procedimientos formalmente definidos para la aplicación de técnicas de MD para la búsqueda de outliers sobre campos alfanuméricos presentes en logs de auditoría de sistemas para constituir una herramienta para un auditor de sistemas.

## 3 RESULTADOS OBTENIDOS/ESPERADOS

### 3.1 Procedimiento Desarrollado

Se determinó que una única técnica no sería suficiente para obtener la calidad de resultados que exige una actividad como es la auditoría de sistemas. Se optó por una solución que integrara varias técnicas, como se menciona en (Britos, 2008), (Shculz, 2008), por un lado tomando algunas de detección de outliers junto a otras técnicas de MD de propósito general. Esto con la finalidad de que la efectividad global a obtener por el procedimiento sea mayor, ya que una técnica determinada puede ser más efectiva que otra en ciertos aspectos (Schaffer, 1994).

En una primera instancia de análisis se seleccionaron técnicas teniendo en cuenta diversas características, para las de

detección de outliers: principalmente que pudieran operar sobre datos alfanuméricos, además, que no requieran una gran cantidad de parámetros y que la determinación de los mismos se pudiera realizar en forma automática. Mientras que para el caso de las técnicas de propósito general se priorizó que pudieran utilizarse en forma complementaria con las del primer grupo para dar lugar a una solución integrada.

Una vez seleccionadas las técnicas se inició con el proceso de análisis en sí a fin de determinar cuáles pasarían a formar parte del procedimiento. Para ello se comenzó por realizar pruebas con BD sintéticas, es decir, generadas artificialmente; sobre éstas se probaron las técnicas de detección de outliers seleccionadas previamente y aquellas que obtuvieron un mejor resultado global, en términos de efectividad en la detección y bajo porcentaje de errores en el proceso; fueron sobre las que continuó el análisis utilizando las técnicas del segundo grupo para buscar reforzar y/o corregir los resultados obtenidos por las técnicas del primer grupo.

Una vez que se finalizó el análisis se procedió con el diseño e implementación del procedimiento, quedando el mismo conformado por dos etapas: una primera en la que se aplican las técnicas de detección de outliers y una posterior en la que aplican las técnicas de propósito general para el refinamiento de los resultados.

Las operaciones que conforman el procedimiento son las siguientes:

- Lectura de la BD objetivo
- Para la aplicación de las técnicas de detección de outliers:
  - Aplicación de LOF (Breunig et al., 2000)
  - Adaptación de los resultados de LOF
  - Aplicación de DBSCAN (Ester et al., 1996)
  - Adaptación de los resultados de DBSCAN
- Unión de los resultados de LOF y DBSCAN

- Aplicación de un conjunto de reglas para la limpieza de los resultados (1)
- Para la aplicación de las técnicas de propósito general:
  - Aplicación del algoritmo TDIDT C4.5 (Quinlan, 1993)
  - Aplicación del modelo de C4.5 sobre la BD
  - Aplicación de la red Bayesiana (Pearl, 1988)
  - Aplicación del modelo de la red Bayesiana sobre la BD
  - Aplicación del algoritmo de extracción de reglas PART (E. Frank & Witten, 1998)
  - Aplicación del modelo de la técnica PART a la BD
- Unión de los resultados de la aplicación de los modelos generados a la BD
- Aplicación de un conjunto de reglas para la limpieza de los resultados de la unión (2)
- Escritura de los resultados generales

En las actividades (1) y (2) se hace mención a un conjunto de reglas que se aplicaron para la limpieza de los resultados, el motivo de tal actividad fue la necesidad, ante la unión de los resultados de dos o más técnicas, de potenciar la correcta detección de outliers y minimizar la cantidad de errores que se pudiera haber acumulado.

El conjunto de reglas al que se hace referencia se generó a medida que se realizó el análisis de las diversas técnicas para el diseño del procedimiento. En el caso de las de la actividad (1) se resolvió el caso en el que las dos técnicas no coincidían con la clasificación para una tupla, en tal caso se recurrió al valor de LOF asignado a la tupla, siendo que si el mismo era mayor a dos umbrales definidos, la tupla se consideraba como outlier, el motivo de tal determinación fue que tal técnica obtuvo mejores resultados en cuanto a la detección de los outliers, en cambio DBSCAN obtuvo mejores resultados en lo referido a los



falsos positivos. El caso de la actividad (2) se realizó algo similar, considerando en este caso que, si la tupla era definida como outlier por la mayoría de los algoritmos utilizados, se mantenía tal clasificación, en caso de que sólo una técnica haya sido la que marcó a la tupla como outlier se pasaba a utilizar el valor de LOF obtenido en la primera etapa por la misma, evaluándolo contra un umbral más alto en el caso de la actividad anterior.

### 3.2 Experimentación

Una vez finalizado el diseño e implementación del procedimiento desarrollado se procedió con la validación del mismo a partir de dos instancias de experimentación. En un primer caso se utilizó una BD obtenida a partir de un repositorio digital (A. Frank & Asuncion, 2010) de la Universidad de California, EEUU. Como la BD “Mushroom” no cuenta en forma nativa con outliers se decidió utilizar las tuplas de una de las clases presentes en la misma para establecer una selección del 5% de tales tuplas para que cumplieran la función de outliers en esta experimentación, estrategia similar a la seguida en otras publicaciones (C. C. Aggarwal & Yu, 2001), (Breunig et al., 2000). Al aplicar sobre la BD con los outliers incluidos se obtuvo un porcentaje de efectividad superior al 70% y un margen de errores menor al 1.5%.

En una segunda instancia de experimentación se utilizó una BD obtenida a partir los logs de auditoría de un sistema real, en este caso los resultados de la ejecución del procedimiento fueron evaluados por los administradores del sistema en cuestión, determinando ellos los casos en los que el procedimiento detectó correctamente outliers y aquellos casos en los que se había generado un error en la clasificación. Los resultados obtenidos fueron ampliamente satisfactorios con porcentajes de efectividad que superaban el 80% y un margen de errores en la clasificación menor al 1%.

Como conclusión se pudo detectar efectivamente outliers en campos alfanuméricos de logs de auditoría de un sistema para colaborar con la tarea de un auditor.

## 4 FORMACION DE RECURSOS HUMANOS

Este proyecto es parte de las líneas de investigación del “Programa de Investigación en Computación” de la Facultad de Ciencias Exactas Químicas y Naturales de la UNaM, con siete integrantes (todos ellos alumnos, docentes y egresados de la carrera de Licenciatura en Sistemas de Información de la facultad de Ciencias Exactas Químicas y Naturales de la Universidad Nacional de Misiones) de los cuales cuatro están realizando su tesis de grado y dos están realizando un doctorado.

## 5 BIBLIOGRAFIA

- Aggarwal, C. C., & Yu, P. S. (2001). Outlier detection for high dimensional data. En *Proceedings of the 2001 ACM SIGMOD international conference on Management of data* (pp. 37–46). New York, NY, USA: ACM.
- Aggarwal, C., & Yu, S. (2005). An effective and efficient algorithm for high-dimensional outlier detection. *The VLDB Journal*, 14(2), 211–221.
- Breunig, M. M., Kriegel, H.-P., Ng, R. T., & Sander, J. (2000). LOF: identifying density-based local outliers. En *Proceedings of the 2000 ACM SIGMOD international conference on Management of data* (pp. 93–104). New York, NY, USA: ACM.
- Britos, P. (2008). *Procesos de Explotación de Información Basados en Sistemas Inteligentes* (Tesis Doctoral). Universidad Nacional de La Plata, La Plata, Buenos Aires, Argentina.

- Ester, M., Kriegel, H., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *Proc. of KDD '96*, 226-231.
- Frank, A., & Asuncion, A. (2010). *UCI Machine Learning Repository*. University of California, Irvine, School of Information and Computer Sciences.
- Frank, E., & Witten, I. H. (1998). Generating Accurate Rule Sets Without Global Optimization. *In: Proc. of the 15th Int. Conference on Machine Learning*.
- Hawkins, D. M. (1980). *Identification of outliers*. Taylor & Francis.
- Hodge, V., & Austin, J. (2004). A Survey of Outlier Detection Methodologies. *Artificial Intelligence Review*, 22, 85-126.
- Knorr, E. M., & Ng, R. T. (1998). Algorithms for Mining Distance-Based Outliers in Large Datasets. *Proceedings of the 24th VLDDDB Conference*, 392-403.
- Knorr, E. M., Ng, R. T., & Tucakov, V. (2000). Distance-Based Outliers: Algorithms and Applications. *The VLDB Journal*, 8, 237-253.
- Kuna, H., Pautsch, J. G. A., Rey, M., Cuba, C., Rambo, A., Caballero, S., García Martínez, R., Villatoro, F. (2012). Comparación de la efectividad de procedimientos de la explotación de información para la identificación de outliers en bases de datos. Presentado en XIV Workshop de Investigadores en Ciencias de la Computación.
- Lee, W., Stolfo, S. J., & Mok, K. W. (1998). Mining Audit Data to Build Intrusion Detection Models. Presentado en AAAI-KDD-98, New York - USA: AAAI.
- Mamčenko, J., & Kulvietienė, R. (2005). From log files to valuable information using data mining techniques. *En Proceedings of the 4th WSEAS/IASME international conference on System science and simulation in engineering* (pp. 216–219). Stevens Point, Wisconsin, USA: World Scientific and Engineering Academy and Society (WSEAS).
- Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann.
- Quinlan, J. R. (1993). *C4.5: programs for machine learning*. Morgan Kaufmann.
- Schaffer, C. (1994). A conservation law for generalization performance. *En Proceedings of the Eleventh International Conference on Machine Learning* (pp. 259-265).
- Shculz, G. (2008). *Un Ambiente Integrado de Clasificación, Selección y Ponderación de Reglas Basado en Sistemas Inteligentes* (Tesis de grado). UBA - Fac. de Ingeniería, Bs. As., Argentina.
- Sykacek, P. (1997). Equivalent Error Bars For Neural Network Classifiers Trained By Bayesian Inference. *IN PROC. ESANN*, 121-126.
- Wu, N., Shi, L., Jiang, Q., & Weng, F. (2007). An Outlier Mining-Based Method for Anomaly Detection. *En 2007 IEEE International Workshop on Anti-counterfeiting, Security, Identification* (pp. 152 -156).
- Yoon, K.-A., Kwon, O.-S., & Bae, D.-H. (2007). An Approach to Outlier Detection of Software Measurement Data using the K-means Clustering Method. *En First International Symposium on Empirical Software Engineering and Measurement, 2007. ESEM 2007* (pp. 443 -445).
- Zhang, Y., Meratnia, N., & Havinga, P. (2007). A Taxonomy Framework for Unsupervised Outlier Detection Techniques for Multi-Type Data Sets. *Department of Computer Science – University of Twente – Netherlands*.

## PATRONES ESTADÍSTICOS RELACIONADOS CON EL PERFIL DEL ALUMNO DE LA FACULTAD DE TECNOLOGÍA Y CIENCIAS APLICADAS

María A. Barrera, M. Soledad Bustos Aguiar, Nancy Lamas, Daniela Lobos Anfuso, Manuel  
Baquinza y Leguizamón María Belén

Departamento Sistemas/Facultad de Tecnología y Ciencias Aplicadas/Universidad Nacional de Catamarca  
Maximio Victoria N° 55 - C.P: 4700 - San Fernando del Valle de Catamarca

Telefono: 03834- 435112 – int 168

[mbarrera@tecno.unca.edu.ar](mailto:mbarrera@tecno.unca.edu.ar), [soledadbustosaguiar@yahoo.com.ar](mailto:soledadbustosaguiar@yahoo.com.ar), [nancylamas@gmail.com](mailto:nancylamas@gmail.com),  
[danielalobosanfuso@gmail.com](mailto:danielalobosanfuso@gmail.com), [mbaquinzay@gmail.com](mailto:mbaquinzay@gmail.com), & [mbelenleguizamon@gmail.com](mailto:mbelenleguizamon@gmail.com)

### Resumen

El rendimiento académico estudiantil ha sido definido como el cumplimiento de las metas, logros u objetivos establecidos en el programa o asignatura que está cursando el alumno.

Se puede analizar el rendimiento académico como una medida de las capacidades del alumno, que expresa lo que éste ha aprendido a lo largo del proceso formativo.

Existen diversos motivos o factores que pueden llevar al alumno a mostrar un pobre rendimiento académico tales como (poca motivación, desinterés, distracciones en clase, etc.)

En este trabajo se propone aplicar un proceso de descubrimiento de conocimiento o KDD (Knowledge Discovery from Data Base) a la Base de Datos de los alumnos de la Facultad de Tecnología y Ciencias Aplicadas de la Universidad Nacional de Catamarca (UNCa), con el propósito de convertir una masa de datos carente de significado en información relevante, es decir, en conocimiento. Dicho conocimiento vendrá dado mediante patrones, tendencias o relaciones.

Estos resultados pueden mejorar el proceso de formación académica, estimular fundamentalmente el desarrollo de aptitudes y de valores, contribuir en la toma de decisiones tácticas y estratégicas de la Facultad, proporcionando un sentido automatizado para la generación de conocimiento y elevando así la calidad de la educación en la Universidad.

**Palabras clave:** KDD (Knowledge Discovery

from Data Base, Data Minig, Patrones, Rendimiento Académico.

### Contexto

El proyecto de investigación: “Patrones estadísticos relacionados con el perfil del alumno de la Facultad de Tecnología y Ciencias Aplicadas”, se desarrolla y ejecuta en ámbitos del Departamento de Informática de la Facultad de Tecnología y Ciencias Aplicadas de la UNCa y es evaluado por la Secretaría de Ciencia y Tecnología de la universidad, donde los integrantes se desempeñan como docentes de la carrera de Ingeniería en Informática.

Así mismo es prioritario, tanto para el Departamento de Informática, como para los demás departamentos de la Facultad, transferir soluciones concretas a las problemáticas que se plantean a los docentes en relación con el rendimiento académico de los alumnos.

Utilizando los conceptos del proceso de KDD y sus técnicas de Data Minig, se decidió tomar como marco de referencia a la Facultad de Tecnología y Ciencias Aplicadas. La aplicación de estas técnicas constituye una herramienta o medio para lograr el objetivo del presente proyecto. Si bien en la Facultad se han puesto en práctica técnicas de DM, no existen antecedentes de aplicación de un proceso de KDD, involucrando a todo el alumnado.

### Introducción

El rendimiento académico estudiantil desde un punto de vista operativo, se ha limitado a la expresión de una nota cuantitativa o cualitativa y se encuentra que en muchos casos es

insatisfactorio lo que se ve reflejado en la deserción (abandono de los estudios universitarios) y el desgranamiento (pérdida de continuidad y/o retraso en el avance) de los alumnos de la carrera

El rendimiento supone la capacidad del alumno para responder a los estímulos educativos. En este sentido, el rendimiento académico está vinculado a la aptitud.

Existen distintos factores que inciden en el rendimiento académico. Desde la dificultad propia de algunas asignaturas, hasta la gran cantidad de exámenes que pueden coincidir en una fecha, pasando por la amplia extensión de ciertos programas educativos. Son muchos los motivos que pueden llevar a un alumno a mostrar un pobre rendimiento académico.

Por otra parte, el rendimiento académico puede estar asociado a la subjetividad del docente cuando corrige. Ciertas materias, en especial aquellas que pertenecen a las ciencias sociales, pueden generar distintas interpretaciones o explicaciones, que el profesor debe saber analizar en la corrección para determinar si el estudiante ha comprendido o no los conceptos.

Como podemos ver, hay muchos motivos por los que el rendimiento del alumno se ve afectado. Es por eso que no podemos simplificar el análisis del mismo a sólo las calificaciones obtenidas en los exámenes a lo largo del cursado de una materia a través de métodos meramente estadísticos.

En la última década, ha existido un gran crecimiento en nuestras capacidades de generar y coleccionar datos en las más diversas áreas: el comercio, la banca, astronomía, física de partículas, química, medicina, departamentos de gobierno, entre otras. Esto debido al gran poder de procesamiento de las máquinas como a su bajo costo de almacenamiento. Sin embargo, dentro de estas enormes masas de datos existe una gran cantidad de información "oculta", de gran importancia estratégica, a la que no se puede acceder por las técnicas clásicas de recuperación de la misma.

A partir de la aplicación de un grupo de técnicas de Data Mining (DM) como el clustering, árboles de decisión, algoritmos de aprendizaje inductivo, sumarización, modelización de dependencias, regresión,

entre otros, se pretende clasificar a los estudiantes encontrando patrones ocultos y reglas que los caractericen. Estos resultados pueden mejorar el proceso de formación académica, estimular fundamentalmente el desarrollo de aptitudes y de valores, contribuir en la toma de decisiones tácticas y estratégicas de la Facultad, proporcionando un sentido automatizado para la generación de conocimiento y elevando de esta manera la calidad de la educación en la Universidad.

### **Líneas de investigación y desarrollo**

El proyecto aborda la aplicación de un proceso de descubrimiento de conocimiento o KDD a la Base de Datos de los alumnos de la Facultad de Tecnología y Ciencias Aplicadas de la UNCa.

El proceso de KDD entre otras sofisticadas técnicas aplica la inteligencia artificial para encontrar patrones y relaciones dentro de los datos permitiendo la creación de modelos, es decir, representaciones abstractas de la realidad.

Este proceso conlleva diversas actividades entre ellas interpretación de los datos sobre los que se va a trabajar, depuración de los mismos, aplicación de algoritmos que ofrece la técnica de DM y análisis de resultados. Hay que destacar que todo este proceso es interactivo e iterativo.

### **Resultados y Objetivos**

El grupo de trabajo ha realizado algunos análisis planteados, pero no posee aún resultados finales, ya que el mismo ha comenzado su ejecución en enero del pasado año. Se tiene como objetivo proveer los conocimientos para la formación científica, tecnológica y complementaria, sobre las tecnologías de Gestión de Datos en Base de Datos (BD), buscando manipular estructuras de almacenamiento de grandes volúmenes de información para lograr acceso rápido y seguro a los datos. Por lo que se pretende investigar y experimentar nuevas técnicas y metodologías para lograr la agilidad en las distintas operaciones relacionadas a las BD. Esto permitirá obtener información, que con simples accesos o consultas a los datos no sería posible y se podrá, de acuerdo a cada



situación, elegir el algoritmo correcto para obtener resultados.

### Objetivo General:

Determinar patrones estadísticos relacionados con el perfil del alumno de la Facultad de Tecnología y Ciencias Aplicadas de la UNCa.

### Objetivos Específicos:

- Determinar el rendimiento académico de los alumnos por año y carrera de la Facultad de Tecnología y Ciencias Aplicadas.
- Descubrir la incidencia en el rendimiento académico, teniendo en cuenta los entornos socio-económico, socio-familiar, socio-sanitario y habitacional de los alumnos.

## Formación de Recursos Humanos

El proyecto cuenta con integrantes que se encuentran en la etapa de desarrollo de sus tesis de posgrado en el marco de la carrera de maestría en Ingeniería del Software en temas específicamente relacionados al área del proyecto y con una becaria de Beca de Estimulo a las Vocaciones Científicas (CIN) 2012 otorgada por la Secretaria de Ciencia y Tecnología de UNCa. Además los mismos son docentes de las cátedras Gestión de Datos y Estructura de Datos de la carrera Ingeniería en Informática de la Facultad de Tecnología y Ciencias Aplicadas; estando ambas materias estrechamente ligadas al tema en cuestión y donde tienen por objetivo proveer los conocimientos para la formación científica, tecnológica y complementaria sobre las tecnologías de Gestión de Datos en Base de Datos, buscando manipular estructuras de almacenamiento de grandes volúmenes de información para lograr acceso rápido y seguro a los datos, por lo consiguiente siempre están abocados a investigar y experimentar nuevas técnicas y metodologías para lograr la agilidad en las distintas operaciones relacionadas a las BD.

Específicamente el proyecto de investigación prevé el programa de capacitación y formación de recursos humanos, que contempla las siguientes actividades:

- Incorporación de alumnos de los últimos años de la carrera de Ingeniería en Informática de la Facultad de Tecnología

y Ciencias Aplicadas de la UNCa., en calidad de auxiliares de investigación.

- Dirección de tesinas de grado de la carrera de Ingeniería en Informática de la Facultad de Tecnología y Ciencias Aplicadas de la UNCa integrados al proyecto.
- Participación de los integrantes del proyecto en cursos de actualización y posgrado en el área de estudio.
- Participación en talleres o workshops de herramientas informáticas relacionadas con las BD y KDD.
- Celebración de convenios con otras universidades, para capacitación de los recursos humanos en KDD.

Para garantizar la capacitación y actualización del equipo de investigación, así como la difusión de los avances y resultados logrados, se propuso la participación en eventos nacionales e internacionales de la especialidad, como congresos, simposios, seminarios y cursos.

## Bibliografía

- CHAPMAN & HALL, *The Top Ten Algorithms in Data Mining*
- DOÑA J.M., QUINTANA O.P., VALESANI M.E., VALLEJOS O.A. *Analysis of Agregation Methods in Incomplete Database Systems. Information Processing and Management of Uncertainty in Knowledge-Based System (IPMU 2008)*. ISBN: 9978-84-612-3061-7
- YANG C. YUAN. *Multiple Imputation for Missing Data: Concepts and New Development.*, SAS Institute Inc., Rockville, MD. P267-25.
- PIATETSKY-SHAPIRO G. FRAWLEY WJ. (1991) - *Knowledge Discovery in Databases Ed. AAAI/MIT Press.*
- THURAISINGHAM, B. (1999) – *Data Mining. Technologies, Techniques Tools and Trends – Ed. CRC Press LLC.*
- WITTEN, IAN H. AND FRANK EIBE (2005) – *Data Mining: Practical Machine Learning Tools and Techniques – Morgan Kaufmann 2° Edición.*
- JEAN - MICHEL – FRANCO (1997) – *El Data Warehouse. El Data Mining – Ed. Gestión 2000*

## INSTRUMENTOS PARA LA GESTIÓN DE PROYECTOS DE EXPLOTACIÓN DE INFORMACIÓN

García-Martínez, R., Merlino, H., Diez, E., Rodríguez, D., Pytel, P.,  
Baldizzoni, E., Arboleya, H., Martins, S., Cartanilica, A.

Grupo Investigación en Sistemas de Información

Departamento de Desarrollo Productivo y Tecnológico. Universidad Nacional de Lanús  
29 de Septiembre 3901 (1826) Remedios de Escalada, Lanús. Argentina. Tel +54 11 6322-9200 Ext. 194  
rgarcia@unla.edu.ar

### RESUMEN

Los Proyectos de Explotación de Información difieren sustancialmente de los de Software convencional. Las herramientas involucradas en los procesos de Ingeniería en Software no son aplicables a este tipo de proyectos. Surge la necesidad de desarrollar un cuerpo de conocimientos para una Ingeniería de Explotación de Información, cuyo eje se centre en el desarrollo de instrumentos que permitan un adecuado control y gestión de los proyectos, contribuyendo a mejorar la productividad de las PYMES del área.

**Palabras clave:** Proyectos de explotación de información, Instrumentos de gestión, test de viabilidad, modelo de ciclo de vida, mapa de actividades, formalismos para documentación de requisitos, método de estimación de recursos.

### CONTEXTO

Durante la etapa de investigación documental de los proyectos UNLa 33A81 y UNLa 33A105 desarrollados por el grupo de trabajo, se observó que en la bibliografía consultada se repetía con insistencia el uso indiscriminado de los términos “minería de datos” (o data mining) y “explotación de información” (o information mining) para referirse al mismo cuerpo de conocimientos. En el marco de este proyecto, se considera que la minería de datos se refiere a la algoritmia para encontrar patrones de conocimiento en masas de datos; mientras que la ingeniería de explotación de información entiende en los procesos y las metodologías utilizadas para ordenar, controlar y gestionar la tarea de encontrar patrones de conocimiento referidos (García-Martínez et al. 2011a).

Los proyectos de explotación de información poseen características muy distintas a los proyectos de desarrollo de software tradicional (Pollo-Cattaneo et al., 2010a), sobre todo en la parte operativa del proyecto. La diferencia se presenta en los procesos de desarrollo y mantenimiento en los cuales el ciclo

de fases de un proyecto de software tradicional: inicio, requisitos, análisis y diseño, construcción, integración y pruebas no resultan naturales en un proyecto de explotación de información (Vanrell et al., 2010a; 2010b; 2012). Por otra parte, al evaluar las principales metodologías existentes para los proyectos de explotación de información (Chapman et al., 1999; SAS, 2008; Pyle, 2003), se observa la falta de herramientas que permitan soportar de forma completa la fase de administración de proyectos (Vanrell 2012).

Durante el desarrollo el proyecto 33A105 “Ingeniería de Proyectos de Explotación de Información para PYMES” se pudo producir un primer ordenamiento del cuerpo de conocimiento existente sobre lo que se ha definido como línea de investigación en Ingeniería de Explotación de Información. Una de las áreas de vacancia identificadas es la de instrumentos que permitan un adecuado control y gestión de proyectos de explotación de información (García-Martínez et al., 2011c).

Una de las lecciones aprendidas sobre desarrollo de software en Informática derivada de los estadios tempranos de la disciplina, es que, la ausencia de una ingeniería de software conllevaba a un desarrollo artesanal de los artefactos software (Ochoa et al., 2008). El desarrollo artesanal implicaba la imposibilidad de poder establecer dentro de valores racionales, parámetros tales como: [a] cantidad y calificación de los recursos humanos a emplear en el proyecto, [b] tiempos de desarrollo, [c] modelos de proceso que guiaran el desarrollo y permitieran establecer hitos de entrega, [d] formalismos de documentación que dieran cuenta de lo hecho en el proyecto de desarrollo del artefacto software y de las decisiones de diseño asumidas, constituyendo el punto de partida para futuras ampliación de funcionalidades, [e] modelos de costo de proyecto (Böehm, 1981), entre otros. De hecho la estimación de estos parámetros se hacía en base a la experiencia de individuos sin ninguna base ingenieril y lo para un grupo de desarrollo podía hacerse en meses para otro podía hacerse en años.

La explotación de información esta en sus primeros estadios y, al igual que lo que ocurría con el desarrollo de artefactos software, adolece de una ingeniería que provea instrumentos para un adecuado control y gestión de proyectos de este tipo. En particular, en el caso de las PYMES del sector, reviste importancia por cuanto contar con herramientas para el control y la gestión de proyectos de explotación de información les permite explorar una nueva línea de servicios informáticos. Disponer de modelos de proceso para este tipo de proyectos, habilita a la PYME disponer de una guía que ordene el trabajo interno y permita establecer y ofrecer al requirente hitos para los entregables del proyecto. Disponer de modelos que permitan cuantificar: la viabilidad, los recursos y los tiempos a emplear en el proyecto de explotación de información, hace predecibles (y acotables) los costos al momento que una PyME ofrezca este tipo de servicios en el marco de aplicaciones gubernamentales o comerciales. Por otra parte disponer de formalismos de documentación facilita a la PYME repetir procesos con el consecuente ahorro de recursos.

## INTRODUCCIÓN

Hace ya un lustro, un estudio de la Universidad de California en Berkeley [Lyman y Varian, 2003] señaló que la información disponible en Internet crecía a razón de 92 petabytes [1015 bytes] por año. En [Maimon y Rokach, 2005] se ha señalado que esta información esta disponible para procesos de descubrimiento de conocimiento con independencia que se encuentre en fuentes estructuradas [Rudin y Cressy, 2003; Moss, 2003] ó desestructuradas [Vuori, 2006].

La inteligencia de negocio propone un abordaje interdisciplinario [dentro del que se encuentra la Informática], que tomando todos los recursos de información disponibles y el uso de herramientas analíticas y de síntesis con capacidad de transformar la información en conocimiento, se centra en generar a partir de éstos, conocimiento que contribuya con la toma de decisiones de gestión y generación de planes estratégicos en las organizaciones [Thomsen, 2003; Negash y Gray, 2008].

La Explotación de Información es la sub-disciplina de los Sistemas de Información que aporta a la Inteligencia de Negocio [Langseth y Vivatrat, 2003] las herramientas para la transformación de información en conocimiento [Mobasher et al., 1999; Srivastava et al., 2000; Abraham, 2003; Coley, 2003]. Ha sido definido como la búsqueda de

patrones interesantes y de regularidades importantes en grandes masas de información [Fayad et al., 1996; Grossman et al., 1998].

Un Proceso de Explotación de Información se define, como un grupo de tareas relacionadas lógicamente [Curtis et al., 1992] que, a partir de un conjunto de información con un cierto grado de valor para la organización, se ejecuta para lograr otro, con un grado de valor mayor que el inicial [Ferreira et al., 2005; Kanungo, 2005]. Adicionalmente, existe una variedad de técnicas de minería de datos, en su mayoría provenientes del campo del Aprendizaje Automático [García-Martínez, 1997; García-Martínez et al., 2003], susceptibles de ser utilizadas en cada uno de estos procesos.

El proyecto que se presenta es continuación del Proyecto UNLa 33A105 “Ingeniería de Proyectos de Explotación de Información para PyMEs” [05/2011 – 04/2013], que fue a su vez continuación del Proyecto UNLa 33A081 “Sistemas de Información para Inteligencia de Negocio” [05/2009 – 04/2011]. Los resultados de esto proyectos son:

- En el área de Fundamentos se ha argumentado la necesidad de desarrollar una Ingeniería de Proyectos de Explotación de Información para PYMES [García-Martínez et al., 2011c]; identificando las áreas prioritarias sobre las cuales trabajar en PyMEs [García-Martínez et al., 2010a].
- En el área de Procesos para Proyectos de Explotación de Información [EI], se ha formulado una propuesta de procesos de EI [Britos y García-Martínez, 2009; Rancan et al., 2010; García-Martínez et al., 2011b; Pollo Cattaneo et al., 2012a], se ha fundamentado la necesidad de definir técnicas para el aseguramiento de la calidad [Diez, et al., 2012], se ha realizado una caracterización empírica de dominios para uso en proyectos de EI [Lopez-Nocera et al., 2011]; se ha trabajado en desarrollar técnicas y una metodología específica para elicitación de requerimientos [Pollo Cattaneo et al., 2010b; Mansilla et al., 2012] identificando posibles formalismos de documentación [Vegega et al., 2012], se han sentado las bases para trabajar en viabilidad y estimación de proyectos de EI [Rodríguez et al., 2010; Pytel et al., 2011a; 2011b; 2011c; 2012], y se ha desarrollado un modelo de proceso de operación para proyectos de EI [Vanrell et al., 2010a; 2010b].
- En el área de Trabajo de Campo se validaron los resultados parciales obtenidos en problemas concretos en los dominios de: tuberías de gas [D’Atri et al., 2009], prevención de daños y averías en la industria automotriz [Flores et al.,



2009], caracterización de problemas de aprendizaje [Jiménez Rey et al., 2009], identificación de patrones de comportamiento de comunidades educativas mediadas por entornos virtuales [Cigliuti et al., 2012], identificación de errores de apropiación de conceptos en el dominio de análisis de sistemas de información [Saavedra-Martínez et al., 2012b; 2012b], prevención de estrés de suelos [Sansón et al., 2009], identificación de datos faltantes con ruido o inconsistentes en auditoría de sistemas [Kuna et al., 2010a; 2010b; 2011] realizando estudios comparativos de distintos procedimientos [Kuna et al., 2012], identificación de causales de abandono de estudios universitarios [Kuna et al., 2009; 2010c], y estimación en gestión de proyectos de software [Bogado, et al., 2011]. Previamente se trabajó en: identificación de caras humanas [Britos et al., 2005], detección de cambios de consumo de usuarios [Grosser et al., 2005; Britos et al., 2008], localización de patrones en eventos meteorológicos [Cogliati et al., 2006a; 2006b], predicción de la salud de una comunidad [Felgaer et al., 2006], detección de daños al corazón [Ferrero et al., 2006], registro de uso de sitios web [Britos et al., 2007], selección de protocolos pedagógicos [Britos et al., 2008a], comprobación de malentendidos en programación [Britos et al., 2008c], y detección de patrones criminales [Valenga et al., 2007a; 2007b; 2008].

### OBJETIVOS E HIPOTESIS DE INVESTIGACION

En este proyecto se busca continuar con el desarrollo y sistematización del cuerpo de conocimiento asociado a la Ingeniería de Explotación de Información iniciado en el proyecto 33A105. En esta etapa, la investigación se focalizará en el desarrollo de herramientas para el Control y la Gestión de Proyectos de Explotación de Información en PyMEs.

La pregunta problema que anima la investigación es: ¿Se puede cubrir la vacancia de herramientas de control y gestión para proyectos de explotación de información?

Entre los supuestos (o hipótesis) que guían el proyecto se encuentran:

*Hipótesis I:* Los proyectos de explotación de información poseen características muy distintas a las de los proyectos de desarrollo de software tradicionales. Las clásicas etapas de análisis, diseño, desarrollo, integración y testeo no encajan con las etapas naturales de los procesos de desarrollo de este tipo de proyectos. En consecuencia, los

métodos de determinación de viabilidad de proyecto, los modelos ciclos de vida para software (p.ej.: cascada, prototipado, ó espiral); los modelos de procesos software (p.ej.: IEEE 1074 o MOPROSOFT); y la natural derivación de estos últimos: los mapas de actividades para proyectos software no son aplicables a este tipo de proyectos.

*Hipótesis II:* Existen metodologías de explotación de información que destacan la importancia del planeamiento de una elicitación de requerimientos a lo largo de todo el proyecto de una manera ordenada, documentada, consistente y trazable. Sin embargo, el abordaje clásico de la ingeniería de requerimientos no se ajusta a los proyectos de explotación de información porque no atiende los aspectos particulares de especificación de requerimientos para este tipo de proyectos. De hecho, las técnicas clásicas no son aplicables al proceso de identificar el problema de explotación de información ni la documentación asociada.

*Hipótesis III:* El proceso de aseguramiento de la calidad es el conjunto de acciones planificadas y sistemáticas implantadas dentro del sistema de calidad, y demostrables si es necesario, para proporcionar la confianza adecuada que una entidad cumplirá los requisitos de calidad de procesos y de productos en el marco de un proyecto de desarrollo. La Ingeniería de Explotación de Información dispone de metodologías que guían el desenvolvimiento de proyectos que se consideran probadas y tienen un buen nivel de madurez en cuanto a los procesos de desarrollo. Sin embargo, estas metodologías no definen procesos específicos para el control y la gestión de proyectos en el área, en particular, para el aseguramiento de la calidad.

*Hipótesis IV:* La determinación de las tareas a desarrollar en un proyecto, de los recursos necesarios y del tiempo que transcurrirá desde el comienzo hasta el final de su realización; es una actividad de control y gestión que permite definir tempranamente la duración del proyecto y formular el correspondiente cálculo de costos. Si bien el método DMCoMo permite estimar proyectos de explotación de información, sus autores han señalado su aplicabilidad solo para proyectos de grandes organizaciones. Esto deja a las PyMEs sin un procedimiento de estimación de proyectos de explotación de información,

*Objetivo General:* El objetivo de este proyecto es desarrollar el cuerpo de conocimiento necesario para las actividades tempranas de control y gestión en un proyecto de explotación de información con focalización en su transferencia al sector PyMEs de informática. Se busca proveer las siguientes



herramientas para proyectos de explotación de información: test de viabilidad, modelo de ciclo de vida, mapa de actividades, procedimiento de derivación del problema de negocio en problema de explotación de información, formalismos para documentación de requisitos, técnicas y métodos de aseguramiento de la calidad del proceso y del producto y método de estimación de recursos.

*Objetivos específicos vinculados a Hipótesis I:*

- 1.- Desarrollar de un Test de Viabilidad de Proyectos de Explotación de Información
- 2.- Desarrollar un Modelo de Ciclo de Vida para Proyectos de Explotación de Información
- 3.- Desarrollar un Mapa de Actividades para para Proyectos de Explotación de Información

*Objetivos específicos vinculados a la Hipótesis II:*

- 4.- Desarrollar un procedimiento basado en formalismos de Ingeniería del Conocimiento que permita derivar el problema de explotación de información a partir de representaciones del problema de negocio.
- 5.- Desarrollar formalismos basados en Ingeniería del Conocimiento que permitan la documentación de requisitos en proyectos de explotación de información

*Objetivo específico vinculado a la Hipótesis III:*

- 6.- Desarrollar técnicas y métodos para aseguramiento de la calidad del proceso y del producto en Proyectos de Explotación de Información.

*Objetivo específico vinculado a la Hipótesis IV:*

- 7.- Desarrollo de un Método de Estimación de Recursos para Proyectos de Explotación de Información en PyMES.

## METODOLOGÍA DE TRABAJO

Se prevé: realizar investigación documental identificando casos de estudio, desarrollar mediante la metodología de prototipado evolutivo las versiones iniciales de los instrumentos combinado con mejora incremental, probar la versión de producción de los prototipos en casos seleccionados.

## RESULTADOS OBTENIDOS/ESPERADOS

Como resultado de este proyecto, se esperan contar al fin de su desarrollo con los siguientes instrumentos para la gestión de proyectos de explotación de información: test de viabilidad, modelo de ciclo de vida, mapa de actividades, procedimiento de derivación del problema de negocio en problema de explotación de información, formalismos para documentación de requisitos,

técnicas y métodos de aseguramiento de la calidad del proceso y del producto, y método de estimación de recursos.

## FORMACIÓN DE RECURSOS HUMANOS

El grupo de trabajo se encuentra formado por tres investigadores formados, dos investigadores en formación y cuatro becarios alumnos de la carrera Licenciatura en Sistemas de la UNLa. En su marco se desarrolla dos Tesis de Doctorado en Ciencias Informáticas y cuatro Trabajos Finales de Licenciatura en Sistemas.

## REFERENCIAS

- Abraham, A. 2003. Business Intelligence from Web Usage Mining. Journal of Information & Knowledge Management, 24: 375-390.
- Böehm, B. 1981. Software Engineering Economics. Prentice Hall. ISBN 0-13-822122-7.
- Böehm, B. 1988. A Spiral Model of Software Development and Enhancement. IEEE Computer. Mayo 1988. Pág. 61-72.
- Bogado, V., Dapozo, G., García Martínez, R. 2011. Estimación en Gestión de Proyectos de Software Basada en Explotación de Información. Proceedings XIII Workshop de Investigadores en Ciencias de la Computación. Artículo 3894.
- Britos, P., Abasolo, M., García-Martínez, R. y Perales, F. 2005. Identification of MPEG-4 Patterns in Human Faces Using Data Mining Techniques. Proceedings 13th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision'2005. Páginas 9-10.
- Britos, P., Cataldi, Z., Sierra, E., García-Martínez, R. 2008a. Pedagogical Protocols Selection Automatic Assistance. Notes in Artificial Intelligence 5027: 331-336.
- Britos, P., García-Martínez, R. 2009. Propuesta de Procesos de Explotación de Información. Proceedings XV Congreso Argentino de Ciencias de la Computación. Workshop de Base de Datos y Minería de Datos. Págs. 1041-1050. ISBN 978-897-24068-4-1.
- Britos, P., Grosser, H., Rodríguez, D., García-Martínez, R. 2008b. Detecting Unusual Changes of Users Consumption. In Artificial Intelligence in Theory and Practice II, ed. M. Bramer, Boston: Springer, 276: 297-306.
- Britos, P., Jiménez Rey, E., García-Martínez, R. 2008c. Work in Progress: Programming Misunderstandings Discovering Process Based On Intelligent Data Mining Tools. Proceedings 38th ASE/IEEE Frontiers in Education Conference.
- Britos, P., Martinelli, D., Merlino, H., García-Martínez, R. 2007. Web Usage Mining Using Self Organized Maps. International Journal of Computer Science and Network Security, 76:45-50. ISSN : 1738-7906.
- Chapman, P., Clinton, J., Keber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R. 1999. CRISP-DM 1.0 Step by step Bgguide. Edited by SPSS. <http://www.crispdm.org/CRISPWP-0800.pdf>. Ultimo acceso Junio 2008.
- Cigliuti, P., Pollo-Cattaneo, F., García-Martínez, R. 2012. Procesos de Identificación de Comportamiento de Comunidades Educativas Centradas en EVEAs. Proceedings del XIV Workshop de Investigadores en Ciencias de la Computación. Pág. 954-956. ISBN 978-950-766-082-5.
- Cogliati, M., Britos, P., García-Martínez, R. 2006a. Patterns in Temporal Series of Meteorological Variables Using SOM & TDIDT. In IFIP International Federation for Information Processing, Volume 217, Artificial Intelligence in Theory and Practice, ed. M. Bramer, Boston: Springer, Pág. 305-314.
- Cogliati, M., Britos, P., García-Martínez, R. 2006b. Análisis del Viento en el Valle del Río Negro Mediante Mapas Auto Organizados y Algoritmos de Inducción. Avances en Energías Renovables y Ambiente Vol. 10. Pág. 11.137-11.144.
- Cooley, R. 2003. The Use of Web Structure and Content to Identify Subjectively Interesting Web Usage Patterns. ACM Transactions on Internet Technology, 32: 93-116.
- Curtis, B., Kellner, M., Over, J. 1992. Process Modelling. Communications of the ACM, 359: 75-90.
- D'Attri, M., Rodríguez, D., García-Martínez, R. 2009. Improving Pipeline Risk Models by Using Data Mining Techniques. 24th World Gas Conference Proceedings CD. Paper 663.
- Diez, E., Pytel, P., Rodríguez, D., García, R., Lacabanne, M., Leonardis, L., Martins, S., Cartanilica, A., García-Martínez, R. 2012. Aseguramiento de la Calidad para Proyectos de Explotación de Información. Proceedings del XIV Workshop de Investigadores en Ciencias de la Computación. Pág. 212-216. ISBN 978-950-766-082-5.
- Fayad, U. M., Piatetsky-Shapiro, G., Smyth, P., Uthurudamy, R. 1996. Advances in Knowledge Discovery and Data Mining. AAAI Press.
- Felgaer, P., Britos, P. and García-Martínez, R. 2006. Prediction in Health Domain Using Bayesian Network Optimization Based on Induction Learning Techniques. International Journal of Modern Physics C 173: 447-455.
- Ferreira, J., Takai, O., Pu, C. 2005. Integration of Business Processes with Autonomous Information Systems: A Case Study in Government Services. Proceedings Seventh IEEE International Conference on E-Commerce Technology. Pág. 471-474.
- Ferrero, G., Britos, P., García-Martínez, R., 2006. Detection of Breast Lesions in Medical Digital Imaging Using Neural Networks. In IFIP International Federation for Information Processing, Volume 218, Professional Practice in Artificial Intelligence, eds. J. Debenham, Boston: Springer, Pág. 1-10.

- Flores, D., García-Martínez, R., Fernandez, E., Merlino, H., Rodríguez, D., Britos, P. 2009. Detección de Patrones para la Prevención de Daños y/o Averías en la Industria Automotriz. *Proceedings XV Congreso Argentino de Ciencias de la Computación. Workshop de Base de Datos y Minería de Datos*. Págs. 1021-1030. ISBN 978-897-24068-4-1.
- García Martínez, R. 1997. *Sistemas Autónomos. Aprendizaje Automático*. Editorial Nueva Librería. ISBN 950-9088-84-6. García Martínez, R. y Britos, P. 2004. *Ingeniería de Sistemas Expertos*. Editorial Nueva Librería. ISBN 987-1104-15-4.
- García Martínez, R., Servente, M. y Pasquini, D. 2003. *Sistemas Inteligentes*. Editorial Nueva Librería. Buenos Aires. ISBN 987-1104-05-7.
- García-Martínez, R., Britos, P., Pesado, P., Bertone, R., Pollo-Cattaneo, F., Rodríguez, D., Pytel, P., Vanrell, J. 2011a. Towards an Information Mining Engineering. En *Software Engineering, Methods, Modeling and Teaching*. Sello Editorial Universidad de Medellín. ISBN 978-958-8692-32-6. Páginas 83-99.
- García-Martínez, R., Britos, P., Pollo-Cattaneo, F., Rodríguez, D., Pytel, P. 2011b. Information Mining Processes Based on Intelligent Systems. *Proceedings of II International Congress on Computer Science and Informatics INFONOR-CHILE 2011*. Pp. 87-94. ISBN 978-956-7701-03-2.
- García-Martínez, R., Lelli, R., Merlino, H., Cornachia, L., Rodríguez, D., Pytel, P., Arbolea, H. 2011c. Ingeniería de Proyectos de Explotación de Información para PYMES. *Proceedings XIII Workshop de Investigadores en Ciencias de la Computación*. Artículo 3761.
- Grossam, 1999. Grossman, R., Kasif, S., Moore, R., Rocke, D., Ullman, J. 1998. *Data Mining Research: Opportunities and Challenges*. <http://www.csl.mtu.edu/cs5811/common/2002-fall-reading-group/2002-10-25/grossman98.ps>. Ultimo acceso 17 de Abril del 2008.
- Grosser, H., Britos, P. y García-Martínez, R. 2005. Detecting Fraud in Mobile Telephony Using Neural Networks. *Lecture Notes in Artificial Intelligence* 3533: 613-615.
- Jiménez Rey, E., Rodríguez, D., Britos, P., García-Martínez, R. 2009. Caracterización de Problemas de Aprendizaje Basada en Explotación de Información. *Proceedings XI Workshop de Investigadores en Ciencias de la Computación*. Pág. 627-629. ISBN 978-950-605-570-7.
- Kanungo, S. 2005. Using Process Theory to Analyze Direct and Indirect Value-Drivers of Information Systems. *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*. Pág. 231-240.
- Kuna, H., Pautsch, G., Rey, M., Cuba, C., Rambo, A., Caballero, S., García-Martínez, R., Villatoro, F. 2012. Comparación de la Efectividad de Procedimientos de la Explotación de Información para la Identificación de Outliers en Bases de Datos. *Proceedings del XIV Workshop de Investigadores en Ciencias de la Computación*. Pág. 296-300. ISBN 978-950-766-082-5.
- Kuna, H., Caballero, S., Rambo, A., Meinel, E., Steinhilber, A., Pautsch, G., García-Martínez, R., Villatoro, F. 2010a. Avances en Procedimientos de la Explotación de Información Para la Identificación de Datos Faltantes, con Ruido e Inconsistentes. *Proceedings XII Workshop de Investigadores en Ciencias de la Computación*. Pág. 137-141.
- Kuna, H., Caballero, S., Rambo, A., Meinel, E., Steinhilber, A., Pautsch, G., Rodríguez, D., García-Martínez, R., Villatoro, F. 2010b. Identification of Noisy Data in Databases by Means of a Clustering Process. En *Ingeniería de Software e Ingeniería del Conocimiento: Tendencias de Investigación e Innovación Tecnológica en Iberoamérica* Editores: R. Aguilar, J. Díaz, G. Gómez, E- León. Pág. 264-273. Alfaomega Grupo Editor. ISBN 978-607-707-096-2.
- Kuna, H., García Martínez, R., Villatoro, F. 2010c. Pattern Discovery in University Students Desertion Based on Data Mining. *Advances and Applications in Statistical Sciences Journal*, 22: 275-286. ISSN 0974-6811.
- Kuna, H., García Martínez, R., Villatoro, F. 2009. Identificación de Causales de Abandono de Estudios Universitarios. *Uso de Procesos de Explotación de Información*. *Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología* 5: 39-44.
- Kuna, H., Pautsch, G., Rey, M., Cuba, C., Rambo, A., Caballero, S., Steinhilber, A., García-Martínez, R., Villatoro, F. 2011. Avances en Procedimientos de la Explotación de Información con Algoritmos basados en la Densidad para la Identificación de Outliers en Bases de Datos. *Proceedings XIII Workshop de Investigadores en Ciencias de la Computación*. Artículo 3745.
- Langseth, J., Vivatrat, N. 2003. Why Proactive Business Intelligence is a Hallmark of the Real-Time Enterprise: Outward Bound. *Intelligent Enterprise* 518: 34-41.
- Lopez-Nocera, M., Pollo-Cattaneo, F., Britos, P., García-Martínez, R. 2011. Un Protocolo de Caracterización Empírica de Dominios para Uso en Explotación de Información. *Proceedings XVII Congreso Argentino de Ciencias de la Computación*. Pág. 1047-1055. ISBN 978-950-34-0756-1.
- Lyman, P., Varain, H. 2003. How Much Information?. *School of Information Management & Systems*. University of California Berkeley. <http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/>. Ultimo acceso 21 de Mayo del 2008.
- Maimon, O., Rokach, L. 2005. *The Data Mining and Knowledge Discovery Handbook*. Springer Science + Business Media Publishers.
- Mansilla, D., Pollo-Cattaneo, M., Pytel, P., García-Martínez, R. 2012. Modelo de Proceso para Elicitación de Requerimientos en Proyectos de Explotación de Información. *Proceedings del XIV Workshop de Investigadores en Ciencias de la Computación*. Pág. 222-225. ISBN 978-950-766-082-5.
- Marbán, O. 2003. *Modelo Matemático Paramétrico de Estimación para Proyectos de Data Mining (DMCoMo)*. Tesis de Doctorado en Informática. Universidad Politécnica de Madrid.
- Mobasher, B. R Cooley and J Srivastava 1999. Creating adaptive web sites through usage-based clustering of URLs. *Proceedings Workshop on Knowledge and Data Engineering Exchange*, Pág. 19-25.
- Moss, L. 2003. Nontechnical Infrastructure of BI Applications. *DM Review* 131: 42-45.
- Negash, S., Gray, P. 2008. Business Intelligence. En *Handbook on Decision Support Systems* 2, ed. F. Burstein y C. Holsapple Heidelberg, Springer, Pág. 175-193.
- Ochoa, A. Fernández, E., Britos, P., García-Martínez, R. 2008. *Metodologías de Ingeniería Informática*. Editorial Nueva Librería. ISBN 978-987-1104-54-3.
- Pollo-Cattaneo, F., Britos, P., Pesado, P., García-Martínez, R. 2010a. Ingeniería de Procesos de Explotación de Información. En *Ingeniería de Software e Ingeniería del Conocimiento: Tendencias de Investigación e Innovación Tecnológica en Iberoamérica* Editores: R. Aguilar, J. Díaz, G. Gómez, E- León. Pág. 252-263. Alfaomega Grupo Editor. ISBN 978-607-707-096-2.
- Pollo-Cattaneo, F., Britos, P., Pesado, P., García-Martínez, R. 2010b. Proceso de Educación de Expositores en Proyectos de Explotación de Información. En *Ingeniería de Software e Ingeniería del Conocimiento: Tendencias de Investigación e Innovación Tecnológica en Iberoamérica* Editores: R. Aguilar, J. Díaz, G. Gómez, E- León. Pág. 01-11. Alfaomega Grupo Editor. ISBN 978-607-707-096-2.
- Pollo-Cattaneo, M., García-Martínez, R., Britos, P., Pesado, P., Bertone, R., Rodríguez, D., Merlino, H., Pytel, P., Vanrell, J. 2012a. Elementos para una Ingeniería de Explotación de Información. *Proyecciones* 101: 67-84. ISSN 1667-8400.
- Pollo-Cattaneo, M., García-Martínez, R., Britos, P., Pesado, P., Bertone, R., Rodríguez, D., Merlino, H., Pytel, P., Vanrell, J. 2012b. Elementos para una Ingeniería de Explotación de Información. *Proyecciones* 101: 67-84. ISSN 1667-8400.
- Pyle, D. 2003. *Business Modeling and Business intelligence*. Morgan Kaufmann Publishers.
- Pytel, P., Britos, P., García-Martínez, R. 2012c. Viabilidad y Estimación de Proyectos de Explotación de Información. *Proceedings del XIV Workshop de Investigadores en Ciencias de la Computación*. Pág. 217-221. ISBN 978-950-766-082-5.
- Pytel, P., Pollo-Cattaneo, F., Rodríguez, D., Britos, P., García-Martínez, R. 2011b. Identificación de Tareas Críticas en una Metodología de Desarrollo de Proyectos de Explotación. *Proceedings XVII Congreso Argentino de Ciencias de la Computación*. Pág. 989-998. ISBN 978-950-34-0756-1.
- Pytel, P., Tomasello, M., Rodríguez, D., Pollo-Cattaneo, F., Britos, P., García-Martínez, R. 2011a. Estudio del Modelo Paramétrico DMCoMo de Estimación de Proyectos de Explotación de Información. *Proceedings XVII Congreso Argentino de Ciencias de la Computación*. Pág. 979-988. ISBN 978-950-34-0756-1.
- Pytel, P., Tomasello, M., Rodríguez, D., Arbolea, H., Pollo-Cattaneo, M., Britos, P., García-Martínez, R. 2011c. Estimación de Proyectos de Explotación de Información. *Estudio Comparado de Modelos Analíticos y Empíricos*. *Proceedings XIII Workshop de Investigadores en Ciencias de la Computación*. Artículo 3848.
- Rancan, C., Pesado, P., García-Martínez, R. 2010. Issues in Rule Based Knowledge Discovering Process. *Advances and Applications in Statistical Sciences Journal* ISSN 0974-6811, 22: 303-314. ISSN 0974-6811.
- Rodríguez, D., Pollo-Cattaneo, F., Britos, P., García-Martínez, R. 2010. Estimación Empírica de Carga de Trabajo en Proyectos de Explotación de Información. *Anales del XVI Congreso Argentino de Ciencias de la Computación*. Pág. 664-673. ISBN 978-950-9474-49-9.
- Rudin, K., Cressy, D. 2003. Will the Real Analytic Application Please Stand Up? *DM Review* 133: 30-34.
- Saavedra-Martínez, P., Pollo-Cattaneo, F., Rodríguez, D., Britos, P., García-Martínez, R. 2012a. Proceso de Identificación de Errores de Apropriación de Conceptos Basado en Explotación de Información. *Proceedings VII Congreso de Tecnología en Educación y Educación en Tecnología*. ISBN 978-987-28186-0-9. Universidad Nacional del Noroeste de Buenos Aires. Pergamino. Buenos Aires. Argentina.
- Saavedra-Martínez, P., Pollo-Cattaneo, M., Pytel, P., Rodríguez, D., García-Martínez, R. 2012b. Detección de Problemas de Aprendizaje Basado en Explotación de Información. *Proceedings del XIV Workshop de Investigadores en Ciencias de la Computación*. Pág. 962-964. ISBN 978-950-766-082-5.
- Sanson, E., Britos, P., Rodríguez, D., García-Martínez, R. 2009. Clasificación Automática para la Prevención del Estrés de los Suelos y la Fatiga de Soja en el Noroeste Argentino. *Proceedings XI Workshop de Investigadores en Ciencias de la Computación*. Pág. 333-335. ISBN 978-950-605-570-7.
- SAS, 2008. SAS Enterprise Miner. SEMMA. <http://www.sas.com/technologies/analytics/datamining/miner/semma.html>. Ultimo acceso Junio 2008.
- Srivastava, J., Cooley, R., Deshpande, M., Tan, P. 2000. Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data. *SIGKDD Explorations*, 12: 12-23.
- Thomsen, E. 2003. BI's Promised Land. *Intelligent Enterprise*, 64: 21-25.
- Valenga, F., Fernández, E., Merlino, H., Rodríguez, D., Procopio, C., Britos, P., García-Martínez, R. 2008. Minería de Datos Aplicada a la Detección de Patrones Delictivos en Argentina. *Proceedings VII Jornadas Iberoamericanas de Ingeniería de Software e Ingeniería del Conocimiento*. Pág. 31-39. ISSN 1390-292X.
- Valenga, F., Perversi, I., Fernández, E., Merlino, H., Rodríguez, D., Britos, P., García-Martínez, R. 2007a. Estudio Preliminar: La Estadística Criminal y el Aporte de la Minería de Datos. En Kaminsky, G., Kosovsky, D., Kessler, G. *El Delito en la Argentina Post-crisis*. pp 11-24. Editado por la Friedrich Ebert Stiftung.
- Valenga, F., Perversi, I., Fernández, E., Merlino, H., Rodríguez, D., Britos, P., García-Martínez, R. 2007b. Aplicación de Minería de Datos para la Exploración y Detección de Patrones Delictivos en Argentina. *Anales del XIII Congreso Argentino de Ciencias de la Computación*. Pág. 258-270. ISBN 978-950-656-109-3.
- Vanrell, J., Bertone, R., García-Martínez, R. 2010a. Modelo de Proceso de Operación para Proyectos de Explotación de Información. *Anales del XVI Congreso Argentino de Ciencias de la Computación*. Pág. 674-682. ISBN 978-950-9474-49-9.
- Vanrell, J., Bertone, R., García-Martínez, R. 2010b. Un Modelo de Procesos de Explotación de Información. *Proceedings XII Workshop de Investigadores en Ciencias de la Computación*. Pág. 167-171.
- Vegega, C., Mansilla, D., Pollo-Cattaneo, M. F., Pytel, P., Rodríguez, D., Diez, E., García-Martínez, R. 2012. Documentación de Requisitos en Proyectos de Explotación de Información. *Proceedings del XIV Workshop de Investigadores en Ciencias de la Computación*. Pág. 226-229. ISBN 978-950-766-082-5.
- Vuori, V. 2006. The Employees as a Source of External Business Information. *Proceedings European Productivity Conference EPC'06*. Pág. 29-36.

## **Elaboración y Ensayo de Aplicación de Algoritmos de Ingeniería del Conocimiento en la Gestión de Redes Sociales.**

**Daniel Xodo – [dxodo@exa.unicen.edu.ar](mailto:dxodo@exa.unicen.edu.ar)**

**Ricardo Puleo Zubillaga - [puleo@speedy.com.ar](mailto:puleo@speedy.com.ar)<sup>1</sup>**

**Gustavo Tripodi - [gtripodi@exa.unicen.edu.ar](mailto:gtripodi@exa.unicen.edu.ar)**

**Gustavo Illescas - [illescas@exa.unicen.edu.ar](mailto:illescas@exa.unicen.edu.ar)**

**Moises Bueno - [bueno@econ.unicen.edu.ar](mailto:bueno@econ.unicen.edu.ar)**

Instituto de Investigación en Tecnología Informática Avanzada (INTIA) – Departamento de Computación y Sistemas - Facultad de Ciencias Exactas - Universidad Nacional del Centro de la Provincia de Buenos Aires (UNCPBA). Teléfono: +54 249-4439680. Dirección postal: Campus Universitario, Paraje Arroyo Seco, (7000) Tandil, ARGENTINA

<sup>1</sup> GESTADII/Departamento Ingeniería Industrial/Facultad Regional Trenque Lauquen/Universidad Tecnológica Nacional. Racedo 298. (02392) 422023

### **Resumen.**

El estado, a través de sus planes sociales, y distintas ONG ofrecen ayuda social a sectores vulnerables de nuestro país. Pero dicha ayuda social demanda que la asignación de esos recursos sea optimizada para que cada grupo familiar obtenga los recursos necesarios de acuerdo a sus necesidades. El objetivo de este proyecto es generar herramientas informáticas de gestión que analice y optimice la asignación de dichos recursos.

### **Palabras clave:**

Indicadores. Optimización. Ontologías. Cluster. Lógica Difusa. Multicriterio.

### **Contexto.**

La propuesta esta contenida en el proyecto MERAIS V (Métodos de Razonamiento Aproximado en la Investigación Socioeconómica), dentro del Instituto de Tecnología Informática Avanzada (INTIA) del Grupo Informática de Gestión, Universidad Nacional del Centro de la provincia de Buenos Aires. Es continuación de los anteriores MERAIS IV (Análisis de la

Aplicabilidad del BSC en el desarrollo de Ontologías en Informática de Gestión (03/C126), MERAIS III (Data Mining Ontology, 03/C189), MERAIS II (Sistema Basado en Conocimiento para la Selección de Técnicas de Análisis de Datos, 03/C155) y MERAIS I (03/C113). Aprobados por Consejo Superior UNCPBA y del proyecto Análisis de la Aplicabilidad del BSC en el desarrollo de Ontologías en Informática de Gestión (Codigo UTN 1147) UTN FRBB:

### **Introducción.**

En la actualidad las redes sociales, en especial las que trabajan en la atención de las necesidades primarias insatisfechas, enfrentan una problemática común, que la demanda supera ampliamente la oferta de recursos. Este hecho hace obligatorio establecer criterios que con asistencia de herramientas de gestión y análisis de procesos, optimicen en el mayor grado posible, la asignación de los escasos recursos.

El rol del estado, a través de sus planes sociales o programas de desarrollo social, que suelen ser puntos “sensibles” en la acción de los municipios, constituye una de las



organizaciones que más demanda la optimización de la asignación de recursos para satisfacer las necesidades primarias de sus ciudadanos.

El presente proyecto está integrado por una serie de etapas tendientes a generar un conjunto de herramientas informáticas de gestión, a fin de analizar y optimizar los procesos de asignación de los recursos, trabajando sobre datos reales de ONG's dedicadas a la ayuda social.

Las etapas pueden establecerse como:

- 1- Relevamiento de la demanda.
- 2- Relevamiento de los recursos (actuales y potenciales) existentes.
- 3- Fijación de criterios de asignación de recursos.
- 4- Implementación.
- 5- Monitoreo y seguimiento, con mejora continua de procesos.

#### **Etapas 1:** Relevamiento de la demanda.

La principal acción en esta etapa consistirá en analizar cuáles son los programas existentes de las organizaciones sociales y relevar, mediante encuestas, cuales son las necesidades presentes en los beneficiarios de esos programas. El relevamiento cubrirá una serie de factores de análisis personales, de las viviendas y del grupo familiar, entendiéndose por tal a todos los individuos que viven bajo el mismo techo independientemente de la condición filiatoria. Se procederá a la categorización de las necesidades y la definición formal de las relaciones: "la necesidad N es evaluada en la pregunta P".

La categorización implica la selección y combinación de factores por métodos que garanticen la equidad y la satisfacción de las condiciones de vida digna a la cual toda persona tiene derecho, y cuya definición puede sufrir variaciones según regiones, subculturas, grupos sociales, etc.

Esta categorización, previa a las decisiones de asignación, conlleva la aplicación de algoritmos de análisis "multiobjetivo" y "multicriterio" a fin de mantener la equidad en la gestión de los recursos. (Barba Romero, 1998), (Jie, 2007), (Saaty, 2008)

**Etapas 2:** Relevamiento de recursos (actuales y potenciales) existentes. En esta etapa serán estudiadas las fuentes y los recursos con que cuentan las distintas organizaciones sociales que asisten a los necesitados. Se procederá a la categorización de los mismos y la definición formal de la relación "el recurso R satisface la necesidad N".

**Etapas 3:** fijación de criterios de asignación de recursos.

Habiendo realizado las etapas 1 y 2, se dispone de la regla "la persona X, a través de la respuesta a la pregunta P, presenta la necesidad N que se satisface con el recurso R". Naturalmente habrá en general varias personas que demanden los mismos recursos.

En esta etapa serán establecidos los criterios de asignación de recursos a fin de optimizar la distribución de los que son demandados concurrentemente.

#### **Etapas 4:** Implementación.

El sistema a desarrollar está basado en la acción inter institucional con el agregado de control conjunto entre las organizaciones participantes.

Son las instituciones las que conocen las necesidades de sus asistidos y los recursos con que cuentan, pero podría darse el caso de que sea "sobreasistido" o "no asistido" a un beneficiario por ignorar lo que sucede con otras instituciones que operan en el mismo sector.

**Etapas 5:** Monitoreo y seguimiento, con mejora continua de procesos.

Lo expuesto pone en evidencia que realizar este proyecto implica lograr un fuerte compromiso de las partes intervinientes, para que el sistema entre en un ciclo de "mejora continua".

## **Líneas de investigación y desarrollo.**

Dada la complejidad y el volumen de los datos con los que se trabajará para elaborar las asignaciones, el proyecto limitará su análisis y aplicación de algoritmos a un segmento muestral a fin de poder operar



experimentalmente con los modelos seleccionados sin que sea requeridos por la naturaleza del modelo a aplicar y dirigidos a:

1. Desarrollar sistemas de indicadores relativos al funcionamiento de una red social (concretamente en el caso de análisis, una ONG de ayuda a carecientes).
2. Elaborar y ensayar algoritmos de formalización y optimización para la gestión de redes mediante algoritmos algebraicos y técnicas probabilistas.
3. Elaborar un SIBO (Sistema de Información Basado en Ontologías) de la red social utilizada para el análisis.
4. Aplicar técnicas multicriterio de análisis de decisiones a la gestión de redes y comparar con las adoptadas y disponibles en las bases de datos.
3. Calcular las asignaciones de ayuda a familias en forma individual y agregada mediante modelos multicriterio AHP (Analytic Hierarchy Process) de asignación de ayudas individuales y aplicarlos a los requerimientos establecidos para determinar las asignaciones apropiadas. Desarrollar indicadores de gestión para evaluar las asignaciones.
4. Comparar mediante Lógica Difusa y Análisis de Agrupamientos las asignaciones realizadas en periodos anteriores con las definidas por el modelo multicriterio a fin de verificar la conveniencia de reemplazar los métodos empíricos de asignación de ayuda por un modelo informatizado propuesto y elaborar un sistema de indicadores.
5. Desarrollar e implementar el sistema de aplicación del modelo de asignaciones para su utilización por las ONG's.

## Resultados y Objetivos.

Los objetivos generales del proyecto son:

1. Analizar y definir las características socioeconómicas de las familias que participan de los programas de ayuda social en función de los parámetros preestablecidos y agruparlas para su tratamiento.
2. Analizar los requerimientos de ayuda económica, por familia y agregados, definirlos mediante ontologías computacionales para su procesamiento y desarrollar modelos de gestión de asignaciones, control, evaluación y pronóstico de oferta y demanda de bienes.

Los objetivos específicos son:

1. Diferenciar, clasificar y agrupar las familias según composición, necesidades permanentes y circunstanciales (enfermedades, nacimientos, etc.) mediante técnicas de Cluster Analysis para operar computacionalmente con las unidades y su agregación a fin de optimizar las asignaciones a realizar.
2. Seleccionar el segmento apropiado para la elaboración de un SIBO (Sistema basado en Ontologías) seleccionando las necesidades permanentes de los grupos en estudio para su diseño.

## Formación de Recursos

### Humanos.

Apellido y nombre	Cargo Docente Actual/Posgrado		Funciones
	Cat.	Dedic.	
Xodo, Daniel	Tit.	Semi.	Director
Tripodi, Gustavo	Adj.	Semi.	Co-Director
Bueno, Moisés	Adj.	Simple	Investig. formado
Matassa, Marcelo	Adj.	Semi	Investig. de apoyo
Illescas, Gustavo	Adj.	Excl.	Investig formado
Lagrang, Iván	JTP	Simple	Becario
Dos Reis, Maria	JTP	Excl.	Investig. de apoyo
Trotti, Marcos	Al.	Simple	Investig. estudiante
Gil, Guillermo	Al.	Simple	Investig. de apoyo
Puleo, Ricardo	Ay. Grad	Simple	Investig. de apoyo
Amador, Adrian	Adj.	Simple	Investig. de apoyo

Tesis			
Apellido y Nombre	Tipo de Tesis	Institución	Año
Illescas Gustavo	Doctoral	Fac.Cs Exactas - UNCPBA	2015
Tripodi Gustavo	Doctoral	Fac.Cs Económicas - UNCPBA	2015
Dos Reis Rosa	Doctoral	Fac.Cs Económicas - UNCPBA	2015
Puleo Ricardo	De Grado	UTN FRTL	2012
Tripodi Josefina	De Grado	Fac.Cs Económicas - UNCPBA	2015

## Referencias.

Barba-Romero, S. 1998. Conceptos y soportes informáticos de la decisión multicriterio discreta. Evaluación y Decisión Multicriterio: reflexiones y Experiencias.

E. Martinez y M. Escudey. Eds. Editorial Universidad de Chile. Santiago, USACH, UNESCO. ISBN:956-7069-28-X  
www.inescc.pt/~ewgmcda/MartinexBook.html

Boratto, M, D. Gimenez, and A.Vidal. 2006. Automatic parametrization on Divide-and-Conquer Algorithms. In International Congress of Mathematicians, pages 495-496, [http://riunet.upv.es/bitstream/handle/10251/12388/tesisMaster\\_DoCarmo\\_Murilo\\_2007.pdf?sequence=1](http://riunet.upv.es/bitstream/handle/10251/12388/tesisMaster_DoCarmo_Murilo_2007.pdf?sequence=1)

Alonso Betanzos A. e.a. 2004. Ingeniería del Conocimiento. Aspectos Metodológicos. Madrid: Pearson. ISBN: 978-842-054-192-1

Chao, Chih-Yang, Yi-Li Huang & Min-Yu Wang. 2006. An application of the Analytic Hierarchy Process(AHP) for a competence analysis of technology managers from the manufacturing industry in Taiwan. World Transactions on Engineering and Technology Education Vol.5N°1 2006. UNESCO International Centre for Engineering

Education.

[www.eng.monash.edu.au/uicee/worldtransactions/WordTransAbstractsVol5No1/11\\_Chao51.pdf](http://www.eng.monash.edu.au/uicee/worldtransactions/WordTransAbstractsVol5No1/11_Chao51.pdf)

Davenport, T.1993. Process Innovation. Harvard Business School Press. Boston. Massachusetts

Del Moral, A.e.a.2007. Gestión del Conocimiento. Thompson. Madrid. ISBN:978-84-9732-548-6

Groppa, O. 2004. Las necesidades humanas y su determinación Los aportes de Doyal y Gough, Nussbaum y Max-Neef al estudio de la pobreza. Instituto Para la Integración del Saber. Universidad Católica Argentina <http://uca.edu.ar/uca/common/grupo32/files/Las-necesidades-Groppa-2004.pdf>

Guarino, N. 1998. Formal Ontology and Information Systems. Proceedings of FOIS'98. National Research Council. LADSEB 1998. <http://citeseer.ist.psu.edu/guarino98formsl.htm>

Jie Lu, Guangquan Zhang, Da Ruan, Fengjie Wu. 2007. "Multiobjective Group Decision Making" Imperial College Press. Capítulo 9, págs 189-206. Linoff, G. and Berry. 1997.Data Mining, Techniques, For Marketing, Sales, and Customer Support, John Wiley & Sons Inc.,

Kaplan N. y D. Norton. 2004. Mapas Estratégicos. Gestión 2000. ISBN: 84-8088-486-X.

Kaplan, R.y D. Norton. 1997. The Balanced Scorecard. Gestión 2000. Barcelona.

Kosko, B. 1992. Neural Networks and Fuzzy Systems, A dynamical Systems Approach to Machine Intelligence

Marakas George. 1999. Decision Support Systems in the 21st Century Prentice Hall Uppersdale River. New Jersey. ISBN: 0-13-744186-X

Pajares Martizanz, G. y M. Santos Penas. 2006. Inteligencia Artificial e Ingeniería del Conocimiento. Alfaomega-Rama. 2006. ISBN 970-15-1166-2. pag. 73-74

Rodríguez Bello, Sandra. 2007. Toma de Decisión Multicriterio con AHP, ANP y Lógica Difusa. Seminario de Investigación. Maestría en Ingeniería de Sistemas y Computación. Universidad Nacional de Colombia.  
<http://www.srodriguezbo.googlepages.com/presentacion.pdf>

Rodríguez, S. Toma Decisión Multicriterio con AHP, ANP y Lógica Difusa. Maestría en Ingeniería de Sistemas y Computación. Universidad Nacional de Colombia. Consultado el 4 de Febrero de 2009 en: [www.srodriguezbo.googlepages.com](http://www.srodriguezbo.googlepages.com)

Saaty, T. 2008. Decision Making with the Analytic Hierarchy Process, Int. J. Services Sciences, Vol. 1, N°. 1, pp. 83-98  
<http://inderscience.metapress.com/media/pgwf2q5yhg2quk9jqift/contributions/0/2/t/6/02t637305v6g65n8.pdf>

Saaty, Thomas L. 2008. Decision making with the analytic hierarchy process, Int. J. Services Sciences, Vol. 1, N°. 1, pp. 83-98  
<http://inderscience.metapress.com/media/pgwf2q5yhg2quk9jqift/contributions/0/2/t/6/02t637305v6g65n8.pdf>

Schreiber, G., Crubézy M., y M. Musen. 2000. A Case Study in Using Protégé 2000 as a Tool for KommonKADS. EKAW' Proceedings of the 12<sup>th</sup> European Workshop on Knowledge Acquisition. London. ISBN: 3-540-41119-4  
<http://www.cs.vu.nl/~schreiber/papers/Schreiber00a.pdf>

Schmoltdt, Daniel L.; D.L. Peterson y Robert L. Smith, 1994. The Analytic Hierarchy Process and Participatory Decisionmaking. Decision Support 2001. Vol 1. p. 129-143  
 <<http://www.srs4702.forprod.vt.edu/pubsubj/pdf/9501.pdf>.>

Spann, M. and Bernd Skiera. 2003. Internet-Based Virtual Stock Markets for Business Forecasting Management Science Vol. 49, No. 10, Special Issue on E-Business and Management Science (Oct., 2003), pp. 1310-1326. Published by: INFORMS  
 Stable URL:  
<http://www.jstor.org/stable/4134008>

Taslicali, A.K. and S. Ercan. 2006. The Analytic Hierarchy & The Analytic Network Processes In Multicriteria Decision Making: A Comparative Study. Journal Of Aeronautics And Space Technologies .Volume 2 Number 4 (55-65)  
[http://www.hho.edu.tr/hutendergi/2006Temmuz/10\\_TASLICALI\\_ERCAN.pdf](http://www.hho.edu.tr/hutendergi/2006Temmuz/10_TASLICALI_ERCAN.pdf).

# Bases de Datos de Objetos No Estructurados

**Anabella De Battista , Andrés Pascal,**

**Pablo Gancharov, Melisa Argüello, Christian Saliwonczyk**

Departamento de Sistemas de Información

Fac. Reg. Concepción del Uruguay

Universidad Tecnológica Nacional

Entre Ríos, Argentina

{debattistaa, pascalj, gancharovp, arguellom, saliwonczyk}@frcu.utn.edu.ar

**Norma Edith Herrera**

Departamento de Informática

Univ. Nac. de San Luis

San Luis, Argentina

nherrera@unsl.edu.ar

**Gilberto Gutierrez**

Facultad de Ciencias Empresariales

Universidad del Bio-Bio

Chillán, Chile

ggutierr@ubiobio.cl

## Resumen

En las bases de datos tradicionales es frecuente el procesamiento de consultas por exactitud o por rango de valores susceptibles de ser ordenados, sobre datos estructurados en registros de tamaño fijo compuestos por campos comparables. La necesidad de almacenar otros tipos de datos tales como los objetos multimediales (imágenes, video, texto) y el hecho de que estos datos no puedan estructurarse, obligó a extender las capacidades de las bases de datos; pero en la mayoría de los casos sólo se permiten el almacenamiento y alguna funcionalidad adicional. Por ello resulta necesario desarrollar nuevos enfoques para almacenar y la buscar objetos no estructurados eficientemente. En estos nuevos modelos la búsqueda exacta carece de interés y en muchos casos se requiere mantener los distintos estados de la base de datos a través de tiempo y no sólo el más reciente, para poder consul-

tar información histórica. Como solución han surgido modelos como el espacial, temporal, espacio-temporal, espacios métricos y el modelo métrico-temporal, que permiten representar y manipular estos tipos de datos. El tema de estudio del *Grupo de Investigación en Bases de Datos (GIBD)*, es el modelado de objetos no estructurados y el procesamiento eficiente de consultas sobre estos tipos de datos.

**Palabras Claves:** Bases de Datos Espaciales, Bases de Datos Espacio-Temporales, Espacios Métricos, Índices, Espacios Métrico-Temporales.

## 1. Contexto

El presente trabajo se desarrolla en el ámbito del proyecto *Métodos de acceso, consultas y aplicaciones en modelos de bases de datos no convencionales* (PID 25-D040) del Grupo de Investigación en Bases de Datos, pertene-



ciente al Departamento Ingeniería en Sistemas de Información de la Universidad Tecnológica Nacional, F. R. Concepción del Uruguay.

## 2. Introducción

Las bases de datos clásicas se organizan bajo el concepto de búsqueda exacta sobre datos estructurados. Esto significa que la información se organiza en registros los cuales se dividen en campos que contienen valores completamente comparables. Una consulta retorna todos aquellos registros cuyos campos coinciden con los aportados en la consulta (búsqueda exacta). Por otro lado, otra característica importante de las bases de datos clásicas es que capturan sólo un estado de la realidad modelada, usualmente el más reciente. Por medio de las transacciones, la base de datos evoluciona de un estado al siguiente descartando el estado previo.

En la actualidad es necesario implementar nuevas estrategias de almacenamiento y búsqueda para nuevos modelos de bases de datos, que permiten almacenar datos no estructurados tales como imágenes, sonido, texto, video, datos geométricos, etc. Las características principales de estos nuevos tipos de datos es que no poseen una estructura uniforme, por lo cual los índices tales como el *B\*-Tree* no se pueden utilizar para hacer más eficiente la búsqueda, las consultas por igualdad carecen de interés, y en algunos casos es un requisito mantener todos los estados de la base de datos y no sólo el más reciente. En este contexto se han generado los nuevos modelos que describimos brevemente a continuación.

Las *Bases de Datos Espaciales* permiten procesar objetos con alguna referencia espacial. Un dato espacial puede ser en su forma más simple un punto, una polilínea o un polígono. La persistencia de estos tipos de datos espaciales se basa no sólo en el valor de ciertos atributos, sino también en la ubicación espacial del objeto. Por ejemplo, podría resul-

tar de interés obtener los terrenos geográficamente adyacentes a uno dado, o encontrar todos los hospitales cercanos a una determinada ruta. Existen muchas aplicaciones para el modelo de bases de datos espaciales; una de las más destacadas son los sistemas de información geográfica (SIG), que realizan el procesamiento de datos geográficos y que almacenan la geometría y los atributos de datos con algún tipo de georreferencia, es decir, situados en la superficie de la tierra y representados bajo una proyección cartográfica. Uno de sus objetivos es resolver problemas complejos de planificación y gestión.

Las *Bases de Datos Temporales* manejan internamente una o más dimensiones temporales, permitiendo asociar tiempos a los datos almacenados. Existen tres clases de bases de datos temporales según el modo en que manejan el tiempo: (a) de tiempo transaccional (transaction time), donde el tiempo se registra de acuerdo al orden en que se procesan las transacciones; (b) de tiempo vigente (valid time), que almacenan el momento en que el hecho ocurrió en la realidad, que puede no coincidir con el momento de su registro; y (c) bitemporales, que integran la dimensión transaccional y la dimensión vigente a través del versionado de los estados. En las consultas se requiere conocer el comportamiento de algún objeto en algún instante dado o durante un intervalo de tiempo determinado. Por ejemplo una consulta temporal podría ser *recuperar la evolución del sueldo de un empleado en un intervalo de tiempo dado, o encontrar todos los empleados que tenían cierta categoría en una fecha dada*.

Los *Espacios Métricos* constituyen un modelo de bases de datos orientado al almacenamiento de objetos no estructurados, que permite realizar consultas por similitud eficientemente. Este tipo de consultas utiliza funciones de distancia para determinar el grado de similitud entre los objetos de la base de datos y el objeto que se consulta. Un *Espacio Métri-*

co se define como un par  $(U, d)$  donde  $U$  es el universo de objetos válidos del espacio y  $d : U \times U \rightarrow R^+$  es una función de distancia definida entre los elementos de  $U$  que mide su similitud (a menor distancia más cercanos o similares son los objetos). Llamaremos base de datos a cualquier subconjunto finito  $X \subseteq U$  cuya cardinalidad es  $|X| = n$ . La función  $d$  cumple con las propiedades características de una función métrica:  $\forall x, y \in U, d(x, y) \geq 0$  (positividad);  $\forall x, y \in U, d(x, y) = d(y, x)$  (simetría);  $\forall x \in U, d(x, x) = 0$  (reflexividad) y  $\forall x, y, z \in U, d(x, y) \leq d(x, z) + d(z, y)$  (desigualdad triangular). En base a este modelo se han desarrollado índices especiales que aumentan la velocidad de respuesta de las búsquedas por similitud.

Estos tres tipos de bases de datos se pueden combinar para resolver consultas complejas que involucran más de un aspecto de los anteriormente descritos. Así han surgido los modelos *Espacio-Temporal* y *Métrico-Temporal*.

Las *Bases de Datos Espacio-Temporales* tratan con objetos que cambian su identidad, su posición o su forma en el tiempo. Las consultas a resolver en este tipo de bases de datos pueden incluir referencias espaciales, tales como posición, intersección, inclusión o superposición, y temporales, tanto respecto al pasado o presente como predicciones del tiempo futuro. Por ejemplo, nos puede interesar saber cuál es la máxima velocidad alcanzada por un objeto en un intervalo de tiempo, o recuperar los objetos que cruzaron una cierta área en un instante de tiempo dado o incluso los que pasarán por un punto en el futuro, si es que mantienen su dirección. Entre las aplicaciones que tratan con este tipo de bases de datos se incluyen las de predicción climática, control de tráfico terrestre o aéreo, aspectos sociales (demografía, salud) y multimedia.

El *Modelo Métrico-Temporal* surge ante la necesidad de aplicaciones donde resulta de interés realizar búsquedas por similitud teniendo en cuenta también la componente temporal.

En este modelo se puede trabajar con objetos no estructurados con tiempos de vigencia asociados y realizar consultas por similitud y por tiempo en forma simultánea. Formalmente un *Espacio Métrico-Temporal* es un par  $(U, d)$ , donde  $U = O \times N \times N$ , y la función  $d$  es de la forma  $d : O \times O \rightarrow R^+$ . Cada elemento  $u \in U$  es una tripla  $(obj, t_i, t_f)$ , donde  $obj$  es un objeto (por ejemplo, una imagen, sonido, cadena, etc) y  $[t_i, t_f]$  es el intervalo de vigencia de  $obj$ . La función de distancia  $d$ , que mide la similitud entre dos objetos, cumple con las propiedades de una métrica (positividad, simetría, reflexividad y desigualdad triangular). Una *consulta métrico-temporal* por rango se define como una 4-upla  $(q, r, t_{iq}, t_{fq})_d$ , tal que  $(q, r, t_{iq}, t_{fq})_d = \{o / (o, t_{io}, t_{fo}) \in X \wedge d(q, o) \leq r \wedge (t_{io} \leq t_{fq}) \wedge (t_{iq} \leq t_{fo})\}$ .

### 3. Líneas de Investigación

Nuestra principal línea de trabajo es el estudio de métodos de acceso, procesamiento de consultas y aplicaciones de bases de datos no tradicionales, centrándonos principalmente en los modelos métrico-temporal y espacio-temporal. Damos a continuación una descripción de las líneas de investigación que actualmente estamos desarrollando.

#### 3.1. Consultas Métrico Temporales sobre Cadenas

Hasta el momento se han propuesto cuatro índices métrico-temporales: el *FHQT-Temporal* [6], el *Historical-FHQT* [2], el *Event-FHQT* [5] y el *Pivot-FHQT* [3] todos ellos han tomado como base el índice para espacios métricos Fixed Height Queries Tree[1], que trabaja con funciones de distancia discretas. Además se han diseñado las variantes *FHQT<sup>+</sup>-Temporal* y *Event-FHQT<sup>+</sup>* que permiten tanto funciones discretas como continuas.

Para probar la eficiencia de los índices se

desarrolló una aplicación que tiene por finalidad permitir efectuar consultas métrico-temporales sobre el sistema de archivos de los sistemas operativos (Windows/Linux). Esta aplicación está orientada a la búsqueda por similitud de archivos y carpetas tanto por nombre como por fecha, con diferentes radios de búsqueda, y utiliza índices métrico-temporales que disminuyen significativamente el tiempo de respuesta.

### 3.2. Búsqueda de Imágenes

En la búsqueda de imágenes por similitud en grandes bases de datos, es tan importante la eficiencia del sistema (recuperar imágenes en un tiempo razonable) como su eficacia (recuperar imágenes que sean realmente de interés). La eficacia depende principalmente del preprocesamiento de las imágenes, de la técnica de extracción de características y de la función de distancia que se emplee. Por otro lado, los factores de mayor relevancia para la eficiencia del proceso son el costo de la función de distancia y el tipo de índice que se utilice para acelerar la búsqueda.

Un verdadero sistema de recuperación de imágenes debe permitir dar una imagen como objeto de consulta y debe poder determinar la similitud entre ese objeto y cada una de las imágenes de la base de datos en forma eficiente, a fin de responder la consulta.

La búsqueda por similitud aplicada a imágenes implica transformar las imágenes en vectores de características que luego se insertan en un índice métrico. Luego, ante una consulta, se transforma la imagen de consulta de la misma manera para poder buscar usando el índice.

Existen dos tareas que son cruciales en este proceso: una es convertir las imágenes en vectores y la otra definir una función de distancia que permita comparar las imágenes. La primera tarea afecta directamente la eficacia del sistema dado que las búsquedas se realizarán en base a las características extraídas de cada ima-

gen. La segunda tarea afecta tanto la eficacia como la eficiencia; la eficacia porque la función de distancia modela formalmente lo que se entiende por similitud y la eficiencia porque el costo de búsqueda en el índice se ve directamente afectado por el costo de cálculo de la función de distancia y por la distribución de distancias que genera.

Si bien hay numerosos trabajos de investigación que se concentran en el preprocesamiento de las imágenes y extracción de características [8], las funciones de distancia [7] y los índices métricos [4], la mayoría lo hace por separado, sin estudiar la integración de estos aspectos.

En esta línea hemos trabajado definiendo un proceso para el tratamiento integral de las bases de datos de imágenes.

### 3.3. Aplicaciones de Bases de Datos Espaciales y Sistemas de Información Geográfica

En el marco de este proyecto se han firmado convenios de colaboración con otras instituciones y grupos de investigación con el fin de prestar servicios relacionados a la temática del grupo. Se colaboró con el Grupo de Estudios de Calidad y Medio Ambiente de esta Facultad en la elaboración de un informe para analizar y describir el sector comercial y de servicios de la ciudad de Concepción del Uruguay a fin de obtener una herramienta de planificación. Con la Secretaría de Desarrollo Social del Municipio de Concepción del Uruguay se firmó un convenio para desarrollar e implementar una herramienta SIG (Sistema de Información Geográfica) para la LINEA 102 (Línea de los Derechos) que sirva como herramienta de planificación y soporte a la toma de decisiones, mediante la visualización en un mapa de la ciudad de las direcciones asociadas a las denuncias telefónicas recibidas por dicho servicio de atención telefónica y la vinculación de esta capa con otras de interés. Con la Facultad de Ciencias de la Salud de la Univ. Nac.

de Entre Ríos se estableció un convenio para el desarrollo y mantenimiento de un servidor de mapas interactivo en el que se visualizan datos georreferenciados resultantes de diversos proyectos de investigación. Actualmente se está trabajando en el desarrollo de un Sistema de Información Geográfica para el municipio de la localidad de Caseros, Entre Ríos, que permitirá georreferenciar la capa catastral de la localidad y asociar dicha base de datos a la gestión de tasas municipales y posteriormente servirá como herramienta de planificación para la gestión municipal.

#### 4. Resultados Esperados

Se espera contar con métodos eficientes, tanto en memoria principal como en memoria secundaria, para el procesamiento de consultas en el ámbito de bases de datos no tradicionales. Esto incluye el diseño de índices, la definición de funciones de distancias adecuadas a la problemática tratada, la definición de nuevas consultas que sean de interés y el desarrollo de aplicaciones en ámbitos reales de uso de los métodos desarrollados.

#### 5. Formación de Recursos Humanos

El trabajo desarrollado hasta el momento forma parte del desarrollo de dos Tesis de Maestría en Ciencias de la Computación. Uno de los integrantes del grupo está desarrollando su Tesis Doctoral sobre la temática de indexación en memoria secundaria de bases de datos textuales, tema íntimamente relacionado a las líneas de estudio de este grupo. El grupo cuenta en la actualidad con tres alumnos becarios que se están formando en estas temáticas y se han desarrollado hasta la fecha cinco tesinas de grado en el marco del proyecto.

#### Referencias

- [1] R. Baeza-Yates, W. Cunto, U. Manber, and S. Wu. Proximity matching using fixed-queries trees. In *Proc. 5th Combinatorial Pattern Matching (CPM94)*, LNCS 807, pages 198–212, 1994.
- [2] A. De Battista, A. Pascal, G. Gutierrez, and N. Herrera. Un nuevo índice métrico-temporal: el historical fhqt. In *Actas del XIII Congreso Argentino de Ciencias de la Computacion*, Corrientes, Argentina, 2007.
- [3] A. De Battista, A. Pascal, N. Herrera, and G. Gutierrez. Metric-temporal access methods. *Journal of Computer Science & Technology*, 10(2):54–60, 2010.
- [4] E. Chavez, G. Navarro, R. Baeza-Yates, and J.L. Marroquin. Searching in metric spaces. *ACM Computing Surveys*, 33(3):273–321, September 2001.
- [5] A. Pascal, A. De Battista, G. Gutierrez, and N. Herrera. Índice métrico-temporal event-fhqt. In *Actas del XIII Congreso Argentino de Ciencias de la Computacion*, La Rioja, Argentina, 2008.
- [6] A. Pascal, De Battista, G. Gutierrez, and N. Herrera. Procesamiento de consultas métrico-temporales. In *XXIII Conferencia Latinoamericana de Informatica*, pages 133–144, San Jose de Costa Rica, 2007.
- [7] Y. Rubner, C. Tomasi, and L.J. Guibas. The earth movers distance as a metric for image retrieval. *International Journal of Computer Vision*, 40:99–121, 2000.
- [8] A. Shahbahrani and D. B. Juurlink. Comparison between color and texture features for image retrieval. In *Proceedings of the 19th Annual Workshop on Circuits, Systems and Signal Processing*, 2008.



## Almacenamiento de datos XML en Oracle 11g

Lic. Silvina Migani, Lic. Cristina Vera, Ing. Carlos Correa, Lic. Liliana Romera

Departamento de Informática / Facultad de Ciencias Exactas, Físicas y Naturales / UNSJ

Av. Ignacio de la Roza 590 (O)

Teléfonos: 4260353 - 4260355

silvina.migani@gmail.com; civera2@yahoo.com.ar; carcorrea@hotmail.com.ar; liliromera@yahoo.com.ar

### Resumen

Sin duda la aparición del lenguaje de marcas extensible, XML (eXtensible Markup Language), generó muchos cambios en el mundo de la informática. Por un lado, la introducción de muchas tecnologías asociadas, y por otro, la transformación o adecuación de las ya existentes. El mundo de las bases de datos no quedó fuera de tal situación. Los fabricantes de los SGBDs tradicionales se vieron en la necesidad de extender sus estructuras habituales para poder manipular datos XML, dando origen a los SGBDs extendidos (SGBDR\_XML). Conjuntamente, surgieron los SGBDs XML nativos (NXDB), que adoptaron el tipo XML como estructura de datos intrínseca.

Este trabajo tiene como finalidad, presentar y explicar algunos aspectos significativos en cuanto a la elección del modo de almacenamiento de datos XML en el Sistema de Gestión de Bases de Datos Oracle 11g.

**Palabras clave:** Bases de Datos – XML

### Contexto

Este artículo expone parte del trabajo realizado dentro de las actividades planificadas en el Proyecto de Investigación “XML: TÉCNICAS DE GESTIÓN E INTERCAMBIO DE DATOS” - 21/ E915, aprobado por el CICITCA en Octubre de 2011. De esta manera, contribuye al objetivo general de dicho proyecto: “Estudiar y experimentar la tecnología XML en distintos tipos de motores de bases de datos”.

### Introducción

XML, al igual que el lenguaje de marcas de hipertexto HTML (Hyper Text Markup Language) sobre el que está basado World Wide Web, tiene sus raíces en la gestión de documentos y está derivado de un lenguaje para estructurar grandes documentos, conocido como lenguaje estándar generalizado de marcas SGML (Standard Generalized Markup Language) [1] [3]. Sin embargo, a diferencia de SGML y HTML, XML puede representar datos estructurados usados en aplicaciones de negocios.

Un archivo escrito en XML además de proporcionar información se describe asimismo, constituyéndose en un formato ideal para materializar el intercambio de datos entre aplicaciones. El receptor puede entender la información recibida y, por lo tanto, procesarla. Tiene el formato de un archivo de texto plano, lo que facilita enormemente la transferencia de información, logrando además independencia con respecto a diferentes plataformas.

La utilización masiva de esta nueva forma de representación de datos en muchos contextos, provocó la necesidad de gestionarlos [2]:

- Los SGBDs tradicionales extendieron sus capacidades para poder soportar datos XML [4][6][9]. En consecuencia, el lenguaje estándar utilizado por los motores de bases de datos relacionales, SQL, se amplió con el objetivo de manejar estructuras XML, dando lugar a la aparición del SQL/XML (ISO/IEC 9075-14:2006) [5].
- Surgió un nuevo tipo de motor de bases de datos, los SGBDs XML nativos. La estructura de datos subyacente en estos motores es precisamente XML. Además, proveen lenguajes XML específicos, tales como XPath y XQuery [2].

Este trabajo aborda específicamente aspectos relacionados al primer grupo, los SGBDR\_XML. Particularmente se eligió Oracle 11g dado su posicionamiento aventajado en el mercado y al importante conjunto de tecnologías relacionadas al almacenamiento y recuperación de datos XML que incluye desde su versión 9iR2, llamado Oracle XML DB [7].

Oracle permite almacenar datos XML fuera de la base de datos, en un repositorio, y acceder a sus datos a través de técnicas basadas en el lenguaje XPath.

Asimismo, permite almacenar documentos XML dentro la base de datos, específicamente en tablas. Para ello provee un tipo de datos nativo llamado XMLType.

**1. Modelos de almacenamiento XMLType**

Este nuevo tipo de dato abstracto permite guardar datos en formato XML. Brindando la posibilidad de utilizar esquemas XML [10], XPath, XQuery [3], XSLT [11], indexación y particionamiento de documentos XML.

En las versiones Oracle Database 9i y 10g los documentos XML se almacenan como CLOB o como objetos. A partir de la versión Oracle Database 11g se agregó una nueva posibilidad, almacenarlos en formato binario (Binary XML), que de hecho es el formato por defecto a partir de la versión 11.2.0.2 [7][8].

*a) Almacenamiento no estructurado*

Los datos XMLType se almacenan como un CLOB (Character Large Object). Los documentos XML son almacenados preservando el documento original, inclusive los espacios en blanco. Así, se mantiene fidelidad de contenido o de documento. Además, Oracle brinda la posibilidad de asociar un esquema en esta forma de almacenamiento. Provee un desempeño muy bueno en las operaciones de inserción y recuperación de los documentos completos.

Sin duda es una opción que proporciona una gran flexibilidad, pero por otro lado, requiere una sobrecarga en el procesamiento cuando se necesita consultar su contenido, como por ejemplo cuando se usan las funciones XMLType.Extract y XMLType.ExistsNode. Evaluar estas funciones requiere construir el árbol XML DOM en memoria y sobre él evaluar las expresiones XPath. Además, toda operación de actualización debe realizarse a nivel de documento [7].

*b) Almacenamiento estructurado*

En este caso, los datos XMLType se almacenan como un conjunto de objetos. Los documentos XML mantienen fidelidad DOM (Documento Object Model). Esta alternativa también es mencionada como almacenamiento Objeto-Relacional (OR) [7].

Es común y conveniente almacenarlos con un esquema asociado, puesto que permite acelerar las

consultas y las actualizaciones de granularidad fina. Esto provoca que las aplicaciones que lo utilizan deban ajustarse a un esquema de datos bien estructurado y rígido.

Por otro lado, presenta varias ventajas en comparación con el almacenamiento no estructurado. Se logra una optimización del manejo de memoria, se reducen los requerimientos de almacenamiento, y se pueden hacer actualizaciones a nivel de detalle.

Concluyendo, las mejoras de esta alternativa tiene su contrapartida, el incremento de la sobrecarga durante la inserción y la recuperación de los datos, además de la reducción de la flexibilidad en términos de estructura.

*c) Almacenamiento binario*

Esta nueva forma de almacenamiento provista por la versión 11g se introduce con la intención de compensar ventajas y desventajas de las dos alternativas anteriores. Así, provee una gran flexibilidad en cuanto a la estructura sin deteriorar el desempeño.

Concretamente, mantiene un buen rendimiento en la actualización, indexación y extracción de fragmentos; así como en las consultas puesto que parsea el documento antes de almacenarlo [7][13].

La Figura 1[13] muestra una comparación bastante ilustrativa entre las tres opciones de almacenamiento descriptas en Oracle 11g.

**2. Elección de la estructura de almacenamiento**

	CLOB	OR	Binary XML
Query	poor	excellent	good/excellent
DML	poor	good/excellent	excellent
document retrieval	excellent	good/excellent	excellent
schema flexibility	good	poor	excellent
document fidelity	excellent	poor	good/excellent

Figura 1: Comparación de los diferentes modelos de almacenamiento [13]

Cada modelo de almacenamiento posee un conjunto de características que lo hacen más o menos apropiado según el tipo de datos a manipular. Sin embargo, aunque la naturaleza de los datos es un aspecto fundamental a considerar, el uso que se les dará a esos datos también constituye un factor decisivo [12].

En cuanto a los datos, se pueden distinguir:

*a) Datos Estructurados*

Corresponde a aquellos datos que tienen una estructura regular y granularidad fina. Los datos contenidos en facturas de ventas y en cuentas de un banco son ejemplos de esta categoría.

*b) Datos Semiestructurados o No Estructurados*

Se caracterizan por tener una estructura menos regular, una granularidad más gruesa y con contenidos mezclados; tales como los datos contenidos en correos electrónicos, libros, currículos y advertencias. En general, los datos están organizados para ser usados por personas.

Por otra parte, como se comentó con anterioridad, las aplicaciones que usan los datos también pueden ser de diferente naturaleza:

### c) Aplicaciones centradas en datos

Estas aplicaciones necesitan conocer la estructura de los datos. En general, los datos son altamente estructurados. Luego, la aplicación puede tomar ventaja en este sentido. Es común que los datos se ajustan a un esquema XML.

### d) Aplicaciones centrada en documentos

En este caso, es común que las aplicaciones necesiten mantener una copia exacta de los documentos, como sucede por ejemplo en el ámbito judicial, médico, etc.

Luego, bajo la hipótesis de contar con datos estructurados, por un lado cabe la situación que las aplicaciones no necesiten conocer esa estructura, y por otro, que sí lo demanden. En el primer caso, el modelo de almacenamiento como CLOB será la alternativa más aconsejable, mientras que en el segundo, el modelo apropiado sería el binario.

## 3. Ejemplo

En cuanto a las experiencias realizadas en el uso de datos XML en Oracle, éstas tienen relación con algunas tareas realizadas desde el proyecto marco. Concretamente se tomó como escenario la situación presentada en el Departamento de Informática de la Facultad de CEFN de la UNSJ, a raíz del proceso de acreditación de sus carreras.

El primer objetivo del ensayo realizado fue capturar datos relacionales del Sistema SIU Guarani de la FCFN de la UNSJ, para posteriormente almacenarlos como datos XML, ofreciendo así una de las cualidades de este tipo de estructura, la conveniencia para el intercambio de datos.

En esta sección se expone una porción exigua del trabajo mencionado en los párrafos anteriores, ya que el objetivo, en este caso, es sólo ilustrar algunos aspectos relacionados al almacenamiento de datos XML en Oracle 11g, finalidad de este artículo.

- Se generó una base de datos llamada SIU\_XML, la cual contendrá algunos datos SIU Guarani versión 2.00.0, a Junio de 2004, por supuesto almacenados como datos XML.

- Se creó la tabla EXAMEN\_MATERIA, ilustrada en la Figura 2. Como se puede visualizar en este caso el modo de almacenamiento optado fue el binario [8]. Dicha elección se debió a que los datos a almacenar son estructurados, además que el uso que se les dio posteriormente necesita conocer su estructura, hacer consultas sobre ellos, y tener cierta flexibilidad.

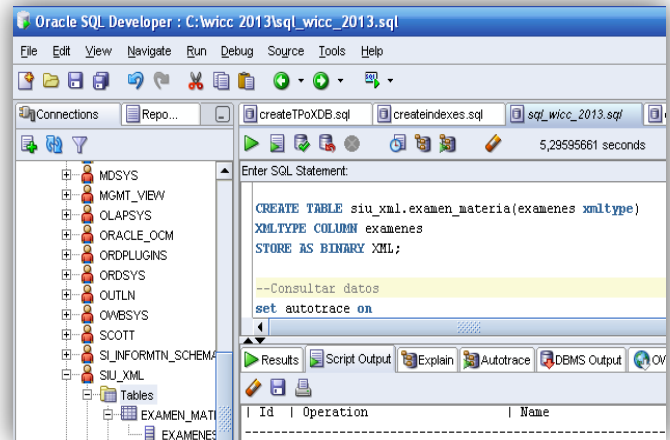


Figura 2: Creación de la tabla EXAMEN\_MATERIA (esquema SIU\_XML)

- Posteriormente se insertaron las tuplas en la tabla EXAMEN\_MATERIA. Los datos agregados corresponden a los Exámenes de la carrera Licenciatura en Ciencias de la Información, período académico 2000-2002. Uno de los documentos XML insertados se muestra en la Figura 3.

```
<?xml version="1.0" encoding="UTF-8" ?>
- <Resultado_Examen_Materia>
  <Materia Nombre_mat="Computación I" />
  <Carrera Nombre_carrera="Licenciatura en Ciencias de la Información" />
  - <Examen Año="2000">
    - <Cantidad>
      <Aprobados>48</Aprobados>
      <Reprobados>59</Reprobados>
    </Cantidad>
  </Examen>
  - <Examen Año="2001">
    - <Cantidad>
      <Aprobados>48</Aprobados>
      <Reprobados>54</Reprobados>
    </Cantidad>
  </Examen>
  - <Examen Año="2002">
    - <Cantidad>
      <Aprobados>46</Aprobados>
      <Reprobados>69</Reprobados>
    </Cantidad>
  </Examen>
</Resultado_Examen_Materia>
```

Figura 3: Ejemplo de Documento XML insertado en la tabla EXAMEN\_MATERIA

Cabe mencionar que se generaron índices específicos<sup>1</sup> que permitieron optimizar la manipulación de los datos XML[12][7][8]. En este caso el tipo de índice creado fue XMLIndex, mostrado en la Figura 4.

<sup>1</sup> Por razones de espacio, en este trabajo no se abordó el tema de índices. Sólo se presenta una breve acotación dada su gran y provechosa utilización.

```
create index idx_examen ON
siu_xml.EXAMEN_MATERIA(EXAMENES) INDEXTYPE IS
XDB.XMLINDEX PARAMETERS
('PATHS (INCLUDE
(/Resultado_Examen_Materia/Nombre_mat));
```

Figura 4: Creación del índice

Específicamente el índice se construyó sobre el elemento Nombre\_mat. Si en la sentencia de creación del índice no se especifica el parámetro PATHS, se crea un índice para cada uno de los elementos del documento XML. Es necesario ser cuidadoso y analizar el uso que se le dará a posteriori a esos datos para decidir sobre qué elementos conviene hacer la indexación, ya que por supuesto tiene asociado un costo, fundamentalmente en cuanto al espacio de almacenamiento.

La Figura 5 muestra una consulta planteada que emplea la función extractValue para recuperar el nombre de las materias.

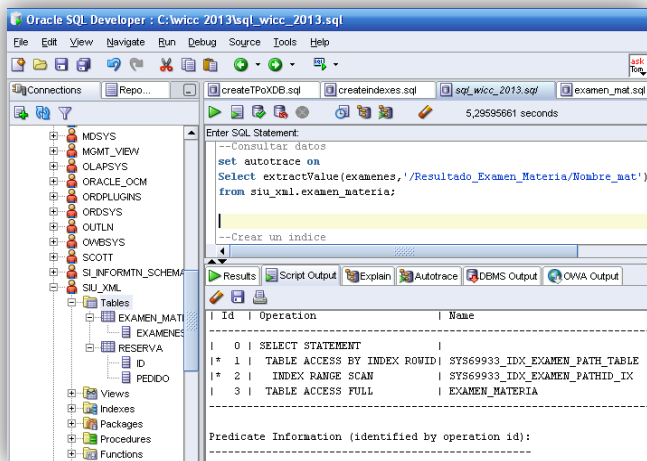


Figura 5: Consulta documento XML

A continuación, en la Figura 6 se puede observar el resultado de dicha consulta.

```
AutoTrace Enabled
EXTRACTVALUE(EXAMENES, '/RESULTADO_EXAMEN_MATERIA/NOMBRE_MAT')
-----
Sistemas de Datos I
Análisis Comparativo de Lenguajes
Modelos y Simulación
Introducción a los Sistemas Digitales
Computación III
Investigación Operativa
Integración Cultural I
Sistemas Administrativos III
Integración Cultural II
Integración Cultural III
Base de Datos con Orientación a Objetos
Optativa I (LIC)
Optativa II (LIC)
Razonamiento Temporal y Programación Lógica con Restricciones
Informática Industrial
Requisito de Idioma
Sistemas de Datos II
Sistemas de Información I
Sistemas de Información II
Informática Legal
Sistemas de Información III
Computación IV
Matemática Básica y Lógica
Introducción a los Sistemas de Información
```

Figura 6: Resultado de la consulta efectuada

Como la consulta de la Figura 5 selecciona los nombres de materias, elemento por el cual se generó el índice, Oracle lo utiliza para acelerar el tiempo de respuesta. La Figura 7 muestra el plan de ejecución que permite observar dicha situación.

Id	Operation	Cost	Card	Bytes	Temp	IO	Time
0	SELECT STATEMENT						
1	TABLE ACCESS BY INDEX ROWID	1	1				00:00:01
2	INDEX RANGE SCAN	1	1				00:00:01
3	TABLE ACCESS FULL	38	426				00:00:01

Figura 7: Plan de Ejecución

## Líneas de Investigación y Desarrollo

La línea de investigación es la tecnología XML en el area de bases de datos, tanto desde el aspecto teórico como también del práctico. En tal sentido se ha trabajado con herramientas que soportan estas tecnologías de maneras muy diferentes:

### Editores XML:

- Altova XMLSpy. Se usó una versión de prueba durante 30 días. Sitio de descarga es <http://www.altova.com>.
- XML Copy Editor: Licencia GPL. Sitio de descarga: <http://xml-copy-editor.sourceforge.net/>.

### Sistemas de gestión de bases de datos que manipulan datos XML:

- El SGBD Oracle 11 g R1 con licencia OTN. Sitio de descarga <http://www.oracle.com.ar>
- El SGBD SQL Server 2012
- El SGBD XML nativo eXist 1.4.0, con licencia GPL. Sitio de descarga <http://www.exist-db.org/download.html>

## Resultados y Objetivos

A continuación se detallan los resultados obtenidos hasta el momento:

- Se ha estudiado y experimentado la mayoría de los temas que estaban planificados. También se ha investigado el motor SQL Server 2012, aspecto no contemplado inicialmente en el proyecto.
- Se dictó el curso: “Base de Datos Relacional-XML” dirigido a empleados de cómputos de



OSSE (Obras Sanitarias Sociedad del Estado, San Juan), de 70 horas presenciales.

- Se dirigió un trabajo final de la carrera Licenciatura en Ciencias de la Información, “Tecnología XML y Bases de Datos: Un ejemplo de aplicación con datos del SIU”. C. Vera y L. Romera.
- Se está asesorando un trabajo final de la carrera LSI, “XML y Bases de datos: Un caso de aplicación”. M. Avendaño.

URL://www.w3.org/TR/2001/REC-xmlschema-1-20010502.

11. w3c. “XSL Transformations (XSLT) “ URL: <http://www.w3.org/TR/xslt.html>
12. Williams. “Performance of Relational Databases versus Native XML Databases”
13. Zhang, Agarwal. “Binary XML Storage and Query Processing in Oracle 11g”.

## Formación de Recursos Humanos

Se propusieron y aceptaron dos Becas de Investigación del CICITCA:

- Beca Interna de Investigación en la CATEGORÍA INICIACIÓN, convocatoria 2012, en el marco de la Ordenanza Nro. 10/05-CS: Licenciada Cristina Vera.
- Beca Interna de Investigación en la CATEGORÍA ALUMNO AVANZADO, convocatoria 2012, en el marco de la Ordenanza Nro. 10/05-CS: Mauro Avendaño, alumno de la carrera LCI.

## Referencias

1. Benz, Durant. “XML Programming Bible”. Wiley Publishing Inc.
2. Chaudhri, Rashid, Zicari. “XML Data Management: Native XML and XML-Enabled Database Systems”. Addison Wesley.
3. Evjenet. “Professional XML”. Wrox Press 2007.
4. Elmasri, Navathe. FUNDAMENTALS OF DATABASE SYSTEMS. Cuarta Edición. Addison Wesley.
5. ISO/IEC 9075-14:2003 Information technology – Database languages – SQL – Part 14: XML-Related Specifications, (SQL/XML)
6. Molina, Ullman, Widom. DATABASES SYSTEMS: THE COMPLETE BOOK. Prentice Hall.
7. Oracle. “Oracle® XML DB Developer's Guide 11g”.
8. Oracle. “SQL Reference 11g”.
9. Silberschatz, Korth, Sudarshan. FUNDAMENTOS DE BASES DE DATOS, Quinta Edición, Mc Graw Hill.
10. w3c. “XML Schema Part 1: Structures W3C Recommendation 2 May 2001”. Fecha de consulta: 10 de Marzo de 2010.

# VISIÓN AMPLIADA DEL MANAGEMENT (EPM)

**Autores:** Doc. Ing. Rodolfo Bollatti (Director Investigación)

Universidad Abierta Interamericana – Buenos Aires  
Tecnología Informática  
Secretaría de Investigación – CAETI

**E-mail:** Rodolfo.Bollatti@UAI.edu.ar

**Telefono:** 011-1536584700

## Resumen

Estudio metodológico y sistémico para la optimización de la planificación estratégica y gestión en organizaciones latinoamericanas a través de la tecnología de información.

Nuestras organizaciones, intentan adoptar las mejores prácticas de Enterprise Performance Management (EPM), con metodologías de implementación, modelos y sistemas de información gerenciales con bases culturales muy diferentes a nuestro contexto, este proceso conlleva como consecuencia implementaciones que fracasan o esfuerzos extraordinarios que sobrepasan ampliamente los beneficios que estas implementaciones proveen a las organizaciones.

El objetivo general de este proyecto es revelar la tendencia actual y futura sobre Enterprise Performance Management, desarrollando metodologías exitosas de implementación a través del análisis de las diferentes vertientes y

disciplinas, tecnologías innovadoras (EPM/BI en nube, análisis de redes sociales ARS, otros) adaptables a las organizaciones y contexto sociocultural latinoamericano.

**Palabras clave:** Enterprise Performance Management, Planeamiento Estratégico, EPM, Business Performance Management, BPM, Corporate Performance Management, CPM, Business Intelligence.

## Contexto

El proyecto se base en dos líneas rectoras: sistemas de información y planificación estratégica. Desarrollado desde la facultad de Tecnología de Información de la Universidad Abierta Interamericana, y articulando objetivos específicos junto a la facultad de ciencias empresariales de UAI. Secretaría de Investigación de UAI (CAETI) contribuye en su financiamiento.

## Introducción

Enterprise Performance Management (EPM) es una serie de procesos y aplicaciones de softwares diseñadas para optimizar la implementación, evaluación y formulación de la estrategia empresarial. Las organizaciones están descubriendo que EPM es más que sólo la planificación, presupuestación, previsión, control de gestión, consolidación, scorecarding o cuadros de mando, es un marco conceptual que unifica y estandariza todos los procesos y sistemas de información gerenciales y de planificación estratégica.

Hay prejuicios y confusión en la definición del término EPM, en gran parte, cuando los ejecutivos al oír por primera vez el término, dicen: "De nuevo otra sigla", para algunos, significa Business Intelligence (BI), y para otros es otro viejo truco de la comunidad de proveedores para vender las tecnologías existentes. EPM es real y es distinto del BI, la principal diferencia es que BI es la tecnología innovadora, mientras que EPM son procesos de negocio que aprovecha la inteligencia de negocios.

Muchas fuentes confunden hoy en día qué es y no es EPM, gran parte de la confusión es que EPM es una nueva categoría para describir las múltiples aplicaciones que ya han sido implementadas, como planificación, presupuestos e informes de KPI.

EPM trasciende la tecnología y se centra en la esencia de los conceptos de planificación estratégica.

Los procesos y las soluciones EPM permiten una excelente gestión, por lo que es más fácil para los ejecutivos de todos los niveles identificar, comunicar y supervisar a los lineamientos claves y esenciales de los valores empresariales de la organización. Una vez que una organización entiende sus factores claves, la forma de optimizar y comunicarlos, entonces se puede ejecutar la estrategia en la dirección correcta.

A veces es más fácil, para clarificar los conceptos la mención de lo que no son. EPM no es:

- Una solución de tecnología o software.
- Business Intelligence
- El presupuesto o la planificación
- Consolidación financiera o la presentación de informes
- Cuadros de mando
- Previsión y modelado de escenarios
- Indicadores clave de rendimiento

EPM cierra la brecha entre estrategia y ejecución, fundamentalmente es la aplicación sistémica de las buenas prácticas de gestión empresarial reforzada por la información oportuna y precisa para comunicarse de manera efectiva, comprender y controlar el desempeño de una organización.

Las soluciones EPM permiten:

- Mejora la comunicación estratégica: proporciona un mecanismo eficaz para comunicar la estrategia y las expectativas de los directivos y el personal de todos los niveles de la organización a través de modelos de planificación y medición de desempeño ligados a los objetivos corporativos.
- Mejora la colaboración: fomenta un intercambio bidireccional de ideas e información, tanto vertical entre los niveles dentro de una organización y horizontalmente entre los departamentos. Permitiendo la generación de estrategias emergentes.
- Mejora el control: permite a los empleados ajustar continuamente el plan y corregir o mejorar las operaciones de una manera oportuna, proporcionándoles información actualizada al día sobre las condiciones del mercado y el estado de las áreas operativas.
- Mejora la coordinación: facilita la coordinación entre las unidades de negocio y funcional de grupos que de otra forma actúan de manera independiente.
- Mejora el conocimiento del planeamiento estratégico: Permiten los registros de los elementos críticos para la dedición estratégica.

Los procesos EPM se enmarcan en 4 pasos esenciales, que cierran el círculo virtuoso del management:

- Estrategia
- Plan
- Medición de desempeño
- Feedback, ajustes y evolución

Las empresas en EU, Canadá y países nórdicos como Dinamarca, son los primeros en adoptar los sistemas EPM para administrar, comunicar, desarrollar y cuantificar sus estrategias, presupuestos y métodos de planeación de recursos para las futuras capacidades. Es muy importante que los países en vía de desarrollo adopten también estas prácticas, en el caso de sus empresas interesa mejorar su rentabilidad o, en el caso de sus gobiernos o entidades sin fines de lucro, optimizar los servicios que presta a sus ciudadanos, comunidades, o entidades reguladoras. Todos los que participan en el mercado se benefician, incluyendo a los empleados, los clientes y los usuarios, no sólo los dueños de las empresas y los inversionistas, asimismo, a una mayor recuperación de la inversión, una mayor generación de riqueza nacional.

Las organizaciones latinoamericanas, intentan adoptar las mejores prácticas de EPM, originadas por metodologías de implementación, modelos y sistemas de información gerenciales con bases culturales muy diferentes a nuestro contexto sociocultural, este proceso conlleva como consecuencia implementaciones que fracasan o esfuerzos extraordinarios que sobrepasan ampliamente los beneficios que estas soluciones proveen a las organizaciones, muchos de estos inconvenientes son originados por:

- Modelos de pensamientos y planificación estratégicos inadecuados para el contexto sociocultural latinoamericano
- Metodologías de implementación EPM inadecuadas para las organizaciones Latinoamérica
  - Deficiente cultura estratégica en directores y gerentes.
  - Escasa información sobre conceptos y mejores prácticas EPM.
  - Insuficiente experiencia en implementaciones de sistemas de información gerenciales.
  - Tecnologías inadecuadas.

Los conceptos e implementaciones sistémicas de EPM deben cambiar los fundamentos ampliamente difundidos por académicos americanos o europeos, para adaptar sus conceptos y principios a la realidad sociocultural latinoamericana, permitiendo implementaciones exitosas en nuestras organizaciones. Las opciones estratégicas convencionales deben ampliarse para poder explicar las nuevas fuentes de rentabilidad observadas en la economía de hoy, manteniendo las nuevas variables impulsoras de la sociedad (arte, cultural, sustentabilidad, etc). De esta forma, las soluciones que permitan combinar los pilares existentes del EPM con las nuevas bases que impulsan las economías regionales, centrando la atención en los consumidores y potenciando nuestras características más virtuosas como sociedad.

Nuestro equipo de proyecto, cuenta con más de 15 años de experiencia y formación académica en dichas soluciones dentro de diferentes empresas en todo el mundo, permitiendo la búsqueda de mejores prácticas de dichas soluciones, con el objetivo de lograr una visión más amplia y profunda del EPM en toda Latinoamérica, con el beneficio directo tanto para nuestras organizaciones como así también para nuestra sociedad.

## Resultados y Objetivos

En la actualidad el proyecto se encuentra en su primera fase de maduración y análisis detallado.

A continuación se describen los correspondientes objetivos específicos a desarrollar:

- Clarificación de conceptos de EPM/BI
- Relación entre Planificación estratégica clásica y EPM
- Ventajas de los diferentes enfoques y visiones a nivel mundial sobre EPM, integración del arte en el pensamiento estratégico.
- Revelar el GAP existente entre la teoría sobre Planificación Estratégica y la



implementación de soluciones EPM en organizaciones latinoamericanas.

e) Problemáticas y dificultades de implementar soluciones EPM sobre Organizaciones Latinoamericanas

f) Análisis de tecnologías facilitadoras a nuestro contexto sociocultural (EPM/BI en la nube, Análisis de redes sociales ARS, otros)

g) Desarrollo de una metodología y soluciones propias para adaptar las implementaciones de sistemas de información de EPM a nuestro contexto sociocultural y organizaciones, centrados en una visión integral valorando los consumidores y las necesidades de nuestra sociedad.

h) Estudio de campo sobre dichas metodologías y prácticas sistémicas. Conclusiones y optimizaciones.

## Formación de Recursos Humanos

A continuación se describen los productos asociados a dicho proyecto:

Producto	Nº
Libros	1
Artículos en Revistas Internacionales con referato	1
Artículos en Revistas nacionales con referato	2
Presentaciones en Congresos	3
Tesis aprobadas	2
Recursos Humanos formados	5
Patentes, convenios, desarrollos transferibles	1

## Referencias

### Síntesis Bibliográfica:

- Eckerson, W. (2004). Best Practices in Business Performance Management: Business and Technical Strategies. 101communications LLC, 2.

- Kaplan, R.S. and Norton, D.P. (1992). The Balanced Scorecard — Measures that Drive Performance. Harvard Business Review

- Information Strategy: A Philosophical Blueprint for Building the - Executive Dashboard, J. Griffin, DM Review, Aug. 2002

- CPM: A Strategic Deployment of BI Applications, Gartner, May 2002

- Strategy Safari, 1998a. Mintzberg con Bruce Ahlstrand & Joe Lampel

- Managing, Mintzberg 2009.

- Miranda, S. (2004). Beyond BI: Benefiting from Corporate Performance Management Solutions. Financial Executive, 20(2) 58–61.

- Griffin, J. (2004). Information Strategy: Overcoming Political Challenges in Corporate Performance Management.

- Power, D.J. (2003). A Brief History of Decision Support Systems.

- Power, D.J. (2003). A Brief History of Decision Support Systems.

- Lingle, J. H., and Schiemann, W. A. (1996). From Balanced Scorecard to Strategic Gauges: Is Measurement Worth It? Management Review,

- David, Fred R.; Fincowsky, Enrique Benjamín Franklin. Conceptos de administración estratégica. 11a.ed.-- México, DF : Pearson Educación de México, 2008

- Hax, Arnoldo C.; Majluf, Nicolás S. Estrategia para el liderazgo competitivo : de la visión a los resultados.-- Barcelona : Granica, c1997

- Thompson, Arthur A.; Strickland, A. J.; Gamble, John E. y otros. Administración estratégica : teoría y casos. 15a. ed.-- México, D.F. : McGraw-Hill Interamericana de México, 2008

- Steiner, George A. Planeación estratégica : lo que todo director debe saber, una guía paso-a-paso.-- México, DF : Compañía Editorial Continental, c2000

# Generador Automático de Modelos de Datos Normalizados en Bases de Datos Relacionales

María A. Paz Menvielle, Juan C. Cuevas, Luis E. Damiano, Roberto M. Muñoz, Sergio Quinteros

Departamento de Ingeniería en Sistemas de Información

Facultad Regional Córdoba – Universidad Tecnológica Nacional

Maestro Marcelo López esq. Cruz Roja Argentina – Ciudad Universitaria - Córdoba

0351 - 4686385

pazmalejandra@gmail.com, juancarloscue@gmail.com, luis.damiano@gmail.com, rmunoz@sistemas.frc.utn.edu.ar, ser.quinteros@gmail.com

## Resumen

El propósito de este proyecto de investigación es diseñar y construir un aplicativo que sea capaz de resolver situaciones reales en la obtención de diseños de modelos de datos en bases de datos relacionales, basado en el concepto informático de sistema experto y partiendo de un conjunto de datos desnormalizado.

El resultado de esta propuesta es lograr:

- Una herramienta a disposición del plantel docente y estudiantes de Gestión de Datos, asignatura de la Carrera de Ingeniería en Sistemas de Información, Universidad Tecnológica Nacional, Facultad Regional Córdoba.

- Un aplicativo para la validación de conocimientos adquiridos por los estudiantes.

- Antecedentes para nuevos estudios relacionados y que amplíen lo logrado con ingeniería del conocimiento.

- La aplicación de conceptos y procedimientos de la inteligencia artificial en el proceso de normalización de una base de datos relacional.

- Una base de conocimiento en las dos especialidades: normalización de modelos de datos e ingeniería del conocimiento, que serán centrales en este proyecto y útiles como recursos de posteriores incursiones en investigaciones futuras.

Para todo esto es necesario desarrollar estudios y aprendizajes en el área de Ingeniería del

Conocimiento y del lenguaje CLIPS en particular, como lenguaje de programación a utilizar.

**Palabras clave:** normalización, modelo relacional, ingeniería del conocimiento, sistemas expertos, estructuras de datos, generador.

## Contexto

En la asignatura curricular Gestión de Datos, del tercer nivel de la Carrera de Ingeniería en Sistemas de Información, se cumple con los contenidos mínimos fijados en la Ordenanza 1150/2007, donde la Universidad Tecnológica Nacional aprueba el Diseño Curricular de la Carrera y fija como contenido mínimo la técnica de “Normalización”, e implementa los descriptores de la Resolución Ministerial 786/2009, del Ministerio de Educación de la Nación, en la que explícitamente un descriptor es “Diseño, Administración y Gestión de Bases de Datos”, en la sub área Bases de Datos del área Tecnologías Aplicadas.

En la Ordenanza 1150/2007 se mencionan, para la asignatura, cinco objetivos generales, donde dos de ellos específicamente puntualizan:

- Desarrollar los conceptos de estructuración de los datos en dispositivos de almacenamiento

- Describir metodologías para el modelado de datos

El estudio de los modelos típicos y de sus variantes, cómo se comportan y el estudio de las

estructuras internas a los modelos y su caracterización, permitirá sistematizar este conocimiento.

El proyecto de investigación y desarrollo ha sido homologado por la Secretaría de Investigación, Desarrollo y Posgrado de la Universidad Tecnológica Nacional, reconocido con el código UTN1702 en el ámbito de la Universidad, por un período de dos años y a partir del 1° de enero de 2013. El proyecto se suma a los ya existentes en el Grupo de Investigación en Inteligencia Artificial, por su relación con la temática y porque está en el contexto de la línea de investigación del grupo.

Otros antecedentes, no menos importantes, son los proyectos de investigación desarrollados en el ámbito de la Facultad Regional Córdoba:

a) Administrador de Base de Datos Relacional TecnoDB. Finalizado en diciembre de 2007.

b) Análisis y aplicación de metodologías para la generación de consultas complejas utilizando esquemas OLAP. Finalizado en diciembre de 2010.

c) PROMETEO-Desarrollo de un método y una herramienta para el aprovechamiento de Metadatos de Base de Datos Relacionales. Finalizado en diciembre de 2010.

## Introducción

En este proyecto se pretende generar un aplicativo que, a partir de un conjunto de datos provistos por un usuario, a través de reglas y hechos expresados en lógica de predicados de primer orden, estructure un modelo de datos en el entorno del Modelo Relacional.

El alcance involucra la obtención de tablas normalizadas, hasta el cumplimiento de la tercera forma normal, y el señalamiento de las relaciones.

Este proyecto podría incluso ser ampliado, para alcanzar la implementación de mayor número de restricciones y reglas disponibles para situaciones especiales provenientes del mundo real.

La propuesta incluye el diseño y la construcción del aplicativo con un lenguaje de programación de sistemas expertos, siendo necesario desarrollar estudios y aprendizajes en el

área de Ingeniería del Conocimiento y del lenguaje CLIPS en particular.

En la implementación de la asignatura de referencia, Gestión de Datos, de la Carrera de Ingeniería en Sistemas de Información de la Universidad Tecnológica Nacional-Facultad Regional Córdoba, se desarrollan sólo temas relacionados a Base de Datos.

Los docentes de esta cátedra desarrollan funciones de administración, diseño y programación en bases de datos, aportando experiencia profesional en la solución de problemas. Uno de los temas que han generado debates en el seno de la cátedra, es el cómo abordar en forma ágil y clara el modelado de una base de datos. La discrepancia de cómo llegar al resultado, el cómo hacerlo, de qué forma, qué pasos realizar, los aspectos procedimentales o funcionales de las mecánicas que llevan al producto terminado, hace dificultoso establecer una metodología para la enseñanza. Se aclara que no está en dudas por ningún docente de la cátedra, que para lograr un modelado de una base de datos hasta tercera forma normal, se debe validar que el conjunto de las relaciones cumplan con satisfacer las reglas que establecen las formas normales, claramente definidas en [1], [2] y [3].

El grupo de investigación se planteó la posibilidad de concretar una investigación que brinde sostén y fundamento a una solución sistemática, viable para el tema de modelar una base de datos relacional.

La motivación de los docentes, por generar nuevos materiales y estrategias para la construcción de un espacio de investigación, permite que integrantes de la cátedra sean coautores en el libro Bases de Datos [4].

En el Grupo de Investigación en Bases de Datos (GIBD), del Departamento de Ingeniería en Sistemas de Información, de la Facultad Regional Concepción de Uruguay, generaron el proyecto 25-D040, referido a Métodos de acceso, consultas y aplicaciones en modelos de bases de datos no convencionales [5].

En cuanto a antecedentes de una herramienta que genere un modelo de datos relacional en forma automática, no se ha encontrado algo

disponible en el ámbito local académico ni en el empresarial. Sólo un producto comercial, generado por la empresa Axesor [6], de Granada (España).

Respecto de los conocimientos sobre la rama de Ingeniería del Conocimiento y Sistemas Expertos como productos, se hizo una investigación preliminar con acceso a bibliografía [7] [8].

El Sr. Sergio Antonio Becerra Zepeda realizó su Tesis para el Posgrado de Maestro en Ciencias Área Computación, Universidad de Colima (México), con el tema Bases de Datos Inteligentes, a utilizar como conocimientos válidos desde el ambiente académico [9].

Los Sistemas Expertos son una rama de la Inteligencia Artificial, que hace un amplio uso del conocimiento especializado para resolver problemas como un especialista humano. Enrique Castillo, José M. Gutiérrez y Ali Hadi definen: “Un sistema experto puede definirse como un sistema informático (hardware y software) que simula a los expertos humanos en un área de especialización dada” [10]. En su dominio de conocimiento, el sistema experto hace inferencias de la misma forma en que un especialista humano inferiría la solución de un problema. Al proceso de construir un sistema experto se lo llama ingeniería del conocimiento, y consiste en la adquisición de conocimiento a partir de un especialista humano o de otra fuente y su codificación en el sistema experto (García Martínez, R. - Britos, P.; 2004) [11]. Los sistemas expertos suelen estar diseñados de manera distinta a los programas convencionales porque los problemas no tienen generalmente una solución algorítmica y dependen de inferencias para obtener una solución razonable, considerando ésta como la mejor que se puede esperar si no hay un algoritmo disponible que ayude a obtener la solución (Giarratano, J. - Riley, G.; 2005) [12].

El lenguaje elegido inicialmente para el desarrollo del proyecto es CLIPS (C Language Integrated Production System), ya que permite la programación con paradigmas múltiples, proporcionando soporte para programación basada en reglas, orientado a objeto y por procedimientos, enfocándose principalmente en

los aspectos de la programación basada en reglas. Es una herramienta para el desarrollo de sistemas expertos, creada por la Software Technology Branch (STB), NASA/Lyndon B. Johnson Space Center (Alanis Barrera, Ricardo; 2005) [13].

Para la planificación, recopilación, generación de instrumentos y el desarrollo de la investigación se trabajará, entre otros, con el material bibliográfico “Metodología de la Investigación”, de César Augusto Bernal [14], y “Metodología de la Investigación”, de Hernández Sampieri, R. y otros [15].

## Líneas de investigación y desarrollo

En el desarrollo de este proyecto se utilizará el método científico, en cuanto a la forma investigativa y el tratamiento de los datos, pero aplicando administración de proyecto con el método espiralado de seguimiento para la planificación del desarrollo, que estará basado en cuatro etapas a ser repetidas en forma cíclica hasta la culminación del proyecto: planeamiento, adquisición del conocimiento, codificación y evaluación.

La investigación abarca dos áreas de la informática, Base de Datos e Ingeniería del Conocimiento, y tomando en cuenta que hay inconmensurable conocimiento disponible en ambos campos, el tipo de investigación a aplicar es transversal, enfocada en conocimientos contemporáneos sobre ambos temas.

Este proyecto se enmarca en los recientes conceptos y métodos de Investigación, Desarrollo e Innovación (I+D+i), ligados a la mirada de desarrollo tecnológico y el ingreso del producto al mercado y su uso, basado en la secuencia: Síntesis y teoría; Explorar, hipotetizar y clarificar; Diseño, desarrollo y prueba; Implementación, estudio de eficacia y mejora; Aumento progresivo y estudio de eficacia.

Las acciones que se contemplan en este proyecto de investigación se describen en los próximos párrafos.



En la primera etapa de trabajo, se está haciendo la recopilación y detección de la existencia de diversas formas y criterios para obtener un modelo de datos normalizado. Según esto el dominio del problema es el área de ingeniería, específicamente en diseño de bases de datos relacionales. Con información recolectada y sistematizada, se hará un primer análisis de variables y se podrá comenzar con la selección de paradigma. Si bien el equipo entiende que el paradigma apropiado es el de programación en formato de Sistemas Expertos, el análisis de las variables, confirmará en su momento esta posición, o permitirá vislumbrar un paradigma diferente.

Determinado el paradigma de trabajo se realizará la selección de herramientas apropiadas para comenzar el desarrollo del producto, produciendo la adquisición del conocimiento y elegir el lenguaje de programación para un sistema experto. Como ya se indicó, se reconoce en CLIPS un lenguaje que puede brindar el potencial para el desarrollo.

Para adquirir conocimiento se utilizarán instrumentos de recolección de información y de medición de variables, dirigida a entrevistas con expertos en Bases de Datos, docentes de Gestión de Datos y diseñadores externos.

Teniendo el material recolectado, los análisis realizados y cerrando la etapa de adquisición de conocimiento, el equipo pasará al desarrollo del prototipo inicial. Se deberán definir las características de diseño y programación, la reingeniería del paradigma, si fuera necesario, la retroalimentación de ideas y conocimientos, perfilando el modelo de diseño y el producto.

Está previsto que durante el segundo año se comience el desarrollo del aplicativo, en donde se encaminará el diseño a una solución definitiva, pensando en el desarrollo de la herramienta de inferencia. Paralelamente se generará el plan de pruebas, sometiendo el producto a las pruebas para observar su comportamiento, produciendo una retroalimentación permanente. El mantenimiento y las mejoras de los detalles detectados, con la intención de que el producto logre un buen

grado de inserción y aceptación, generando innovación.

Un próximo paso será la obtención de un sistema mejorado y estable, que sea capaz de inferir o solucionar un conjunto de modelos y estructuras, incluyendo características deseables en una interfaz gráfica de usuario. Posteriormente está planificada la sociabilización del proyecto de investigación, a nivel académico y empresarial.

## Resultados y Objetivos

El objetivo principal planteado es construir una herramienta capaz de resolver situaciones reales, obteniendo diseños de modelos en bases de datos relacionales, basada en el concepto de Sistema Experto y partiendo de un conjunto de datos desnormalizado.

El grupo se plantea los siguientes objetivos operacionales:

- \*Permitir a los usuarios el cotejo y validación de los resultados obtenidos con la propuesta de normalización de la base de datos que brinde el producto, en relación a los datos suministrados.

- \*Profundizar en el estudio y aprendizaje sobre Ingeniería del Conocimiento, con el fin de tener una visión acabada sobre el diseño del aplicativo.

- \*Lograr que la herramienta emita un diagnóstico respecto de situaciones dudosas, realice sugerencia en relación a su base de conocimiento y efectúe preguntas adicionales sobre información de los datos suministrados.

- \*Lograr incrementar el conocimiento avanzado sobre el modelado de bases de datos relacionales

- \*Crear un método estandarizado para el abordaje y diseño de la normalización de bases de datos relacionales.

- \*Brindar una herramienta didáctica al profesorado de la Cátedra de Gestión de Datos.

- \*Introducir otra tecnología entre los recursos didácticos de que se dispone, pretendiendo que el aplicativo se encuadre como un producto de Tecnología de la Información y la Comunicación.

- \*Difundir el conocimiento obtenido en el ámbito académico y público en general.

- \*Transferir el conocimiento logrado en la temática de Base de Datos, Ingeniería del

Conocimiento, Diseño de Sistema Expertos, al ámbito académico y de investigación.

\* Obtener un producto simple de utilizar y didáctico.

## Formación de Recursos Humanos

El conocimiento obtenido y las conclusiones que se logren, en relación a temas de bases de datos y en sistemas expertos, aportarán seguramente al enriquecimiento en la formación de los integrantes del equipo de investigación. En dicho equipo trabajan una estudiante del último año de la Carrera de Ingeniería en Sistemas de Información, una ingeniera con título en trámite y tres estudiantes del cuarto nivel de la carrera, iniciando su formación en investigación científica y tecnológica. Los estudiantes podrán hacer la Práctica Supervisada de quinto año y acceder a Becas de Investigación de la Universidad.

Está planteado el desarrollo de la Tesis de la Maestría en Ingeniería en Sistemas de Información de un integrante docente, cursada en la Universidad Tecnológica Nacional.

## Referencias

- [1] Elmasri, R. y Navathe, S. - Fundamentos de Sistemas de Bases de Datos-5ta Edición – Edit. Pearson-Impreso en España. 2007. ISBN: 978-84-7829-085-7.
- [2] Date, Christopher - Introducción a los Sistemas de Bases de Datos - Volumen 1 - Quinta Edición - Impreso en Estados Unidos. Edit. Addison Wesley Iberoamericana. 1993 – ISBN: 0-201-51859-7.
- [3] Silberschatz y otros- Fundamentos de Bases de Datos - Quinta Edición- Impreso en Estados Unidos. Edit. Mc Graw Hill- 2006. ISBN: 84-481-4644-1.
- [4] Reinoso, E.; Maldonado, C.; Muñoz, R.; Damiano, L.; Abrutsky, M. - Bases de Datos - Edit. AlfaOmega Editores- Argentina, 2012- ISBN: 978-987-1609-31-4.
- [5] Grupo de Investigación en Bases de Datos (GIBD) - Departamento Ingeniería en Sistemas de Información (DISI), Facultad Regional Concepción del Uruguay. Métodos de acceso, consultas y aplicaciones en modelos de bases de datos no convencionales.

Proyecto culminado: 25-D028 - Nuevas tendencias y aplicaciones en bases de datos.

<http://sistemas.frcu.utn.edu.ar/isi/index.php/gibd> - Accedido en mayo de 2012.

[6] Axesor - Sistemas Expertos en Normalización y Codificación de Bases de Datos- Producto Marketing Intelligence -

<http://marketing-intelligence.axesor.es/tratamiento-cualificacion/normalizacion> . Accedido en mayo 2012.

[7] Enrique Castillo, Ángel Cobo, José Manuel Gutiérrez y Rosa Eva Pruneda. Introducción a las Redes Funcionales con Aplicaciones – Un Nuevo Paradigma Neuronal. Editorial Paraninfo. Madrid, España. 1999. ISBN 84-283-2525-1.

[8] Ponce Cruz, Pedro - Inteligencia Artificial con Aplicaciones a la Ingeniería - Editorial ALFAOMEGA Grupo Editor, 2010 - ISBN 9786077854838.

[9] Becerra Zepeda, Sergio Antonio- Bases de Datos Inteligentes- Tesis para Maestro en Ciencias Área Computación - Universidad de Colima, México -Junio 1999.

[http://digeset.uco.mx/tesis\\_posgrado/Pdf/Sergio%20Antonio%20Becerra%20Zepeda.pdf](http://digeset.uco.mx/tesis_posgrado/Pdf/Sergio%20Antonio%20Becerra%20Zepeda.pdf). Accedido en febrero 2012.

[10] Castillo, Enrique; Gutiérrez, José Manuel y Hadi, Ali S. Sistemas Expertos y Modelos de Redes Probabilísticas. Editorial Academia de Ingeniería. Madrid, España. 1998. ISBN 84-600-9395-6.

[11] García Martínez, Ramón - Britos, Paola Verónica - Ingeniería de Sistemas Expertos - Editorial Nueva Librería, 2004 - ISBN 9789871104154.

[12] Giarratano, Joseph - Riley, Gary - Sistemas Expertos Principios y Programación - Editorial Cengage Learning / Thomson Internacional, 2005 - ISBN 9789706860590

[13] Alanis Barrera, Ricardo- Sistemas Expertos e Inteligencia Artificial CLIPS - Burgos CyL., Febrero de 2005 - Universidad de Burgos: <http://es.scribd.com/doc/6732509/Clips>. Accedido en mayo de 2012.

[14] César Augusto Bernal - Metodología de la Investigación - Pearson Educación de México, 2006 - ISBN 970-26-0645-4.

[15] Hernández Sampieri, Roberto; Fernández Collado, Carlos; Baptista Lucio, Pilar - Metodología de la Investigación - MacGraw-Hill Interamericana Editores - México, 2006 - ISBN: 9701057538.

# Bases de Datos NoSql en Cloud Computing

Adriana Martín<sup>1</sup>, Susana Chávez<sup>2</sup>, Nelson Rodríguez<sup>3</sup>, Adriana Valenzuela<sup>4</sup>, María Murazzo<sup>5</sup>

Departamento e Instituto de Informática - F.C.E.F. y N. - U.N.S.J.

Complejo Islas Malvinas. Cereceto y Meglioli. 5400. Rivadavia. San Juan

<sup>1</sup>arianamartinsj@gmail.com <sup>2</sup>schavez@iinfo.unsj.edu.ar <sup>3</sup>nelson@iinfo.unsj.edu.ar

<sup>4</sup>franciscaadriana.valenzuela@gmail.com <sup>5</sup>marite@unsj-cuim.edu.ar

## Resumen

Las bases de datos NoSQL son sistemas de almacenamiento de información que **no cumplen con el esquema entidad-relación**. Mientras que las tradicionales bases de datos relacionales basan su funcionamiento en tablas, joins y transacciones. Las bases de datos NoSQL **no imponen una estructura de datos en forma de tablas y relaciones entre ellas sino que proveen un esquema mucho más flexible**.

Las bases NoSQL son adecuadas para una escalabilidad realmente enorme, y tienden a utilizar modelos de consistencia relajados, no garantizando la consistencia de los datos, con el fin de lograr una mayor performance y disponibilidad. A esto se agrega el inconveniente de que no tienen un lenguaje de consulta declarativo, por lo que requiere de mayor programación para la manipulación de los datos.

En general se pueden mencionar Sistemas NoSQL clasificados en cuatro categorías:

- Framework **Map-Reduce** (usado por aplicaciones que hacen procesamiento analítico online - OLAP), Por ejemplo Hadoop.
- Almacenamiento **Clave-Valor** (sistemas que tienden al procesamiento de transacciones online - OLTP), Por ejemplo: Google

BigTable, Amazon Dynamo, Cassandra, Voldemort, HBase.

- Almacenamiento de **Documentos**  
Por ejemplo: CouchDB, MongoDDB, SimpleDB
- Sistemas de base de datos **Gráficas**. Por ejemplo: Neo4j, FlockDB, Pregel.

Con respecto al almacenamiento en **Columnas** que en general son tratados como Sistemas NoSQL, no son más que una forma de organización de un sistema de base de datos relacional. Sin embargo por la alta performance para cierto tipo de aplicaciones son considerados como del tipo almacenamiento Clave-Valor.

En resumen, para **manipular enormes cantidades de información de manera muy rápida** los Sistemas NoSQL trabajan mejor que los sistemas de base de datos tradicionales, sin embargo para muchísimas aplicaciones la solución está en las bases de datos tradicionales.

**Palabras clave:** NoSql, Map- Reduce, Cloud Computing.

## Contexto

El presente trabajo se encuadra dentro del área de Bases de Datos y Minería de Datos, y se enmarca dentro del proyecto de investigación Implantación de un ambiente de Cloud Computing para integración de recursos, el cual tiene como unidades ejecutoras al Departamento e Instituto de Informática

de la FCEfyN de la UNSJ. Los trabajos iniciados en el citado proyecto tienden al desarrollo de aplicaciones sobre Cloud Computing, y almacenamiento de grandes cantidades de información y rápido acceso sobre el mismo

## Introducción

Las bases de datos NoSQL surgen como una solución a los constantes requerimientos de procesamiento y análisis a gran escala de enormes cantidades de datos, y para los cuales los sistemas tradicionales de base de datos son insuficientes.

El termino **NoSQL** ha evolucionado no para significar *no lenguaje sql*, sino para referirse a sistemas que *no son DMBS* (database management system) *tradicionales*.

El framework MapReduce es una gran alternativa a los sistemas tradicionales. En base a este framework Google ha desarrollado **Hadoop**, un sistema open source usado ampliamente. Se caracteriza por:

- No existe un modelo de datos, los datos se almacenan en archivos HDFS (Hadoop Distributed File System).
- Los usuarios proveen las funciones específicas los datos usando el Framework MapReduce. Las funciones disponibles son: map(), reduce(), reader(), writer() y combine().
- El sistema provee procesamiento de datos "glue" escalable y con tolerancia a fallos (Glue procesa los datos a través de las funciones garantizando la tolerancia a fallos).

El almacenamiento Clave-Valor, está diseñado para el procesamiento de transacciones online (OLTP). Este tipo de aplicaciones son operaciones pequeñas sobre algunos datos de una base de datos masiva. Por esto, son mas simples

Para ello suelen almacenar toda la información que puedan en memoria y

están preparadas para escalar horizontalmente sin perder rendimiento.

Si pretendemos desarrollar una aplicación que requiera la lectura/escritura de cantidades de datos y pueda dar servicio a millones de usuarios sin perder rendimiento, entonces debemos plantearnos el uso de una base de datos NoSQL. Las grandes redes sociales como [facebook](#) y [twitter](#) o el propio [Google](#) las utilizan como medio fundamental de almacenamiento de información.

Se puede utilizar una base de datos NoSQL para almacenar toda la información de una aplicación para aquellas funcionalidades que requieren millones de consultas en tiempo real.

Existe una gran variedad de bases de datos basadas en documentos NoSQL para clasificar la información en un formato estructurado, mientras afrontamos la estructura flexible de los puntos de datos individuales. Muchos entornos NoSQL también brindan apoyo para consultas Map-Reduce exhaustivas y para el procesamiento, lo que las hace ideales para el procesamiento de una gran cantidad de datos en un formato resumido.

## Líneas de investigación y desarrollo

Afortunadamente existen varias bases de datos NoSql conocidas. Mientras todas ellas comparten muchas características, también hay algunas diferencias significativas

Base de Datos	Características	Aplicabilidad
<b>Mongo DB</b>	<ul style="list-style-type: none"> <li>• Escrito en C++Algunas características SQL (Query, index)</li> <li>• Protocolo binario</li> <li>• Replicación maestro-esclavo</li> </ul>	<p>Para queries dinámicasNo para Map/Reduce</p> <p>Cuando necesitas CouchDB con muchos cambios</p> <p>Para muchas</p>



	<ul style="list-style-type: none"> <li>• Sharding</li> <li>• Permite ejecutar Javascript</li> <li>• Rendimiento sobre características</li> <li>• En 32bits sólo 2.5 Gb</li> <li>• GridFS para almacenar BigData</li> <li>• Indexing geoespacial</li> </ul>	cosas que ahora se hacen con MySQL		<ul style="list-style-type: none"> <li>• en memoria con backup en disco</li> <li>• Sin disk-swap</li> <li>• Replicación maestro-esclavo</li> <li>• Clave-Valor, pero valores pueden ser List, Hashes, Sets,</li> <li>• Transacciones</li> </ul>	tamaño)Bases de datos de tiempo real	
<b>Riak</b>	<ul style="list-style-type: none"> <li>• Escrito en Erlang y CTolerancia a fallos</li> <li>• Protocolo binario o HTTP/REST</li> <li>• Pre y postcommits en JS y Erlang</li> <li>• Map/Reduce en JS o Erlang</li> <li>• Soporte objetos grandes</li> <li>• Versión opensource y Enterprise</li> <li>• Búsqueda Full text, indexing,... con Riak Search Server</li> <li>• Multireplicación</li> </ul>	Estilo Cassandra pero sin su complejidad	Escalabilidad, disponibilidad y tolerancia a fallos	<b>HBase</b>	<ul style="list-style-type: none"> <li>• Escrito en JavaBillones de filas x millones de columnas</li> <li>• Protocolo HTTP/REST y Thrift</li> <li>• Basada en Google Big Table</li> <li>• Map/Reduce con Hadoop</li> <li>• Optimizaciones para queries en tiempo real</li> <li>• Gateway Thrift de alto rendimiento</li> <li>• HTTP soporta XML, protobuf y binario</li> <li>• Módulos para Cascading, Hive y Pig</li> <li>• Shell basada en JRuby</li> <li>• Rendimiento random-acces como MySQL</li> </ul>	Analíticas Stock prices Comunicación en tiempo real
<b>CouchDB</b>	<ul style="list-style-type: none"> <li>• Escrito en ErlangProtocolo HTTP/REST</li> <li>• Replicación bidireccional con detección de conflictos</li> <li>• MVCC</li> <li>• Versiones previas de documentos disponibles</li> <li>• Necesita compactación de vez en cuando</li> <li>• Soporta attachment</li> <li>• Incluye librería JQuery</li> </ul>	Para acumular datos que sólo cambian ocasionalmente con queries predefinidasPara sistemas que necesitan versionado	Sistemas CRM Sistemas con replicación	<b>Neo4J</b>	<ul style="list-style-type: none"> <li>• Escrito en JavaBase de datos de grafos</li> <li>• Protocolo HTTP/REST o Java</li> <li>• Funcionamiento standalone o embebido en Java</li> <li>• Full ACID</li> <li>• Lenguaje de queries pattern-maching</li> <li>• Web de administración incluida</li> <li>• Path-finding algoritmos</li> </ul>	Mejor opción para Map/ReduceAlmacenamiento y análisis ficheros de log
<b>Redis</b>	<ul style="list-style-type: none"> <li>• Escrito en C++Muy rápida</li> <li>• Protocolo estilo Telnet</li> <li>• Bases de datos</li> </ul>	Para datos cambiantes almacenados en memoria (no de gran				

	<ul style="list-style-type: none"> <li>• Indexado de claves y relaciones</li> <li>• Optimizado para lecturas</li> <li>• Transacciones en API Java</li> <li>• Scripts en Groovy</li> <li>• Backup online, monitorización y alta disponibilidad en versión comercial AGPL</li> </ul>	
<b>Cassandra</b>	<ul style="list-style-type: none"> <li>• Escrito en JavaLo mejor de BigTable y Dinamo</li> <li>• Protocolo binario (Thrift)</li> <li>• Tuneable para distribución o replicación</li> <li>• Búsqueda por columnas o rango de claves</li> <li>• Características BigTable</li> <li>• Índices secundarios</li> <li>• Escrituras más rápidas que lecturas</li> <li>• Map/Reduce con Hadoop</li> </ul>	<p>Mas escritura que lectura (logging) Cuando todos los componentes son Java</p> <p>Análisis tiempo real</p>
<b>Membase</b>	<ul style="list-style-type: none"> <li>• Escrito en Erlang y CCompatible Memcached pero con persistencia y clustering</li> <li>• Protocolo memcached</li> <li>• Acceso muy rápido por clave</li> <li>• Persistencia en disco</li> <li>• GUI para gestión del cluster</li> <li>• Actualizaciones de SW sin parar la BD</li> </ul>	<p>Aplicaciones con acceso de muy baja latenciaAplicaciones con alta concurrencia y alta disponibilidad</p> <p>Online gaming (Zinga)</p>

En general, en el área de la computación distribuida Map se utiliza para fraccionar una operación compleja entre varios nodos y Fold/Reduce para recoger los resultados y unificarlos.

Por su parte, los [frameworks MapReduce](#) toman la base de las operaciones mencionadas anteriormente para crear una operación genérica y más compleja, cuyo funcionamiento es realmente útil para las bases de datos NoSQL: en vez de usarse sobre listas de valores unidimensionales, ésta toma como parámetros entrantes una lista de tuplas de tipo (*clave, valor*) y devuelve una lista de valores. Entre las operaciones [map](#) (distribuida) y [reduce](#) (normalmente localizada) se genera una lista de tuplas (*clave, valor*) con valores temporales, de las que [reduce](#) filtra solamente las que tengan una determinada clave.

En definitiva, MapReduce es fundamental en las bases de datos NoSQL para permitir la utilización de funciones de agregación de datos, ya que al carecer de esquema son mucho más complicadas que en las bases de datos relacionales clásicas [RDBMS](#) (*Relational Database Management System*).

## RESULTADOS Y OBJETIVOS

### Resultados Obtenidos

Se han publicado siete (7) trabajos de investigación en diferentes Congresos y Jornadas, y tres (3) trabajos de divulgación: un trabajo en el Symposium Internacional. La computadora contra el estigma de la discapacidad. Montevideo. 2009[26], otro trabajo en el Workshop de Investigadores en Ciencias de la Computación 2010 [1], otro en el WICC 2011 [8], dos (2) trabajos en el Congreso Argentino de Ciencias de la Computación 2010 [2] [3], otro en Perú [4] otro en V Seminario Internacional “De legados y Horizontes para el Siglo XXI”, 2010, organizado por RUEDA [4], y otro en el 1º Seminario Argentina Brasil de Tecnologías de la Información y la Computación [7]. El trabajo de divulgación es citado en [6]. También se ha desarrollado una tesina en Windows Azure, otra con Google AppEngine, otra

sobre Eye OS, y tres basadas en aplicaciones para lengua de señas.

### Resultados Esperados

El objetivo del grupo de investigación en esta línea, es la comparación en varios aspectos de las distintas Bases de Datos NoSql, y la integración de distintas tecnologías Open Source como las que propone Google.

### FORMACION DE RECURSOS HUMANOS

El proyecto marco sobre el que se realizan las investigaciones comenzó 2010, las publicaciones y trabajos de divulgación se han desarrollado en colaboración con becarios y alumnos avanzados, como [1] [2] [3] [4] [5] [6] [7] [8].

Se desarrolló una beca de investigación de alumno avanzado sobre Mobile Cloud Computing y se está desarrollando otra sobre orquestación en servicios Web. Se aprobaron tres (3) tesinas de tecnicatura sobre el tema de lengua de señas, dos de ellas sobre plataforma móvil y la otra implementando servicios Web. Por otro lado también se han aprobado 3 (tres) tesinas de licenciatura sobre Cloud Computing, sobre distintas tecnologías como Windows Azure y Google App Engine, una de las cuales se integra con una plataforma móvil con el OS Android. Además se encuentra en desarrollo y se proyectan dos de licenciatura y otra de tecnicatura, y se espera realizar alguna tesis de maestría y aumentar el número de publicaciones. Por otro lado también se prevé la divulgación de varios temas investigados por medio de cursos de postgrado y actualización o publicaciones de divulgación.

### REFERENCIAS

- [1] Murazzo, Rodríguez. "Mobile Cloud Computing". WICC 2010. Calafate. Mayo 2010.
- [2] Murazzo, Millán, Rodríguez, Segura, Villafañe. Desarrollo de aplicaciones para

Cloud Computing. CACIC 2010. Morón. Oct. 2010.

[3] Murazzo, Rodríguez, Millán, Segura y Villafañe. "Plataformas Educativas Implementadas Con Cloud Computing". XVI Congreso Argentino de Ciencias de la Computación – CACIC 2010, Workshop de Tecnologías Informáticas Aplicadas a la Educación. Morón. Oct. 2010.

[4] Murazzo, Rodríguez. "Una propuesta para el desarrollo de aplicaciones para Mobile Cloud Computing". Congreso Internacional de Computación y Telecomunicaciones – COMTEL 2010, Lima, Perú. Oct. 2010.

[5] Millán, Murazzo, Rodríguez. "Plataformas Educativas Implementadas Con Mobile Cloud Computing". V Seminario Internacional "De legados y Horizontes para el Siglo XXI", organizado por RUEDA. Tandil. Sep. 2010.

[6] Rodríguez, Murazzo, Ene. "Cloud Computing". X Workshop de investigadores en Ciencias de la Computación y Sistemas de Información. San Juan. Mayo 2009.

[7] Nelson R. Rodríguez, María A. Murazzo, Cecilia di Sciocio. "Integración de Computación móvil con Cloud Computing". 1º Seminario Argentina Brasil de Tecnologías de la Información y la Computación; bajo el lema "Las TIC como oportunidad de integración". Rosario Nov. 2011.

[8] Rodríguez, Chavez, Martin, Murazzo, Valenzuela. "Interoperabilidad en Cloud Computing". XII Workshop de investigadores en Ciencias de la Computación y Sistemas de Información. Rosario. Mayo 2011.

[9] Rodríguez, Villafañe, Murazzo, Gallardo, Tarrachano. "GAE, una estrategia para complementar SaaS y PaaS a través de la Web". 2do SABTIC. Tres de Maio, Brasil. Agosto 2012.

### BIBLIOGRAFIA CONSULTADA

Antonopoulos- Gillan "Cloud Computing Principles, Systems and Applications" Springer 2010- 978-1-84996-241-4

### LINK DE INTERES

<http://www.nosql.es/blog/nosql/mapreduce.html>

## PRÁCTICAS Y APLICACIONES DE INGENIERÍA DE REQUISITOS EN PROYECTOS DE EXPLOTACIÓN DE INFORMACIÓN

Pollo-Cattaneo, M., Pytel, P., García-Martínez, R., Vegega, C., Amatriain, H., Ramón, H., Mansilla, D., Deroche, A., Cigliuti, P., Saavedra-Martínez, P., Garbarini, R., Rodriguez, D., Britos, P., Tomasello, M.

Grupo de Estudio en Metodologías de Ingeniería de Software

Universidad Tecnológica Nacional.  
Facultad Regional Buenos Aires.  
Medrano 951 (C1179AAQ) Ciudad Autónoma de Argentina. Buenos Aires Tel +54 11 4867-7511  
fpollo@posgrado.frba.utn.edu.ar

Grupo Investigación en Sistemas de Información

Departamento de Desarrollo Productivo y Tecnológico  
Universidad Nacional de Lanús  
29 de Septiembre 3901 (1826) Remedios de Escalada, Lanús. Argentina.  
Tel +54 11 6322-9200 Ext. 194  
rgarcia@unla.edu.ar

Grupo de Investigación en Explotación de Información

Universidad Nacional de Río Negro  
San Martín esq. Pellegrini (8430) El Bolsón. Río Negro. Argentina.  
TE + 54 11 02944 49-8939  
paobritos@gmail.com

### RESUMEN

La Explotación de Información es la sub-disciplina de la Informática que aporta las herramientas de análisis y síntesis para extraer conocimiento no trivial que se encuentra implícitamente en los datos disponibles en distintas fuentes de información. Aunque la Ingeniería en Software provee muchos métodos, técnicas y herramientas para la construcción de software tradicional, estos en general no son aplicables a proyectos de Explotación de Información por poseer características diferentes. En especial se ha detectado la falta de metodologías que permitan cubrir adecuadamente la gestión de los requisitos de un Proyecto de Explotación de Información, permitiendo tanto la identificación y formalización de las necesidades del cliente como sus expectativas, restricciones y los principales conceptos que son necesarios para realizar el proyecto.

En este contexto, este proyecto se enfocará en sistematizar el cuerpo de conocimiento existente en la Ingeniería en Software y la Ingeniería del Conocimiento para sentar las bases para el desarrollo de una Ingeniería de Requisitos con particular énfasis en Proyectos de Explotación de Información. De esta manera se espera también producir un enriquecimiento en los campos. En relación a la transferencia específica se puede indicar la generación de un Modelo de Proceso para la gestión de Requisitos.

**Palabras clave:** Proyecto de Explotación de Información, Requerimiento, Ingeniería en Software, Information Mining.

### CONTEXTO

Este proyecto de investigación se desarrolla en el marco de la cooperación existente entre el Grupo de Investigación en Sistemas de Información (GISI) del Departamento de Desarrollo Productivo y Tecnológico de la Universidad Nacional de Lanús, el Grupo de Estudio en Metodologías de Ingeniería de Software (GEMIS) de la Facultad Regional Buenos Aires de la Universidad Tecnológica Nacional, y el Grupo de Investigación en Explotación de Información de la Universidad Nacional de Río Negro (SAEB-UNRN).

### INTRODUCCIÓN

La Inteligencia de Negocio propone un abordaje interdisciplinario (dentro del que se encuentra la Informática), que se centra en generar conocimiento que contribuya con la toma de decisiones de gestión y generación de planes estratégicos en las organizaciones [Thomsen, 2003]. La Explotación de Información (EdI) es la sub-disciplina de la Informática que aporta a la Inteligencia de Negocio [Negash & Gray, 2008] las herramientas de análisis y síntesis para extraer conocimiento no trivial que se encuentra (implícitamente) en los datos disponibles de diferentes fuentes de información [Schiefer et al., 2004]. Para un experto, o para el responsable de un Sistema de Información, normalmente no son los datos en sí lo más relevante, sino el conocimiento que se encierra en sus relaciones, fluctuaciones y dependencias. Un Proceso de Explotación de Información puede definirse como un conjunto de tareas relacionadas lógicamente [Curtis et



al., 1992], que se ejecutan para lograr, a partir de un conjunto de información con un grado de valor para la organización, otro conjunto de información con un grado de valor mayor que el inicial [Ferreira et al., 2005; Kanungo, 2005]. Identificado el problema de Inteligencia de Negocio, un Ingeniero de Explotación de Información decide la secuencia de Procesos de Explotación de Información que deben ser ejecutados para obtener una solución al problema de Inteligencia de Negocio. Cada Proceso de Explotación de Información tiene varias técnicas de Minería de Datos que pueden ser elegidas para realizarlas [García-Martínez et al., 2011a]. Muchas de esas técnicas vienen del campo del Aprendizaje Automático [García-Martínez et al., 2003].

En [García-Martínez et al., 2011b] se ha observado el uso indiscriminado de los términos “Minería de Datos” (o “Data Mining” en inglés) y “Explotación de Información” (o “Information Mining” en inglés) para referirse al mismo cuerpo de conocimientos. Esto es un tipo de confusión similar a utilizar como sinónimos “Ciencias de la Computación” y “Sistemas de Información”. La Minería de Datos está relacionada a la tecnología (algoritmos) necesaria para la obtención de conocimiento y la Explotación de Información está relacionada con los procesos y las metodologías necesarias para realizar el proyecto con éxito. De esta forma, la Minería de Datos se encuentra más cerca de las tareas de programación mientras que la Explotación de Información está cercana a la Ingeniería en Software.

Aunque la Ingeniería en Software y la Ingeniería del Conocimiento [García-Martínez & Britos, 2004] provee muchos métodos, técnicas y herramientas, estos no son útiles ya que no se ocupan de los aspectos específicos que poseen los Proyectos de Explotación de Información (PEI). Mientras, que las herramientas tradicionales de elicitación de la Ingeniería en Software se enfocan en la descripción de los diferentes tipos de requerimientos haciendo hincapié en las características que debe cumplir el producto software final [Wieggers, 2003], un Proyecto Explotación de Información no busca la

construcción del sistema software sino la aplicación de un proceso que convierta los datos disponibles en conocimiento. Por lo tanto, se considera necesario el desarrollo y la validación de métodos, técnicas y herramientas que puedan asistir a los practicantes del área de ingeniería de requisitos y proveer la necesaria objetividad, racionalidad, generalización y confiabilidad a los Proyectos de Explotación de Información.

En este contexto, se ha identificado como un problema abierto la necesidad de organizar un nuevo cuerpo de conocimiento relacionado a la identificación, formalización y administración de los deseos y necesidades de los interesados en un Proyecto de Explotación de Información. Una de las razones para desarrollar una Ingeniería de Requisitos de Explotación de Información ha sido el descubrimiento de una falta de actividades y técnicas asociadas a la ejecución de las fases correspondientes en las metodologías, normalmente aplicadas para el desarrollo de Proyectos de Explotación de Información. Si bien se destacan las metodologías SEMMA [SAS, 2008], P3TQ [Pyle, 2003] y CRISP-DM [Chapman et al., 2000] las cuales se consideran probadas y tienen un buen nivel de madurez, éstas dejan de lado aspectos a nivel operativo sobre las tareas necesarias para elicitar, formalizar y administrar los requisitos. Por otro lado, existen aproximaciones que tratan de integrar el conocimiento, como es el ciclo de vida propuesto por [Kimball et al., 2011] que es utilizado en iniciativas de Data Warehouse & Business Intelligence. Sin embargo, es necesario adaptar este ciclo de vida para considerar las particularidades de los Proyectos de Explotación de Información.

En este sentido, se han detectado la falta de procesos que permitan cubrir adecuadamente los siguientes aspectos de la Ingeniería de Requisitos en Proyectos de Explotación de Información:

- identificación de las principales necesidades del cliente, sus expectativas, restricciones y los principales conceptos que son necesarios para realizar el proyecto [Pollo-Cattaneo et al., 2009].
- el proceso de elicitación de requisitos mediante la transformación de la

experiencia en el uso de las herramientas disponibles para la elicitación en el dominio de los sistemas tradicionales [Mansilla et al., 2012].

- el proceso de formalización de requisitos a través de la utilización de técnicas de representación de conocimiento provistas por la Ingeniería del Conocimiento [Vegega et al., 2012].

### OBJETIVOS E HIPOTESIS DE INVESTIGACION

Este proyecto se inscribe en una línea de investigación que busca adaptar buenas prácticas ingenieriles (es decir procesos, métodos, técnicas y herramientas) existentes en la Ingeniería en Software y la Ingeniería del Conocimiento como también desarrollar nuevas para ser utilizadas en la Gestión de Requisitos de los Proyectos de Explotación de Información teniendo en cuenta las particularidades de las Pequeñas y Medianas Empresas.

Entre los supuestos (o hipótesis) que guían el proyecto se encuentran los siguientes:

- *Hipótesis I:* Los proyectos de Explotación de Información poseen características muy distintas a las de los proyectos de desarrollo de software tradicional y no tradicional. En consecuencia, muchas de los procesos, métodos, técnicas y herramientas no pueden ser aplicables a proyectos de Explotación de Información.
- *Hipótesis II:* A pesar de existir metodologías para el desarrollo de proyectos de Explotación de Información, muchas de estas metodologías dejan de lado aspectos a nivel operativo por no indicar el método y las técnicas que deberían ser utilizadas para la gestión de requisitos.
- *Hipótesis III:* Los riesgos y problemas encontrados durante el desarrollo de un proyecto de Explotación de Información pueden ser reducidos mediante la aplicación de buenas prácticas provenientes de la Ingeniería en Software y la Ingeniería del Conocimiento.

Teniendo en cuenta dichas hipótesis se buscará:

- *Objetivo General:* En el marco de proyectos de Explotación de Información, este proyecto se enfocará en sistematizar el

cuerpo de conocimientos existente en la Ingeniería en Software y la Ingeniería del Conocimiento para sentar las bases para el desarrollo de prácticas ingenieriles para la Gestión de Requisitos en Proyectos de Explotación de Información.

Asociados a este objetivo general se definen los siguientes objetivos específicos:

- *Objetivo Específico 1:* Identificar las principales actividades, técnicas y herramientas para ser utilizadas en cada fase del proceso de elicitación de requisitos de un proyecto de Explotación de Información.
- *Objetivo Específico 2:* Identificar las principales actividades, técnicas y herramientas para ser utilizadas en cada fase del proceso de formalización de requisitos de un proyecto de Explotación de Información.
- *Objetivo Específico 3:* Proponer un modelo que permita la gestión de requisitos en proyectos de Explotación de Información dentro del marco de un ciclo de vida definido ad-hoc.

### FUNDAMENTACION E IMPORTANCIA

La necesidad de desarrollar una Ingeniería de Requisitos para Proyectos de Explotación de Información surge del relevamiento efectuado en el campo metodológico, en el que se identifica la carencia de técnicas asociadas a la ejecución de las primeras fases planteadas en las metodologías identificadas. En este contexto, este proyecto promueve el desarrollo y la validación de métodos, técnicas y herramientas, conllevando a una mejora en el campo de la Ingeniería de Requisitos. Los métodos con abordaje ingenieril permiten dotar al proceso de desarrollo de: objetividad, sistematicidad, racionalidad, generalidad y fiabilidad, contribuyendo al avance del conocimiento científico mediante el uso de técnicas consistentes. De esta manera, los resultados de este proyecto pueden ser transferidos, en sus distintos niveles de gestión, para formular políticas en varias dimensiones. El crecimiento de los proyectos informáticos con aplicación de técnicas inteligentes en la actualidad delimita la importancia de utilizar

eficientemente los recursos disponibles para el desarrollo de las actividades correspondientes.

### **METODOLOGÍA DE TRABAJO**

Las tareas a realizar para desarrollar los procesos serán las siguientes:

- a) Se realizará una investigación documental exploratoria tratando de identificar los conceptos de interés para caracterizar un proyecto de Explotación de Información.
- b) Se buscará identificar relaciones existentes entre los diversos conceptos identificados, sus posibles fuentes y su posible aplicación para realizar la gestión de requisitos de un proyecto de Explotación de Información.
- c) Para el proceso de elicitación de requisitos se realizarán las siguientes actividades:
  1. Se identificarán y analizarán las técnicas y herramientas existentes en la Ingeniería en Software aplicables para proyectos de Explotación de Información.
  2. Se determinará el orden de las técnicas y herramientas para realizar la elicitación de requisitos en proyectos de explotación de información indicando las entradas y salidas necesarias.
  3. Se realizarán pruebas de concepto para validar el proceso de elicitación desarrollado.
- d) Para el proceso de formalización de requisitos se realizarán las siguientes actividades:
  1. Se identificarán y analizarán las técnicas y herramientas existentes en la Ingeniería en Software e Ingeniería del Conocimiento aplicables para proyectos de Explotación de Información.
  2. Se determinará el orden de las técnicas y herramientas para realizar la formalización de requisitos en proyectos de Explotación de Información indicando las entradas y salidas necesarias.
  3. Se realizarán pruebas de concepto para validar el proceso de formalización desarrollado.
- e) A partir de los resultados de los proceso de elicitación y formalización de requisitos, se propondrá un modelo que permita la gestión de requisitos en proyectos de Explotación

de Información dentro del marco de un ciclo de vida definido ad-hoc.

- f) Se realizarán pruebas de concepto para validar el modelo de proceso desarrollado.

Para el desarrollo de las tareas propuestas se utilizarán los siguientes materiales:

- Metodologías CRISP-DM [Chapman, et al., 2000], P3TQ [Pyle, 2003] y SEMMA [SAS, 2008].
- Ciclo de Vida para iniciativas de Data Warehouse & Business Intelligence [Kimball et al., 2011].
- Trabajos previos de los integrantes del proyecto relacionados con la Explotación de Información.
- Métodos, técnicas y herramientas existentes en Ingeniería en Software.
- Métodos, técnicas y herramientas existentes en Ingeniería del Conocimiento.
- Casos de estudio tomados tanto de proyectos reales desarrollados por los integrantes del proyecto como de trabajos de tesis de maestría y/o especialización disponibles (como por ejemplo son [Cigliuti et al., 2012] y [Saavedra-Martínez et al., 2012]).
- Reportes sobre proyectos de Explotación de Información de alumnos de la carrera Ingeniería en Sistemas de Información de la UTN-FRBA y UTN-FRCU.
- Hemerotecas y Repositorios Científicos accesibles por Internet.
- Acceso a bibliotecas digitales de IEEE, ACM y SCOPUS.

### **RESULTADOS OBTENIDOS/ESPERADOS**

Como resultado de este proyecto, se esperan los siguientes aportes:

- Generar un modelo de proceso para la gestión de requisitos dentro de proyectos de Explotación de Información con la especificación de un ciclo de vida ad-hoc, la cual será transferible a una Software Factory.
- Mejorar la comprensión de las potenciales aplicaciones de la disciplina por parte de los docentes integrantes del proyecto.

**FORMACIÓN DE RECURSOS HUMANOS**

Este proyecto busca tanto la obtención de nuevos conocimientos como la motivación de los implicados para que asciendan dentro del escalafón de la carrera de investigadores.

Los recursos humanos deben poseer una firme vocación de trabajar en el área de la ingeniería en software, los sistemas inteligentes con su vinculación a la Explotación de Información, de manera de canalizar y proveer una base sustentable de aporte al proyecto. De esta manera, se logra un doble beneficio, el proyecto obtiene e incorpora el conocimiento tecnológico de los recursos humanos en el área de la especialidad, a la vez que plantea un esquema de formación de especialistas de punta en el proceso de gestión.

Además, en el marco de este proyecto de investigación se prevé la radicación de una Tesis de Doctorado en Ciencias Informáticas y cinco tesis de magister en distintas áreas de la informática en codirección con otras instituciones. También se espera que los alumnos becarios que participan del proyecto tengan la posibilidad de articular su Trabajo Final de Carrera de grado y posgrado con temas relacionados a este proyecto.

**REFERENCIAS**

- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. & Wirth, R. (2000). CRISP-DM 1.0 Step-by-step Data Mining Guide. <http://tinyurl.com/crispdm> Último acceso Enero de 2013.
- Cigliuti, P., Pollo-Cattaneo, F., García-Martínez, R. (2012). Procesos de Identificación de Comportamiento de Comunidades Educativas Centradas en EVEAs. Proceedings del XIV Workshop de Investigadores en Ciencias de la Computación. Pág. 954-956. ISBN 978-950-766-082-5.
- Curtis, B., Kellner, M., Over, J. (1992). Process Modelling. Communications of the ACM, 35(9): 75-90.
- Ferreira, J., Takai, O. & Pu, C. (2005). Integration of Business Processes with Autonomous Information Systems: A Case Study in Government Services. Proceedings Seventh IEEE International Conference on E-Commerce Technology. Pág. 471-474.
- García Martínez, R. y Britos, P. (2004). Ingeniería de Sistemas Expertos. Editorial Nueva Librería.
- García-Martínez, R., Servente, M. y Pasquini, D. (2003). Sistemas Inteligentes. Editorial Nueva Librería. ISBN 987-1104-05-7
- García-Martínez, R., Britos, P., Pollo-Cattaneo, F., Rodríguez, D., Pytel, P. (2011a). Information Mining Processes Based on Intelligent Systems. Proceedings of II International Congress on Computer Science and Informatics (INFONOR-Chile 2011). Pág. 87-94. ISBN 978-956-7701-03-2.
- García-Martínez, R., Britos, P., Pesado, P., Bertone, R., Pollo-Cattaneo, F., Rodríguez, D., Pytel, P., Vanrell, J. (2011b). Towards an Information Mining Engineering. En Software Engineering, Methods, Modeling and Teaching. Sello Editorial Universidad de Medellín. ISBN 978-958-8692-32-6. Pág. 83-99.
- Kanungo, S. (2005). Using Process Theory to Analyze Direct and Indirect Value-Drivers of Information Systems. Proceedings of the 38th Annual Hawaii International Conference on System Sciences. Pág. 231-240.
- Kimball, R., Ross, M., Thornthwaite, W., Mundy, J., & Becker, B. (2011). The data warehouse lifecycle toolkit. Wiley & Sons.
- Mansilla, D., Pollo-Cattaneo, M., Pytel, P., García-Martínez, R. (2012). Modelo de Proceso para Elicitación de Requerimientos en Proyectos de Explotación de Información. Proceedings del XIV Workshop de Investigadores en Ciencias de la Computación. Pág. 222-225. ISBN 978-950-766-082-5
- Negash, S. & Gray, P. (2008). Business Intelligence. En In Handbook on Decision Support Systems 2, ed.eds. F. Burstein y C. Holsapple (Heidelberg, Springer), Pág.Pp. 175-193.
- Pollo-Cattaneo, F., Britos, P. Pesado, P., García-Martínez, R. (2009). Metodología para Especificación de Requisitos en Proyectos de Explotación de Información. Proceedings XI Workshop de Investigadores en Ciencias de la Computación. Pág. 333-335. ISBN 978-950-605-570-7.
- Pyle, D. (2003). Business Modeling and Business intelligence. Morgan Kaufmann Publishers.
- Saavedra-Martínez, P., Pollo-Cattaneo, M., Pytel, P., Rodríguez, D., García-Martínez, R. (2012). Detección de Problemas de Aprendizaje Basado en Explotación de Información. Proceedings del XIV Workshop de Investigadores en Ciencias de la Computación. Pág. 962-964. ISBN 978-950-766-082-5.
- SAS (2008). SAS Enterprise Miner: SEMMA. <http://tinyurl.com/semmaSAS> Último acceso Enero de 2013.
- Schiefer, J., Jeng, J., Kapoor, S. & Chowdhary, P. (2004). Process Information Factory: A Data Management Approach for Enhancing Business Process Intelligence. Proceedings 2004 IEEE International Conference on E-Commerce Technology. Pág.Pp. 162-169.
- Thomsen, E. (2003). BI's Promised Land. Intelligent Enterprise, 6(4): 21-25.
- Vegega, C., Mansilla, D., Pollo-Cattaneo, M. F., Pytel, P., Rodríguez, D., Diez, E., García-Martínez, R. (2012). Documentación de Requisitos en Proyectos de Explotación de Información. Proceedings del XIV Workshop de Investigadores en Ciencias de la Computación. Pág. 226-229. ISBN 978-950-766-082-5.
- Wieggers, K. (2003). Software Requirements. Microsoft Press.



# Geometría Computacional y Bases de Datos

Maria Gisela Dorzán, Susana Esquivel, Edilma Olinda Gagliardi,  
Pablo Palmero y Maria Teresa Taranilla

Departamento de Informática, Facultad de Ciencias Físico Matemáticas y Naturales  
Universidad Nacional de San Luis, Argentina  
{mgdorzan, esquivel, oli, prpalmero, tarani }@unsl.edu.ar

Gregorio Hernández Peñalver

Departamento de Matemática Aplicada, Facultad de Informática  
Universidad Politécnica de Madrid, España  
gregorio@fi.upm.es

Maria Mercedes Guasch y Maria Rosana Piergallini

Escuela de Informática y Tecnología  
Universidad Nacional del Noroeste de la Provincia de Buenos Aires, Argentina  
mariarosanapier@hotmail.com, mercedes\_guasch@yahoo.com

## Resumen

El proyecto *Tecnologías Avanzadas de Bases de Datos*, en el que se enmarca la línea de investigación *Geometría Computacional y Bases de Datos*, investiga sobre el diseño y desarrollo de herramientas para administrar eficientemente sistemas de bases de avanzadas. En la línea de investigación se vinculan las disciplinas Bases de Datos, Geometría Computacional y Metaheurísticas, utilizando métodos y herramientas provistas por tales disciplinas en la resolución de los problemas surgidos en las temáticas de investigación.

**Palabras clave:** Bases de datos, Geometría Computacional, Metaheurísticas, Bases de datos espacio-temporales.

## Contexto

El proyecto *Tecnologías Avanzadas de Bases de Datos* de la Universidad Nacional de San Luis tiene como objetivo principal el estudio de bases de datos avanzadas, y en este marco investigamos el diseño y desarrollo de herramientas para administrar eficientemente

sistemas de bases de datos no estructuradas.

En el proyecto existen tres líneas de investigación, orientadas al desarrollo de nuevos modelos para administrar y recuperar información almacenada en repositorios de datos no estructurados, donde los escenarios de exploración requieren modelos tales como las bases de datos espacio-temporales, bases de datos de texto, bases de datos de imágenes, bases de datos de sonidos, espacios métricos, entre otros.

En diversas aplicaciones del campo de las Ciencias de la Computación son necesarios repositorios de datos no tradicionales, y aparecen nuevos modelos de bases de datos para buscar y administrar información en ellos. Así, surge el estudio de modelos como las bases de datos espacio-temporales. También, se requiere la construcción y manipulación de diferentes objetos y estructuras geométricas de utilidad en diversas áreas de aplicación, por ejemplo, robótica, sistemas de información geográfica, visión artificial, computación gráfica, computación móvil, entre otras, y que se relacionan con las bases de datos mencionadas. En particular, las estructuras geométricas que estudiamos deben cumplir con propiedades deseables, y algunos de los

problemas relacionados con la optimización de las mismas son problemas NP-duros, por tanto en la búsqueda de soluciones aproximadas utilizamos técnicas metaheurísticas. Por otro lado, en la línea de investigación *Geometría Computacional y Bases de Datos* hemos vinculado temáticas que surgen de las disciplinas Bases de Datos, Geometría Computacional y Metaheurísticas. Desde el año 2002, desarrollamos nuestro trabajo de investigación en forma conjunta con investigadores afines de proyectos locales y de la Universidad Politécnica de Madrid, en el marco de los convenios de cooperación institucional. En este trabajo, presentamos los temas en estudio, junto con las propuestas más recientes de interés.

## Introducción

Numerosas aplicaciones requieren guardar y consultar información histórica y actual, relacionada a los cambios de forma y/o posición que tuvieron los objetos estudiados en diferentes escenarios a lo largo del tiempo. Para tales requerimientos, es necesario contar con los modelos de bases de datos espacio-temporales y con herramientas teóricas, de base, que permitan modelar estos tipos de datos, realizar operaciones sobre ellos, definir lenguajes de consulta, analizar su expresividad, entre otras propiedades.

La Geometría Computacional estudia el diseño de algoritmos para resolver problemas geométricos. En esta disciplina se identifican conceptos, propiedades y técnicas que apuntan al desarrollo de algoritmos eficientes. Esto conduce al análisis y estudio de estructuras de datos geométricas y problemas relacionados con dichas estructuras [BKOS97]. Los problemas se estudian desde un punto de vista combinatorio o algorítmico. Los aspectos algorítmicos de los problemas aparecen en la búsqueda de soluciones, exactas o aproximadas, para una estructura geométrica particular. En este contexto aparecen problemas que son NP duros, o bien, son problemas para los cuales no se conocen algoritmos eficientes que permitan alcanzar soluciones exactas. De todos modos, es de

interés obtener soluciones a tales problemas, aunque sean aproximadas a las óptimas. En nuestro trabajo hacemos uso técnicas metaheurísticas, ya que las mismas permiten encontrar soluciones cercanas a las óptimas. Una metaheurística se puede considerar como una metodología de alto nivel que puede ser utilizada como una estrategia que guía las heurísticas subyacentes para resolver problemas de optimización específicos, combinando inteligentemente diferentes conceptos de diversos campos. Estos métodos son simples de implementar y han demostrado ser exitosos en encontrar, de forma eficiente, buenas soluciones para problemas de optimización NP-duros [MF04] [BFM97].

Los problemas que estudiamos corresponden a estructuras geométricas, obtenidas a partir de un conjunto de puntos u objetos en el plano, y para las cuales se busca optimizar alguna propiedad que permita valorar la calidad de dichas estructuras. Los criterios de optimización consisten en minimizar o maximizar alguna propiedad de las estructuras geométricas. Los problemas de optimización relacionados a triangulaciones, pseudo-triangulaciones y poligonizaciones, son de interés debido a que son utilizadas en diversos campos: detección de colisiones, visibilidad y vigilancia, problemas de rigidez de estructuras, entre otros.

Entre los problemas de optimización estudiados están: la Triangulación de Peso Mínimo (*Minimum Weight Triangulation, MWT*) y la Pseudo-Triangulación de Peso Mínimo (*Minimum Weight Pseudo-Triangulation, MWPT*). Dichos problemas minimizan la suma de las longitudes de las aristas proveyendo así una medida de calidad para determinar cuán “buena” es la estructura. La complejidad del cálculo de MWT fue uno de los problemas abiertos más estudiados en Geometría Computacional hasta que Mulzer y Rote demostraron que la construcción de MWT es un problema NP-duro [MR06]. Por otra parte, la complejidad de la pseudo-triangulación de peso mínimo aún no está resuelta. Levcopoulos y Gudmundsson muestran una 12-aproximación de una pseudo-triangulación que puede ser calculada con

complejidad de  $O(n^3)$  [GL07]. Los autores dan una aproximación de  $O(\log n w(\text{MST}))$  de una pseudo-triangulación de peso mínimo en un tiempo de  $O(n \log n)$  donde  $w(\text{MST})$  es el peso del Árbol de Expansión Mínimo (*Minimum Spanning Tree, MST*) el cual es un subconjunto de la estructura obtenida.

Otro problema en estudio es la Triangulación de Dilación Mínima (*Minimum Dilation Triangulation, MDT*) donde la dilación mide la calidad de conexión entre puntos de la triangulación. La dilación entre un par de puntos,  $u$  y  $v$ , de una triangulación  $T$  se define como la razón entre el camino más corto entre  $u$  y  $v$  en  $T$  y la distancia euclídea entre dicho par de puntos. La máxima dilación entre todos los pares de puntos en  $T$  se llama la dilación de  $T$  ( $\Delta(T)$ ). La mejor posible dilación de cualquier triangulación de un conjunto de puntos  $S$  se conoce como dilación de  $S$  ( $\Delta(S)$ ). La complejidad del problema de encontrar  $\Delta(S)$  para un conjunto de puntos  $S$  se desconoce y por lo tanto, nos centramos en el desarrollo de algoritmos aproximados para encontrar triangulaciones de alta calidad de dilación mínima.

Con respecto a poligonizaciones, estudiamos el problema de obtener poligonizaciones de área mínima/máxima para un conjunto de puntos en el plano. Este problema es NP-duro, según demostró Fekete [Fe00].

Debido a la complejidad de los problemas antes mencionados, aplicamos técnicas metaheurísticas para obtener soluciones aproximadas a las óptimas.

En la línea de investigación estudiamos el diseño y desarrollo de índices espacio-temporales, aplicables a diversos escenarios de movimiento (redes, espacios libres de obstáculos, etc.), considerando la geometría como una disciplina marco en la cual se formalizan aspectos propios de los problemas involucrados. En este contexto, nos proponemos el estudio de optimizaciones de estructuras geométricas que están relacionadas con las bases de datos ya mencionadas y al desarrollo de herramientas para la visualización de estructuras geométricas y

aplicaciones vinculadas a las bases de datos espacio-temporales.

### Línea de investigación

La Geometría Computacional estudia problemas desde un punto de vista geométrico, dedicándose al diseño y análisis de algoritmos y/o estructuras geométricas adecuados para su resolución. Algunos de los problemas de estudio la disciplina provienen de diversas áreas, por ejemplo en los sistemas de información geográfica, problemas de referidos a localización, planificación de rutas y búsqueda geométrica; en robótica, problemas de planificación de movimientos, de visualización y detección de colisiones, entre otros. En particular, los problemas de interés en la línea de investigación son aquellos vinculados al diseño de índices espacio-temporales para resolver en forma integral consultas espacio-temporales y su vinculación con la problemática de objetos móviles; y al tratamiento de problemas geométricos NP-duros. Para estos últimos, en la búsqueda de soluciones proponemos utilizar técnicas metaheurísticas, ya que proporcionan estrategias adecuadas para la obtención de soluciones aproximadas de múltiples problemas de optimización.

Como objetivos específicos de estudio podemos enumerar los siguientes:

i) Indexación espacio-temporal sobre objetos en movimiento para diversos escenarios. Desarrollo de las estructuras de almacenamiento, los algoritmos de consulta y la evaluación experimental, mostrando el desempeño de los distintos índices en aplicaciones de diferentes magnitudes respecto de la población de objetos en movimiento. Desarrollo de aplicaciones con herramientas de Geometría Computacional y Bases de Datos Espacio-Temporales. Desarrollo de una herramienta integral en el ámbito de la Salud para el seguimiento de focos epidémicos utilizando base de datos espacios-temporales.

ii) Estudio de configuraciones geométricas generales de puntos en el plano, que cumplan medidas de calidad como: peso, dilación, área, entre otras. Optimización de Triangulaciones,

Pseudo-triangulaciones y Poligonizaciones considerando medidas de calidad mínimas o máximas, aplicando técnicas metaheurísticas y diversas estrategias algorítmicas.

iii) Diseño y desarrollo de herramientas para la generación, visualización y manipulación de diferentes configuraciones geométricas de conjuntos de puntos en el plano.

## Resultados Obtenidos

Para la resolución de problemas de optimización de Triangulaciones y Pseudo-triangulaciones de Peso Mínimo aplicamos las técnicas metaheurísticas: Optimización basada en Colonia de Hormigas (Ant Colony Optimization, ACO) y Recocido Simulado (Simulated Annealing, SA), técnicas determinísticas Voraces (Greedy) y Triangulación de Delaunay. Se llevó a cabo el estudio, adecuación y evaluación experimental de las técnicas metaheurísticas mencionadas para la búsqueda de triangulaciones y pseudo-triangulaciones que cumplan la propiedad. Los resultados obtenidos para los problemas MWT y MWPT utilizando la técnica metaheurística ACO están publicados en [DGLH11b] [GDLH11] [DGLH12]. Se llevó a cabo una evaluación experimental y análisis de los resultados obtenidos con la técnica Recocido Simulado [DGLH11a] [DGLH11c].

Para el problema Triangulación de Dilación Mínima (Minimum Dilation Triangulation - MDT) todavía no se conoce un algoritmo que lo resuelva en tiempo polinomial y tampoco se ha demostrado que sea NP-duro. Para este problema se estudiaron, desarrollaron y aplicaron diferentes tipos de técnicas: Greedy, Local Search, Iterated Local Search, Simulated Annealing y Random Local Search proponiendo para cada estrategia un conjunto de operadores adecuados. Debido a la complejidad que implica la puesta a punto de los parámetros de técnicas metaheurísticas, se utilizó Optimización de Parámetros Secuencial (Sequential Parameter Optimization - SPO) para el ajuste de los parámetros requeridos por Simulated Annealing. Se realizó un análisis experimental en el cual se compararon dichos

algoritmos con otras técnicas, como por ejemplo, Delaunay. Se crearon las instancias de prueba, ya que para estos problemas no se encontraron disponibles ningún tipo benchmark con el cual comparar nuestros resultados. Las conclusiones fueron afirmadas desarrollando un estudio estadístico aplicando diferentes test estadísticos y métodos de visualización. Actualmente se encuentra en evaluación el artículo "Approximated algorithms for the Minimum Dilation Triangulation Problem" en la revista Journal of Heuristics (<http://www.springer.com/mathematics/applications/journal/10732>) donde se plasmaron todos los resultados obtenidos.

Para el problema de optimización de poligonizaciones de puntos en el plano, se utilizaron técnicas metaheurísticas basadas en Colonia de Hormigas (Ant Colony Optimization, ACO) y técnicas Voraces (Greedy) para aproximar poligonizaciones de área mínima. Los resultados obtenidos se publicaron en [TGH11].

Por otra parte, se ha implementado una herramienta para la generación y visualización de triangulaciones pseudo-triangulaciones y poligonizaciones de conjuntos de puntos en el plano. [DGP10].

Se está desarrollando una aplicación en el ámbito de la Salud para el seguimiento de focos epidémicos utilizando base de datos espacios-temporales y herramientas de Geometría Computacional.

Como trabajo futuro, específicamente para problemas geométricos, continuaremos con el estudio de los problemas de optimización considerando otras técnicas metaheurísticas, que puedan adecuarse para su resolución. Se llevarán a cabo, la evaluación de las distintas técnicas para determinar su funcionalidad e impacto en la comunidad científica y el análisis que incluirá el correspondiente tratamiento estadístico, y comparativo con otros algoritmos de tipo aproximado o de tipo exacto.

## Formación de Recursos Humanos

La formación del grupo de trabajo en la Universidad Nacional de San Luis, se



consolida con actividades de cooperación mutua e intercambio recíproco de información científica, tecnología y desarrollo de nuevos conocimientos con investigadores locales y de otras universidades.

Entre las actividades desarrolladas, se mencionan:

i) formación de recursos humanos plasmada en trabajos finales de Licenciatura en Ciencias de la Computación, Tesis de Maestría en Ciencias de la Computación y Tesis Doctorales en desarrollo, ii) direcciones de becas de investigación de Ciencia y Técnica-FCFM y N-UNSL y de CONICET, iii) realización y dirección de pasantías de investigación con docentes de otras universidades (UNNOBA), iv) actividades de formación académica, cursos de posgrado y de especialización, v) actividades de divulgación científica, conferencias y publicaciones en congresos y revistas en el ámbito nacional e internacional.

Nos proponemos continuar con las actividades integradoras y relacionadas al presente proyecto, proponiendo actividades de formación académica, de formación de recursos humanos locales y de otras universidades nacionales, investigación, desarrollo, y otras actividades académico-científicas vinculantes.

## Referencias

- [BFM97] Bäck T., Fogel D., Michalewicz Z. *Handbook of Evolutionary Computation*. IOP Publishing Ltd and Oxford University Press. 1997.
- [BKOS97] de Berg, M., van Kreveld, M., Overmars, M., Schwarzkopf O. *Computational Geometry: Algorithms and Applications*. Springer-Verlag, Heidelberg, 1997.
- [DGPH10] Dorzán M.G., Gagliardi, E. O., Palmero P.R., Hernández Peñalver, G. *Una herramienta para la generación y visualización de Triangulaciones y Pseudotriangulaciones*. XVI Congreso Argentino de Ciencias de la Computación 2010 (CACIC 2010), 2010
- [DGLH11a] Dorzán M.G., Gagliardi E.O., Leguizamón M.G., Hernández Peñalver G. *Approaches for MWT and MWPT Problems*. XVII Congreso Argentino de Ciencias de la Computación 2011 (CACIC 2011), 2011
- [DGLH11b] Dorzán M.G., Gagliardi E.O., Leguizamón M.G., Hernández Peñalver G. *Using ACO metaheuristic for MWT problem*. XXX International Conference of the Chilean Computer Science Society. ISBN 978-0-7695-4689-6. Chile. 2011
- [DGLH11c] Dorzán M.G., Gagliardi E.O., Leguizamón M.G., Hernández Peñalver G. *Metaheuristic approaches for MWT and MWPT Problems*. XIV Encuentros de Geometría Computacional. pág: 79-82. 2011
- [DGLH12] Dorzán M.G., Gagliardi E.O., Leguizamón M.G., Hernández Peñalver G. *Approximations on Minimum Weight Triangulations and Minimum Weight Pseudo-Triangulations using Ant Colony Optimization Metaheuristic*. *Fundamenta Informaticae*. ISSN: 0169-2968 (Print), 1875-8681 (Online). Volume 119, number 1, pp 1-27.
- [Fe00] Fekete, S. P. *On simple polygonalizations with optimal area*. *Discrete and Computational Geometry*, 23, pp. 73-110. 2000.
- [GDLH11] Gagliardi E. O., Dorzán M. G., Leguizamón M. G. y Hernández Peñalver. G.; *Approximations on Minimum Weight Pseudo-Triangulation problem using Ant Colony Optimization*. XXX International Conference of the Chilean Computer Science Society. ISBN 978-0-7695-4689-6. Chile. 2011
- [GL07] Gudmundsson J., Levkopoulos C.; *Minimum weight pseudo-triangulations*. *Computational Geometry. Theory and applications*. Elsevier Vol. 38-pages 139-153, 2007.
- [MR06] Mulzer W., Rote G. *Minimum weight triangulation is NP-hard*. In *Proceedings of the 22nd Annual ACM Symposium on Computational Geometry*. 2006.
- [MF04] Michalewicz Z., Fogel D., *How to Solve It: Modern Heuristics*, Springer, 2004.
- [TGH11] Taranilla, M. T.; Gagliardi, E. O.; Hernández Peñalver G. "Approaching Minimum Area Polygonization". XVII Congreso Argentino de Ciencias de la Computación 2011 (CACIC 2011), 2011.

## Minería de Datos en Bio-Ciencias

Prato Laura Beatriz<sup>1</sup>, Fresno Rodríguez Cristóbal<sup>2</sup>, Fernandez Elmer Andres<sup>2</sup>, Zingaretti María Laura<sup>1</sup>,  
Ribero Gabriela<sup>1</sup>, Villoria Liliana<sup>1</sup>

<sup>1</sup>Instituto AP de Ciencias Básicas y Aplicadas - Universidad Nacional de Villa María

<sup>2</sup>Grupo de Minería de Datos - Universidad Católica de Córdoba

Universidad Nacional de Villa María: Av. Arturo Jauretche 1555 – Villa María – Córdoba – Argentina  
0353-4539106 / 141

[lprato@unvm.edu.ar](mailto:lprato@unvm.edu.ar) - [cristobalfresno@gmail.com](mailto:cristobalfresno@gmail.com) – [elmerfer@gmail.com](mailto:elmerfer@gmail.com)

### Resumen

El campo de las Bio-ciencias está en pleno desarrollo y expansión. La variedad de tecnologías disponibles y aplicaciones están generando cantidades abrumadoras de datos que necesitan de protocolos, conceptos y métodos que permitan un análisis uniforme y asequible.

Otra característica distintiva de estos ámbitos es su condición multidisciplinaria, donde interactúan (y cada vez más) disciplinas como la biología, la matemática, la estadística, la informática, la inteligencia artificial; y sus aplicaciones sobre la agronomía, la salud humana y animal y el medio ambiente; por lo que cualquier esfuerzo tendiente a aumentar el nivel de comunicación y entendimiento entre las distintas disciplinas redundará en beneficios.

La Minería de Datos, concepto que aglutina una variedad de metodologías analíticas, proporciona un marco conceptual y metodológico para el abordaje del análisis de datos y señales en distintas disciplinas. Sin embargo cada campo de aplicación presenta desafíos específicos que deben ser abordados particularmente desde la racionalización de los conceptos específicos del ámbito.

En este proyecto se integrarán las experiencias y criterios de distintas disciplinas que están involucradas en el desarrollo experimental en bio-ciencias. La finalidad es elaborar protocolos y metodologías de análisis, desarrollar métodos analíticos para generar

nuevas estrategias diagnósticas, predictivas a partir de los datos recogidos que permitan extraer conocimiento en problemas biotecnológicos que se basen en investigación sólida de los procedimientos estadísticos/bioinformáticos relevante para el manejo de datos experimentales.

### Palabras clave:

Bioinformática - minería de datos - biotecnología - inteligencia artificial

### Contexto

Esta línea se inserta en un proyecto de desarrollo conjunto entre la Universidad Nacional de Villa María y la Universidad Católica de Córdoba. Ha sido evaluado por evaluadores externos, participa del Programa de Incentivos de la Nación para Docentes Investigadores, y es financiado por la Universidad Nacional de Villa María

### Introducción

El dominio de técnicas de análisis de Bio-datos, como la capacidad para diseñar y analizar experimentos cuyos resultados puedan ser luego transformados en nuevos dispositivos o técnicas de diagnóstico, es fundamental para hacer un uso más eficiente de los recursos destinados a la investigación y al desarrollo de productos biotecnológicos (vacunas, dispositivos de diagnóstico biomédico, mejoramiento de especies,

evaluación de drogas, mejoramiento vegetal y animal, calidad de alimentos, etc).

Los pasos involucrados en la adquisición, análisis e interpretación de Bio-datos son numerosos y su correcto abordaje es crucial para el éxito de la aplicación [Buckingham2003]. El modelado y la búsqueda de patrones tendientes al diseño de un método de diagnóstico o caracterización implican la utilización de estrategias analíticas capaces de interrelacionar una gran cantidad de variables con un evento particular. Los métodos de PLS y SVM son métodos modernos y prometedores para el análisis y modelado de señales y datos biomédicos/biotecnológicos, más aun si existen mecanismos que permitan modelar en cierto grado (a través de núcleos adecuados) algún tipo de conocimiento específico (prior-knowledge). La inclusión de conocimiento específico en sistemas de análisis inteligente es un tema candente en investigación de Minería de Datos y Reconocimiento de Patrones. Dominar este tipo de técnicas permitirá evaluar rápidamente su aplicabilidad en distintos ámbitos. Unos de los ámbitos que está demandando este tipo de metodologías de análisis es la genómica y proteómica de alto rendimiento, donde Argentina cuenta actualmente con secuenciadores masivos y está adquiriendo nuevo equipamiento a través de las convocatorias a plataforma PPL dado que están consideradas áreas de vacancia. En este contexto y en otros como la agrobiotecnología, la generación masiva de datos experimentales está requiriendo de la adaptación y/o creación de herramientas específicas para satisfacer los requerimientos de estas tecnologías y poder extraer información biológica relevante. En este proyecto se pretende desarrollar metodologías estadístico/computacionales para proveer de técnicas de minería de datos a estos sectores de investigación y

productivos de la región a través de un enfoque multidisciplinar, integrando conceptos y algoritmos provenientes de la Bioinformática, la Estadística, la Bioingeniería, la Biología y la Inteligencia Artificial, disciplinas que en general se tratan independientemente.

## Líneas de investigación y desarrollo

El proyecto sigue la línea de investigación del campo de las bio-ciencias, en especial relacionada a la bioinformática, la bioingeniería y la biotecnología. En ese sentido, se ha comenzado a trabajar en la exploración de librerías en R para análisis cluster, en procesamiento distribuido en bioinformática utilizando el lenguaje r, en exploración e implementación de herramientas de desarrollo en un sistema de análisis estadístico de experimentos proteómicos. Todas estas líneas tienden a determinar los mejores y más efectivos métodos de análisis de la información procedente de datos biológicos, relacionados a agronomía, veterinaria, salud, ambiente.

## Resultados y Objetivos

### Objetivos:

#### Científicos

1. Extender los principios del Descubrimiento de Conocimiento en bases de datos y Minería de Datos en las Biociencias (Bioingeniería/Biotecnología/Bioinformática)
2. Estudiar y comparar métodos de clasificación/predicción basados en inteligencia artificial y en estadística sobre Bio-datos.
3. Formar recursos humanos nacionales en bioinformática.

#### Tecnológicos

4. Desarrollar herramientas y aplicaciones de bioinformáticas (librerías, software) para facilitar el procesamiento de datos biotecnológicos.
1. Editar un material de referencia sobre el uso de las aplicaciones desarrolladas.

**Resultados:**

Algunos integrantes disertaron en distintos Congresos Nacionales sobre temas relacionados a la temática propuesta, tal el caso de Laura Prato, quien disertó sobre “Aplicaciones bioinformáticas en experimentos con proteínas” en la X Semana de la Ciencia y la Tecnología en Villa María, en Junio 2012; y Cristobal Fresno dictó un curso sobre “Lenguaje R aplicado a Bioinformática” en el Workshops RSG-Argentina – Edición 2012, en Oro Verde, Entre Ríos en Septiembre 2012.

Se realizaron las siguientes Presentaciones en Congresos: en el 3er Congreso Argentino de Bioinformática y Biología Computacional, Oro Verde, Septiembre 2012 (Sesión de Posters), se presentaron los siguientes trabajos:

- GOboot: towards a robust SEA analysis. Cristobal Fresno, Andrea Llera, María Romina Girotti, María Pía Valacco, Juan A Lopez, Laura Zingaretti, Laura Prato, Osvaldo Podhajcer, Mónica G Balzarini, Federico Prada y Elmer Fernández.
- One vs One Artificial Neural Network strategy for gene expression multiclass classification. Remón L, Juárez L, Arab Cohen D, Fresno C, Prato L, Villoria L y Fernández E.
- SVM Tree with Optimal Multiclass Partition applied to Gene expression signature classification. Pellarolo M, Arab Cohen D, Fresno C, Prato L y Fernández E.

Y en las VII Jornadas de Investigación, Villa María, Noviembre 2012, se presentaron:

- Minería de datos en bio-ciencias: bioinformática, bioingeniería y biotecnología, Elmer A Fernández, Laura Prato, Cristóbal Fresno
- Exploración de distintas librerías y funciones de r para análisis cluster María Laura Zingaretti, Laura Prato, Elmer Fernández
- Procesamiento distribuido en bioinformática utilizando el lenguaje r, Mayco Fraccaroli,

Emiliano Marzioni, Nicolás Ferreyra, Cristóbal Fresno, Laura Prato

**Formación de Recursos Humanos**

La estructura del equipo de trabajo tiene dos investigadores especialistas en la temática, un investigador encargado de la gestión, dos investigadores más relacionados al área técnico-informática y dos investigadores del área de desarrollo informático. También intervienen alumnos de grado en formación y para el desarrollo de sus respectivas tesinas.

Dos de los alumnos que intervienen han obtenido Beca para las Vocaciones Científicas, del CIN; y dos resultaron favorecidos con el Premio INTEL a la excelencia académica 2012 y 2013.

**Referencias**

- Gianola, D, Perez-Enciso M, Toro MA. 2003. On marker assisted prediction of genetic value: beyond the ridge. *Genetics* 163: 347–365.
- Gianola D, Fernando RL, Stella A. 2006. Genomic-assisted prediction of genetic value with semiparametric procedures. *Genetics* 173: 1761–1776.
- Churchill, Doerge RW. 1994. Empirical Threshold Values for Quantitative Trait Mapping. *Genetics* 138: 963-971.
- Kang M, Balzarini M. Guerra J. 2004. Genotype-by-Environment interaction. In A. Saxton (ed.) *SAS Genetic Book*, SAS Institute, Cary NC (in press).
- 1-Buckingham S. (2003) Programmed for success. *Nature*. V425 pp 209-215
- Balzarini y cols (2004). *Info-Gen: Software para análisis de datos genéticos*. Facultad de Ciencias Agropecuarias. Universidad Nacional de Córdoba. Argentina. Copyright 362964 CESSI.
- U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, R. Uthurusamy, editors, *Advances in Knowledge Discovery and Data Mining*, AAAI/MIT Press, 1996



- Fernández EA, Balzarini M. (2005) Biotecnología, Genómica y Minería de Datos. Enviado a Journal of Basic and Applied Genetics.
- Fernández EA, Balzarini M. (2005a) Improving Cluster Identification and Visualization in SOM like algorithms applied to Gene Expression Pattern Analysis. Enviado a BMC Bioinformatics.
- Fernández EA, (2005) Minería de Datos en Biotecnología. Aceptado en el Congreso Argentino de Bioingeniería 2005.
- Fernández EA. (2003) Minería de Datos y redes Neuronales para el análisis de la cinética de diálisis. Tesis de Doctorado, Universidad de Santiago de Compostela, España
- Fernández EA, Valtuille R, Willshaw P, Perazzo CA. (2001) Detection of Abnormality in the ECG Without Prior Knowledge by Using the Quantization Error of a Self-Organizing Map (SOM), tested on the European Ischemia Database, Medical & Biological Engineering & Computing. 39:330-337
- Cristóbal Fresno, Bioengineer; Andrea S Llera, Ph.D.; María R Girotti, Ph.D.; María P Valacco, Ph.D.; Juan A López; Osvaldo L Podhajcer, Ph.D.; Mónica G Balzarini, Ph.D.; Federico Prada, Ph.D.; Elmer A Fernández, The Multi-Reference Contrast Method: facilitating set enrichment analysis, Ph.D, Computers in Biology and Medicine. IN PRESS
- Fernandez EA, Souza Net EP, Abry P, Macchiavelli R, Balzarini M, Cuzin B, Baude C, Frutoso J, Gharib C Assessing erectile neurogenic dysfunction from heart rate variability through a Generalized Linear Mixed Model framework.. Computer Methods and Programs in Biomedicine doi:10.1016/j.cmpb.2009.11.001
- Fernández Elmer A, Girotti María R., López Juan A, Llera Andrea S., Podhajcer Osvaldo L, Cantet Rodolfo J. C. and Balzarini Mónica Improving 2D-DIGE protein expression analysis by two-stage linear mixed models: Assessing experimental effects in a melanoma cell study Bioinformatics
- Kohonen, T. (1997) Self-Organizing Maps (Springer, Berlin).
- Laciari E, Campos RJ, Brooks DH. Evaluación del daño miocárdico en enfermos chagásicos crónicos a partir del análisis del electrocardiograma de alta resolución... XV Congreso de Bioingeniería SABI 2005
- Podhajcer O. (2002). Estudios de genómica funcional mediante el uso de matrices de ADN (microarrays). Conferencia plenaria. Actas, XXXI Congreso Argentino de Genética, La Plata, 17-20 de setiembre de 2002.
- Schena, M., Shalon, D., Davis, R.W., Brown, P.O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science, 270, 467-470.

## Evaluación de la calidad de la Información extraída por wrappers, de un sitio web

VARGAS, Alejandro, SÁNCHEZ RIVERO David,  
VALDEZ Angel, BERNECHEA Miguel, CASTILLO Natalia & COLQUI Reinaldo  
Evaluación y Calidad Web (EvalCalWeb) / Facultad de Ingeniería /  
Universidad Nacional de Jujuy  
Av. Italia y Av. Martiarena / S. S. de Jujuy / Provincia de Jujuy  
Tel. 388-4221591

lavargas@fi.unju.edu.ar, vdsanchezrivero@fi.unju.edu.ar, a.roberto.valdez@gmail.com,  
pi77co@hotmail.com, natyc48@hotmail.com, mycorreo20@yahoo.com.ar

### Resumen

La complejidad creciente de la estructura y la cantidad de datos presentes en un sitio web determinado, torna necesaria la existencia de herramientas para la recuperación de información (RI), la cual se considera pertinente y adecuada, para su posterior análisis. En tal sentido los wrappers, programas para extracción de datos de la web, cumplen tal función, y pueden ser generados, mediante herramientas, en forma automática o desarrollados en forma artesanal (utilizando los lenguajes de programación python o perl, por ejemplo). Los wrappers son los encargados de transformar la información semi-estructurada (presente en un sitio web) en información estructurada, a través del lenguaje XML (eXtensible Markup Language).

El carácter dinámico de los sitios web posiblemente degrade la calidad de la información extraída por los wrappers, programas que trabajan en base a ciertos criterios, como ser color, posición en la página, fuente, tags, entre otros; los cuales pueden cambiar por el dinamismo propio del sitio.

Los resultados del presente trabajo, van a permitir establecer un criterio de evaluación y comparación de la calidad de los datos extraídos de un sitio web, a medida que este presenta cambio y/o modificaciones.

**Palabras clave:** Extracción de datos. Datos semi-estructurados. Calidad de datos. Medidas de calidad. Wrappers.

### Contexto

El proyecto parte de un trabajo anterior denominado “Criterios de búsqueda y extractores de datos aplicados en los portales de Bibliotecas Digitales BTC y BDBComp”, llevado a cabo por integrantes del grupo de trabajo, en conjunto con el Departamento de Ciencia da Computação - Universidade Federal de Minas Gerais – Brasil, en convenio con la Universidad Nacional de San Luis – Argentina, en el año 2010.

El proyecto de investigación se desarrolla en la Facultad de Ingeniería, de la Universidad Nacional de Jujuy, aprobado por Resolución del Concejo Superior 0167/12. El Proyecto posee Categoría “B”, otorgado por la Secretaría de

Ciencia y Técnica y Estudios Regionales, dependiente de la Universidad Nacional de Jujuy.

## Introducción

Internet es un gran repositorio de datos, con altas tasas de crecimiento día tras día, donde la información no mantiene un lineamiento de estructura de datos estándar, y no hay forma de manipularla y sobre todo se hace difícil llegar a tener un control sobre ella, en forma eficiente. Estos datos no pueden ser restringidos a través de un esquema tradicional y es conveniente contar con un formato de estructura muy flexible para el intercambio de datos entre bases de datos dispares [Buneman, 1997]. Los tipos de datos expuestos en la web, son llamados Semi-Estructurados, ya que tienen, como característica principal, el poseer una estructura implícita, donde la extracción de los mismos se hace en forma manual o en forma automática, a través de programas denominados wrappers. Existen patrones para la extracción de datos, que determinan los elementos relevantes a extraer, como ser el tipo de fuente, el color, la localización dentro de las páginas web, interpretación de tags y otras. Abiteboul [Abiteboul, 1997] determina un nuevo abordaje para la especificación de patrones de extracción de información de la web.

A pesar de la existencia de nuevas herramientas de búsqueda en la web, el acceso

a datos semi-estructurados sigue siendo inadecuado, ya que no se cuenta con el uso de estructuras de datos estándar, y por lo tanto diversas áreas, relacionadas con la temática de Base de Datos, enfrentan nuevos desafíos en la revisión de temas tradicionales como la integración de la información, modelos de datos y lenguajes de consulta aplicados en el contexto de la Web [Lima et al., 1999]. De acuerdo con los casos de estudio planteados por Florescu [Florescu et al., 1998], orientados al gerenciamiento de la información en la Web, en los cuales describe tres clases de tareas: Extracción e Integración de la Información; Modelos de datos y lenguajes de consulta para la Web; y Construcción y reestructuración de Sitios Web.

Los wrappers son programas capaces de reconocer y extraer objetos de interés dentro de páginas o sitios web. Laender [Laender et al., 2002] detalla una caracterización de los diferentes tipos de extractores de datos, según la técnica principal utilizada para la generación de wrappers. El trabajo de Chang [Chang et al., 2006] establece tres enfoques: la dificultad de un extractor de información, la técnica utilizada, y por último, el esfuerzo del usuario en el proceso de llevar dicho extractor a otro dominio.

Los wrappers son responsables de retornar los resultados extraídos en forma estructurada, y para realizarlo pueden utilizar el formato XML [Bray et al., 2000], pueden

ser programados en lenguaje python, perl, u otros lenguajes similares.

No solo se debe tener en cuenta la extracción de los datos de las páginas web y completar, de esa manera, el nuevo repositorio inicial de datos; sino también, y no es un tema menor, verificar la calidad de la información extraída, ya que la misma puede estar incorrecta, o bien que los datos no sean pertinentes de acuerdo a los temas de interés buscados [Golpher et al., 2001]. Las comparaciones para determinar la calidad de la extracción de datos pueden estar incorporadas dentro del wrapper, o pueden ser aplicadas en forma independiente sobre los resultados. Podemos mencionar la que utilizan heurísticas, o determinar la calidad a través de funciones estadísticas, de inducción, de posicionamiento de los objetos en las páginas origen y su estructura u otras.

### **Líneas de Investigación y Desarrollo**

Las líneas de investigación son:

- Aplicación de Técnicas de Recuperación de Información.
- Estudio de problemas relacionados a la búsqueda, extracción, consulta, modelado, almacenamiento, transformación e integración de datos disponibles en la web.

Se investigarán, en principio, diferentes criterios de evaluación de la información extraída por los wrappers, y se dispondrá de la aplicación de funciones y/o algoritmos que

sirvan como patrones de evaluación para los resultados obtenidos.

### **Resultados y Objetivos**

En el trabajo anterior “Criterios de búsqueda y extractores de datos aplicados en los portales de Bibliotecas Digitales BTC (Biblioteca de Trabajos Científicos) y BDBComp (Biblioteca Digital Brasileira de Computação)”, el grupo de trabajo desarrolló un prototipo de extractores de datos en tiempo real, el que extrae el 100% de los datos de las páginas de un repositorio de datos, además se efectuó la clasificación de la información y la visualización de los resultados, generando diferentes archivos con formato XML.

El equipo de investigación se encuentra desarrollando actividades de investigación y desarrollo en el diagnóstico, análisis y descripción de las herramientas que extraen datos de la web. Como así también se está trabajando en la búsqueda, descripción e instalación de herramientas, basadas en software libre, para la extracción de datos y así poder generar los wrappers, sobre un sitio determinado.

Finalmente se procura establecer un estado de la cuestión y analizar en detalle los procesos de ejecución de los wrappers y así poder comparar criterios de evaluación de datos extraídos de un sitio web.



## Formación de Recursos Humanos

El proyecto de investigación se encuentra conformado por dos docentes investigadores categorizados, tres alumnos y un egresado de la carrera de Ingeniería Informática, en proceso de formación, abocados en tareas de investigación y programación.

En el presente año, se realizaron diferentes reuniones con alumnos y egresados interesados en incorporarse al proyecto y se establecieron líneas de investigación para generar Proyectos Finales, en la carrera de Ingeniería en Informática de la Facultad de Ingeniería de la Universidad Nacional de Jujuy.

Por otra parte el desarrollo de las tareas de investigación ha generado, en el año 2012, un anteproyecto de tesis en la Maestría en Ingeniería de Software, de la Universidad Nacional de San Luis.

Se prevee una mayor interacción y colaboración con el Laboratorio de "Base de Datos" de la Universidad Federal de Minas Gerais, dirigida por el Prof. Dr. Alberto Laender, de la Universidade Federal de Minas Gerais, Brasil

## Referencias

- [Buneman, 1997] Buneman, Peter. Semistructured Data. In Proc. Of the ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems (PODS), pages 117-121, Tucson, Arizona, 1997.
- [Abiteboul, 1997] Abiteboul, Serge, Querying semi-structured data. In Proc. Of the Int. Conf. On Database Theory (ICDT), Delphi, Greece, 1997.
- [Lima et al., 1999] Lima, F; Casanova, M.A. & Melo, R. N., Revisitando Técnicas de Bancos de Datos no contexto da Web, 1999.
- [Florescu et al., 1998] Florescu, D., Levy, A. & Mendelzon, A., "Database Techniques for the World-Wide Web: A Survey", SIGMOD Record, vol. 27, n. 3, Setembro 1998.
- [Laender et al., 2002] Laender, A.H.; Ribeiro-Neto, A.S.; Da Silva, A.S & Teixeira, J.S. A Brief Survey of Web Data Extraction Tools. SIGMOD Record, 31(2): 84-93, 2002.
- [Chang et al., 2006] Chang Chia-Hui; Kayed, M.; Girgis, M.R. & Shaalan, K., A survey of Web Information Extraction Systems. Journal IEEE Transactions on Knowledge and Data Engineering. Volume 18(10):1411-1428, 2006.
- [Bray et al., 2000] Bray T., Paoli J, Sperberger C.M., and Maler E.; Extensible Markup Language (XML) version 1.0 (second edition). Technical Report REC-xml-20001006, World Wide Web Consortium. W3C, October 2000.
- [Golpher et al., 2001] Golpher, P.B.; da Silva, A.S.; Laender, A.H.F.; Ribeiro-Neto, B.A., Bootstrapping for Example-Based Data Extraction. In Proceeding of the Tenth ACM International Conference on Information and Knowledge Management, pages 371-378, Atlanta, Georgia, 2001.

## MINERÍA DE DATOS APLICADA A LA CONSERVACIÓN *EX SITU* DE RECURSOS FITOGENÉTICOS DE SAN JUAN.

Karina Fernández<sup>1</sup>, Carola Meglioli<sup>2</sup>, Raúl Klenzi<sup>1</sup>

<sup>1</sup>Departamento de Informática. Facultad de Ciencias Exactas, Físicas y Naturales. Universidad Nacional de San Juan. Avda Ignacio de la Roza y Meglioli. Rivadavia.

<sup>2</sup>Área Conservación de Recursos Fitogenéticos. Instituto de Investigación y Desarrollo Agroindustrial Hortícola Semillero. Coll 3671 (oeste). Rivadavia (5407). San Juan. karinaferh@gmail.com; caromeglioli@yahoo.com.ar; rauloscarklenzi@gmail.com

### Resumen:

La digitalización y posterior análisis de datos biológicos es una actividad creciente en la actualidad. En la presente propuesta, se proyecta realizar tareas de agrupamiento y clasificación sobre los datos pertenecientes al área de conservación de recursos fitogenéticos del Instituto de Investigación y Desarrollo Agroindustrial Hortícola Semillero (INSEMI), de la provincia de San Juan. Se trabajará con una herramienta específica de Minería de Datos con conectividad al banco de datos sobre los recursos fitogenéticos conservados, donde se evaluarán diferentes algoritmos de segmentación, clasificación y visualización, que orienten la toma de decisiones con respecto a: a) la planificación de las campañas de recolección de recursos fitogenéticos, b) el manejo interno de las colecciones en cuanto al fortalecimiento de las mismas, y c) su manipulación posterior.

### Contexto

La actividad de conservación *ex situ* de recursos fitogenéticos en la provincia de San Juan, está siendo fortalecida mediante la ejecución del Proyecto "Desarrollo competitivo del sector semillero de la provincia de San Juan", el cual es financiado por el Banco Mundial y

con aportes locales en el marco del Ministerio de Producción y Desarrollo Económico del Gobierno de la Provincia, los cuales son administrados por el Programa de Servicios Agrícolas Provinciales (PROSAP/BIRF 7597-AR). Este proyecto se lleva a cabo con centro en el Instituto de Investigación y Desarrollo Agroindustrial Hortícola Semillero (INSEMI).

Por otro lado, en el ámbito de la Universidad Nacional de San Juan (UNSJ), se encuentra en ejecución el Proyecto "Minería de Datos en la Determinación de Patrones de Uso y Perfiles de Usuarios (21/E889)", financiado por CICITCA-UNSJ trienio 2011-2013, con el cual y atendiendo a la temática abordada, existe la posibilidad de un trabajo común de fuerte integración. En este contexto se cuenta con el aporte de la beca nacional TIC` s otorgada a la Alumna Karina Fernández, periodo 2012/2013 para el estudio de minería de datos en la segmentación y clasificación de un banco de germoplasma.

En el corto plazo se contará además con aportes del Ministerio de Ciencia, Tecnología e Innovación Productiva (MINCYT), para el fortalecimiento de las bases de datos y para la formación de

recursos humanos, ya que el INSEMI ha sido adherido al Sistema Nacional de Datos Biológicos (SNDB – Res 05/12)

### Introducción

Se entiende por Recursos Genéticos, la variabilidad genética acumulada en todos los organismos vivos a lo largo de millones de años de evolución (REGENSUR, 2007). Entonces los Recursos Fitogenéticos se definen como cualquier material genético de origen vegetal de valor real o potencial para la alimentación y la agricultura, y que constituyen la materia prima utilizada por los agricultores, por consiguiente resulta fundamental su conservación y el mantenimiento para una producción agrícola sostenible (FAO, 1996; UICN, 1980).

La documentación de la información recopilada y generada es fundamental para la toma de decisiones sobre el manejo de las colecciones en un centro de conservación de recursos y sobre su valor de uso (Engels et al., 2007). Además, el valor del germoplasma conservado aumenta en la medida que se conoce, de ahí la importancia de una adecuada documentación de los datos recopilados y generados (Rao et al., 2007). El sistema de documentación se utiliza para la recuperación, el almacenamiento, el mantenimiento o la actualización, procesamiento y análisis, e intercambio de datos con otros centros de conservación (Rivera-Gutiérrez et al., 2003).

El éxito de la revolución digital y el crecimiento de Internet aseguran que grandes volúmenes de datos de alta dimensión, están disponibles en todo lo

que nos rodea. Esta información se mezcla a menudo, con la participación de diferentes tipos de datos tales como texto, imagen, audio, voz, hipertexto, gráficos y componentes de vídeo entremezcladas unas con otras. Sin embargo, a menudo la mayor parte de estos datos no son de mucho interés para la mayoría de los usuarios. Descubrimiento de Conocimiento en Bases de Datos (KDD) es un análisis automático y exploratorio que permite el modelado de depósitos de datos de gran tamaño. KDD es el proceso organizado de identificar patrones válidos, novedosos, útiles y comprensibles a partir de extensos y complejos conjuntos. Data Mining (DM) es el núcleo del proceso de KDD, que implica la inferencia de los algoritmos que analizan y modelan los datos, y permite descubrir patrones previamente desconocidos y de interés para los usuarios. Se trata de un área de investigación y desarrollo interdisciplinaria que abarca diversos dominios, y lejos de estar saturada, se amplía con nuevas técnicas y orientaciones (Mucherino A., y otros. 2009; O. Maimon O. Rokach L. 2010; Witten, Frank, Hall, 2011).

El procesamiento de información basada en minería de datos, está orientada a análisis de atributos pertenecientes al área de conservación *ex situ* de recursos fitogenéticos de San Juan. Se mantienen colecciones de herbario y de semillas tanto de especies nativas como también de cultivos tradicionales, importantes para la provincia. Este material genético responsable de las características de una planta que se transmite de una generación a la siguiente, para el futuro

beneficio de la humanidad y del ambiente.

En este contexto, contar con datos digitalizados y una base de datos actualizada, ágil y de fácil acceso, que contenga información georreferenciada y vinculante sobre los recursos fitogenéticos de San Juan, resulta además, una herramienta de alto valor, dado que no se han encontrado aún en la zona, instituciones que generen, mantengan y actualicen en forma continua datos biológicos.

Se plantea aplicar minería de datos a los diferentes registros colectados en el área de conservación de recursos fitogenéticos de San Juan. Las tareas principales radicarán en el uso de algoritmos específicos que permitan tareas de segmentación y clasificación de datos (Figuerola et al., 2005) provenientes de las colecciones de semillas de especies correspondientes a la geografía de San Juan. Para ello se hará uso de la herramienta de software Rapid Miner cuya última versión disponible es la 5.3.005. Se trata de un entorno de aplicación de algoritmos de aprendizaje de máquina y minería de datos, fácil de instalar y de ejecutar en cualquier plataforma y sistema operativo. Allí se pueden aplicar todos los pasos involucrados en la minería de datos desde el pre-procesamiento hasta la visualización de resultados al evaluar diferentes estrategias de segmentación, clasificación y reglas de asociación mediante una interfaz amigable (North 2012) y que se ofrece bajo una licencia AGPL versión 3.0 (software libre).

Se espera segmentar y clasificar datos biológicos provenientes de los Recursos

Fitogenéticos conservados en San Juan para obtener agrupaciones requeridas para la planificación, manejo, regeneración y transferencia del germoplasma conservado.

### **Líneas de investigación y desarrollo**

El equipo de trabajo desarrolla dos grandes líneas de trabajo: a) Digitalización y análisis de datos en el ámbito de la conservación de recursos fitogenéticos locales, b) Minería de Datos.

### **Resultados y Objetivos**

Los datos sobre los recursos fitogenéticos conservados se encontraban 100% almacenados en planillas de recolección y cuadernos de campo, por lo que la primer tarea correspondió a la digitalización de los mismos para permitir posteriormente la migración de los mismos a una base de datos adecuada y poder realizar los análisis propuestos.

Hasta la fecha se ha logrado un elevado porcentaje de la sistematización y posterior digitalización de la información correspondiente a las colecciones de semillas de la flora nativa, como así también de las colecciones de semillas de los cultivos tradicionales. Estos registros son caracterizados por aproximadamente 80 atributos multivaluados. Estos datos han sido categorizados en planillas Excel, a partir de las cuales serán trasladados a la base de datos que se establecerá para el área de conservación de recursos fitogenéticos del INSEMI y desde donde la herramienta de aplicación RM tomará los datos, previa compatibilización en los



tipos de datos asociados a los atributos, que permita a su vez, viabilizar la aplicación de diferentes algoritmos de segmentación y clasificación. Esto último está iniciándose y los resultados que se obtengan de estos análisis definirán los próximos pasos a seguir dentro del ámbito de la minería de datos.

Row No.	Elements	Clusters	Pages	Fecha recolección	Responsable	Otro sistema	Cultivo	Estado	Nombre ridge	Familia	Espesor	Tipo de suelo
1	1	1	7	20060111 0:00:00 ART	Juan Scaglia	Cristian Piedra	CHESA-PNICEAE	Casapalma pangaparana	Chalabera	CHESA-PNICEAE	Scaglia 29	Italo
2	2	1	7	27620111 0:00:00 ART	Matin Hadad	Juan Scaglia	CHESA-PNICEAE	Cerdum pracoce (R)P14	Brea	CHESA-PNICEAE	Hadad 47	Italo
3	3	1	7	30120111 0:00:00 ART	Juan Scaglia	?	CHESA-PNICEAE	Zucupala pastosa Cm	Jarilla muato	CHESA-PNICEAE	Scaglia 79	Italo
4	4	4	22	18010112 0:00:00 ART	Juan Scaglia	?	CHESA-PNICEAE	Cerdum pracoce soto j	Brea	CHESA-PNICEAE	Scaglia 47	Italo
5	5	1	34	04010111 0:00:00 ART	Matin Hadad	Juan Scaglia	CHESA-PNICEAE	Cerdum pracoce	Brea	CHESA-PNICEAE	Hadad 27	Italo
6	6	1	7	20010112 0:00:00 ART	Juan Scaglia	Gerardo de la Vega	CHESA-PNICEAE	Cerdum pracoce var. Pral	Brea	CHESA-PNICEAE	Scaglia 44	Italo
7	1	1	7	20050111 0:00:00 ART	Matin Hadad	Juan Scaglia	PAPULONICEAE	Ramonea grisea	Chica	PAPULONICEAE	Hadad 21	Italo
8	2	1	7	21040111 0:00:00 ART	Juan Scaglia	?	PAPULONICEAE	Gedifera bicolorata	Challar	PAPULONICEAE	Scaglia 9	Italo
9	1	1	128	21040111 0:00:00 ART	Matin Hadad	Uta Korte	IMBOSACEAE	Prosope diptera Phil	Higadoo temar	IMBOSACEAE	Hadad 14	Italo
10	2	1	7	18100111 0:00:00 ART	Juan Scaglia	?	IMBOSACEAE	Mitrospora carinata (P) Lam	?	IMBOSACEAE	Scaglia 34	Italo
11	3	1	1	21030111 0:00:00 ART	Matin Hadad	?	IMBOSACEAE	Acacia atomifera	Esposito	IMBOSACEAE	Hadad 5	Italo
12	4	3	60	19050111 0:00:00 ART	Juan Scaglia	Mariano Perez	IMBOSACEAE	Acacia vico	Vico	IMBOSACEAE	Scaglia 69	Italo
13	5	1	7	27620111 0:00:00 ART	Juan Scaglia	Rosendo Brucala, José C	IMBOSACEAE	Acacia atomifera	Esposito negro	IMBOSACEAE	Scaglia 14	Italo
14	5	1	4	24050111 0:00:00 ART	Matin Hadad	Tarino Ribas	IMBOSACEAE	Acacia atomifera Benth	Armo negro	IMBOSACEAE	Hadad 7	Italo
15	7	4	18	18010111 0:00:00 ART	Juan Scaglia	Rosendo Brucala, Carlos	IMBOSACEAE	Prosope flexuosa	Algarrobo negro	IMBOSACEAE	Scaglia 63	Italo
16	8	4	18	12050111 0:00:00 ART	Juan Scaglia	Rosendo Brucala	IMBOSACEAE	Prosope flexuosa	Algarrobo negro	IMBOSACEAE	Scaglia 77	Italo
17	9	4	15	14010111 0:00:00 ART	Juan Scaglia	Rosendo Brucala, Javier	IMBOSACEAE	Prosope flexuosa	Algarrobo blanco	IMBOSACEAE	Scaglia 73	Italo
18	10	1	8	22040111 0:00:00 ART	Matin Hadad	Laura Iglesias, NEREA AH	IMBOSACEAE	Acacia fucicola	Garabato	FABACEAE	Hadad 17	Italo
19	11	1	7	24050111 0:00:00 ART	Matin Hadad	Juan Scaglia	IMBOSACEAE	Acacia fucicola	Garabato	IMBOSACEAE	Hadad 24	Italo
20	12	1	18	08050111 0:00:00 ART	Matin Hadad	?	IMBOSACEAE	Acacia sica	Vico	IMBOSACEAE	Hadad 20	Italo
21	13	1	8	02020111 0:00:00 ART	Matin Hadad	?	IMBOSACEAE	Acacia fucicola	Garabato	IMBOSACEAE	Hadad 16	Italo
22	14	1	7	24050111 0:00:00 ART	Matin Hadad	Juan Scaglia	IMBOSACEAE	Acacia senna (Giles et Noel) Arnoy	Armo	IMBOSACEAE	Hadad 11	Italo
23	15	1	7	01060111 0:00:00 ART	Juan Scaglia	?	IMBOSACEAE	Acacia swain	Esposito	IMBOSACEAE	?	Italo

Figura 1

En la Figura1 se puede apreciar la lectura realizada desde RM de una porción de la fuente de datos original excell. En esta primer tanda de registros cargados se observa la existencia de 80 atributos entre los que figuran datos asociados al personal de la campaña de recolección, lugar físico de ubicación de los datos originales, que si bien pueden ser de suma importancia para la determinación de responsabilidades en el INSEMI puede tener escasa importancia a la hora comenzar con las tareas de segmentación y clasificación. Por lo expuesto anteriormente, el grupo de trabajo está en plena tarea de definición de atributos relevantes inherentes a la etapa de preprocesamiento de DM. Se proseguirá con el relevamiento y el análisis de las herramientas de software y necesidades de hardware asociado, y con la generación de conocimiento relevante de los datos que favorezca la

toma de decisiones para el manejo de las colecciones conservadas mediante la aplicación de herramientas específicas de minería de datos.

**Formación de Recursos Humanos**

El equipo de trabajo está formado por investigadores del Departamento de Informática de la Universidad Nacional de San Juan y del instituto de Investigación y Desarrollo Agroindustrial Hortícola Semillero (INSEMI), becarios y adscriptos a los proyectos. También se cuenta con apoyo de investigadores de INTA.

**Tesis de posgrado en curso:**

“Acondicionamiento de semillas de Poaceae nativas para su aprovechamiento en la recuperación de suelos degradados en zonas áridas”. Doctorado en Agronomía. Facultad de Ciencias Agrarias, Universidad Nacional de Cuyo (Mendoza). Alumna: Lic. Carola Meglioli. Director: Dr. Carlos Parera.

**Tesis de grado en curso:**

Tras la carga de la totalidad de los registros existentes en papel a formato digital se llevará adelante el trabajo final de grado: “Minería de datos en la segmentación y clasificación de un banco de germoplasma”. Licenciatura en Sistemas de Información. Departamento de Informática. Facultad de Ciencias Exactas, Físicas y Naturales, Universidad Nacional de San Juan. Alumna: Karina Fernández. Director: Mg. Raúl Klenzi, Codirector: Lic. Carola Meglioli.

Se pretende además, y en el marco de un trabajo integrador de titulación intermedia de la carrera licenciatura en ciencias de la computación, desarrollar una

herramienta de software e interfaz adecuada, que permita la carga directa de la información durante la campaña de recolección, en dispositivos tipo tablets o celular avanzado a cargo del personal afectado, logrando una mayor agilidad al momento de carga y procesamiento de datos

### Referencias

- ° Engels, J.M.M. y Visser, L. (eds.). 2007. Guía para el manejo eficaz de un banco de germoplasma. Manuales para Bancos de Germoplasma No. 6. Bioversity International, Roma, Italia. ISBN 978-92-9043-767-3
- ° FAO. 1996. Plan de Acción Mundial para la Conservación y la utilización sostenible de los Recursos Fitogenéticos para la Alimentación y la Agricultura y la Declaración de Leipzig. Cuarta Conferencia Técnica Internacional sobre Recursos Fitogenéticos, Leipzig, Alemania, 17-23 de junio, 64 pp.
- ° Maimon, O. Rokach L. 2010. "Data Mining and Knowledge Discovery Handbook" –Springer
- ° Mucherino A., Papajorgji P., Pardalos P. 2009. "Data Mining in Agriculture" (Springer Optimization and Its Applications Volume 34)
- ° North M. 2012. "Data Mining for the Masses" ISBN: 0615684378. A Global Text Project Book.
- ° Peña, D. 2002. Análisis de datos multivariantes. Ed. Mc. Graw Hill, España.
- ° Rao, N.K., J. Hanson, M.E. Dulloo, K. Ghosh, D. Novell y M. Larinde. 2007. Manual para el manejo de semillas en bancos de germoplasma. Manuales para Bancos de Germoplasma No. 8. Bioversity International, Roma, Italia. ISBN 978-92-9043-757-4.
- ° REGENSUR/PROCISUR. 2007. Acceso a los recursos genéticos: estado de situación en los países del cono sur.
- ° Rivera-Gutiérrez H. F., Suárez-Mayorga A. M., Varón-Londoño A. 2003. Estándar para la documentación de metadatos de conjuntos de datos relacionados con biodiversidad, versión (electrónica). Instituto de Investigación de Recursos Biológicos Alexander von Humboldt, Bogotá, Colombia, 61 p.
- ° Unión Mundial para la Conservación de la Naturaleza (UICN). 1980. Estrategia Mundial para la Conservación.
- ° Witten I.; Frank E., Hall M. 2011 "Data Mining Practical Machine Learning Tools and Techniques", Third Edition (The Morgan Kaufmann Series in Data Management Systems) Morgan Kaufmann

# Indexando Texto en Memoria Secundaria

**Gonzalo Navarro**

Departamento de Ciencias de la Computación  
Universidad de Chile, Chile  
gnavarro@dcc.uchile

**Nieves Rodríguez Brisaboa**

Facultad de Informática  
Universidad de A Coruña, España  
brisaboa@udc.es

**Norma Herrera, Carina Ruano, Darío Ruano, Ana Villegas, Susana Esquivel**

Departamento de Informática  
Universidad Nacional de San Luis, Argentina  
{nherrera, cmruano, dmruano, anaville, esquivel}@unsl.edu.ar

## Resumen

La próxima generación de administradores de bases de datos deberá ser capaz de indexar datos no estructurados (datos multimedia) y responder consultas sobre estos datos con tanta eficiencia como actualmente responden consultas de búsqueda exacta sobre bases de datos relacionales. Si bien existen numerosas técnicas de indexación diseñadas para esta problemática, mejorar la eficiencia de las mismas es de vital importancia. Nuestro ámbito de investigación es el estudio de índices eficientes para datos no estructurados.

## 1. Contexto

El presente trabajo se desarrolla en el ámbito de la línea Técnicas de Indexación para Datos no Estructurados del Proyecto Tecnologías Avanzadas de Bases de Datos (22/F014), cuyo objetivo es realizar investigación básica en problemas relacionados al manejo y recuperación eficiente de información no tradicional.

## 2. Introducción

Las bases de datos actuales han incluido la capacidad de almacenar datos no estructurados tales como imágenes, sonido, texto, video, datos geométricos, etc. La problemática de almacenamiento y búsqueda en estos tipos de base de datos difiere de las bases de datos clásicas, dado que no es posible organizar los datos en registros y campos, y aun cuando pudiera hacerse, la búsqueda exacta carece de interés. Es en este contexto donde surgen nuevos modelos de bases de datos capaces de cubrir las necesidades de almacenamiento y búsqueda de estas aplicaciones. Nuestro interés se basa en el diseño de índices para estas nuevas bases de datos, centrándonos en bases de datos de texto.

Un base de datos de texto es un sistema que mantiene una colección grande de texto, y provee acceso rápido y seguro al mismo. Sin pérdida de generalidad, asumiremos que la base de datos de texto es un único texto  $T = t_1, \dots, t_n$  posiblemente almacenado en varios archivos. Asumiremos que  $T$  está formado por símbolos de un alfabeto  $\Sigma$  de tamaño  $\sigma$ , donde  $t_n = \$ \notin \Sigma$  es un símbolo menor en

orden lexicográfico que cualquier otro símbolo de  $\Sigma$ , denotaremos con  $T_{i,j}$  a la secuencia  $t_i, \dots, t_j$ , con  $1 \leq i \leq j \leq n$ . Un sufijo de  $T$  es cualquier string de la forma  $T_{i,n} = t_i, \dots, t_n$  y un prefijo de  $T$  es cualquier string de la forma  $T_{1,i} = t_1, \dots, t_i$  con  $i = 1..n$ . Un patrón de búsqueda  $P = p_1 \dots p_m$  es cualquier string sobre el alfabeto  $\Sigma$ .

Construir un índice sobre  $T$  tiene sentido cuando  $T$  es grande, cuando las búsquedas son más frecuentes que las modificaciones (de manera tal que los costos de construcción se vean amortizados) y cuando hay suficiente espacio como para contener el índice. Un índice debe dar soporte a dos operaciones básicas: *count*, que consiste en contar el número de ocurrencias de un patrón  $P$  en un texto  $T$  y *locate*, que consiste en ubicar todas las posiciones del texto  $T$  donde el patrón de búsqueda  $P$  ocurre.

Entre los índices más populares para texto encontramos el *arreglo de sufijos*, el *trie de sufijos* y el *árbol de sufijos*. Estos índices se construyen basándose en la observación de que un patrón  $P$  ocurre en el texto si es prefijo de algún sufijo del texto.

**Arreglo de sufijos:** un arreglo de sufijos  $A[1, n]$  es una permutación de los números  $1, 2, \dots, n$  tal que  $T_{A[i],n} \prec T_{A[i+1],n}$ , donde  $\prec$  es la relación de orden lexicográfico [12]. Buscar un patrón  $P$  en  $T$  equivale a buscar todos los sufijos de los cuales  $P$  es prefijo, los cuales estarán en posiciones consecutivas de  $A$ .

**Trie de Sufijos:** un trie de sufijos es un *Trie* construido sobre el conjunto de todos los sufijos del texto, en el cual cada hoja mantiene el índice del sufijo que esa hoja representa [18]. El trie de sufijos resuelve eficientemente búsquedas de patrones en un texto basándose en la observación anterior y utilizando la eficiencia del Trie para resolver búsquedas de prefijos en un conjunto de string.

**Árbol de sufijos:** un árbol de sufijos es un Pat-Tree [4] construido sobre el conjunto de todos los sufijos de  $T$  codificados sobre alfabeto binario. Cada nodo interno mantiene el número de bit del patrón que corresponde utilizar en ese punto para direccionar la búsqueda y las hojas contienen una posición del texto que representa al sufijo que se inicia en dicha posición [18].

Mientras que en bases de datos tradicionales los índices ocupan menos espacio que el conjunto de datos indexado, en las bases de datos de texto el índice ocupa más espacio que el texto, pudiendo necesitar de 4 a 20 veces el tamaño del mismo [4, 12]. Una alternativa para reducir el espacio ocupado por el índice es buscar una representación compacta del mismo, manteniendo las facilidades de navegación sobre la estructura. Pero en grandes colecciones de texto, el índice aún comprimido suele ser demasiado grande como para residir en memoria principal. Es por ello que el desarrollo de índices comprimidos en memoria secundaria es un tema de creciente interés. Entre los índices para texto en memoria secundaria más relevantes encontramos:

**String B-Tree** [3]: consiste básicamente en un B-Tree en el que cada nodo es representado como un Pat-Tree [4]. Este índice requiere tanto para *count* como para *locate*  $O(\frac{m+occ}{b} + \log_b n)$  accesos a memoria secundaria en el peor caso, donde *occ* es la cantidad de ocurrencias de  $P$  en  $T$  y  $b$  es el tamaño de páginas de disco medido en enteros. No es un índice comprimido y su versión estática requiere en espacio de 5 a 6 veces el tamaño del texto más el texto.

**Compact Pat Tree** [2]: representa un árbol de sufijos en memoria secundaria y en forma compacta. Si bien no existen desarrollos teóricos que garanticen el espacio ocupado por el índice y el tiempo insumido en resolver la búsqueda, en la práctica el índice tiene un muy



buen desempeño requiriendo de 2 a 3 accesos a memoria secundaria tanto para *count* como para *locate*, y ocupando entre 4 y 5 veces el tamaño del texto más el texto.

**Disk-based Compressed Suffix Array** [11]: adapta el autoíndice comprimido para memoria principal presentado en [17] a memoria secundaria. Requiere  $n(H_0 + O(\log \log \sigma))$  bits de espacio (donde  $H_k \leq \log \sigma$  es la entropía de orden  $k$  de  $T$ ). Para la operación *count* realiza  $O(m \log_b n)$  accesos. Para la operación *locate* realiza  $O(\log n)$  accesos lo cual es demasiado costoso.

**Disk-based LZ-Index** [1]: adapta a memoria secundaria el autoíndice comprimido para memoria principal presentado en [15]. Utiliza  $8 n H_k(T) + o(n \log \sigma)$  bits; los autores no proveen límites teóricos para la complejidad temporal, pero en la práctica es muy competitivo.

### 3. Líneas de Investigación

Nuestra principal línea de trabajo es el estudio de algoritmos de indexación sobre bases de datos no estructurados, centrándonos principalmente en el diseño de índices para bases de datos textuales. Describimos a continuación las líneas de investigación que actualmente estamos desarrollando.

#### 3.1. Locally Compressed SA

El Locally Compressed Suffix Array (LC-SA) [5] es una técnica para compresión de arreglos de sufijos. Un arreglo de sufijos  $A$  construido sobre un texto  $T$  de longitud  $n$  es compresible si  $T$  lo es. La entropía de orden  $k$  de  $T$  ( $H_k$ ) se refleja en  $A$  formando secuencias largas  $A[i, i + l]$ , denominadas *pseudo-repeticiones* que aparecen en otro lugar  $A[j, j + l]$  con todos los valores incremen-

tados en uno, es decir:  $A[j + s] = A[i + s] + 1$  con  $0 \leq s \leq l$ .

Si particionamos  $A$  en *pseudo-repeticiones* de tamaño maximal, el número de partes que obtendríamos sería a lo más  $nH_k + \sigma^k$ , para algún  $k$  [16]. Esta propiedad ha sido usada por varios autores para comprimir un arreglo de sufijos  $A$  [9, 10]. El LCSA es una técnica para compresión de arreglos de sufijos que consiste en convertir las *pseudo-repeticiones* en repeticiones reales, que luego son factorizadas usando Re-Pair [8].

Re-Pair es un compresor basado en diccionario que permite una rápida descompresión local usando solamente el diccionario. La técnica consiste en encontrar el par de símbolos más frecuente y reemplazarlo con un nuevo símbolo. Podemos resumir el proceso realizado por Re-Pair en los siguientes pasos:

1. Encontrar el par de símbolos  $ab$  más frecuente en  $T$ .
2. Crear un nuevo símbolo  $s$  mayor que cualquier símbolo en  $T$  y agregar al diccionario  $R$  la regla  $s \rightarrow ab$ .
3. Reemplazar en  $T$  toda ocurrencia de  $ab$  por  $s$ .
4. Repetir hasta que todos los pares tengan frecuencia 1.

El resultado de este algoritmo de compresión es el diccionario de reglas  $R$  más una secuencia de símbolos  $C$  (símbolos originales y nuevos) que es el texto  $T$  ya comprimido. Notar que podemos representar  $R$  en un vector de pares de manera tal que la regla  $s \rightarrow ab$  esté representada en  $R[s - \sigma] = a : b$ .

Cualquier segmento de  $C$  puede ser rápida y fácilmente descomprimido de la siguiente manera: para descomprimir  $C[i]$  primero verificamos el valor de  $C[i]$ . Si  $C[i] < \sigma$ , entonces es un símbolo original de  $T$ , por lo tanto no corresponde hacer nada más. Caso contrario obtenemos los símbolos que corresponden a  $C[i]$  en  $R[C[i] - \sigma]$  y los expandimos recursivamente. Esto permite reproducir  $u$  caracteres de  $T$  en  $O(u)$  unidades de tiempo.

### 3.2. LCSA + CPT

Como mencionamos anteriormente, el Compact Pat Tree (CPT) consiste en representar un árbol de sufijos en memoria secundaria y en forma compacta.

En [7] hemos presentado una modificación en el diseño del CPT que permite mantener la representación del arreglo de sufijos subyacente en el CPT separada de la representación del árbol propiamente dicho. Esto nos permite reducir el espacio total requerido por el índice comprimiendo dicho arreglo de sufijos. Para ello estamos trabajando en la incorporación de la técnica LCSA en el CPT.

Como primer paso se deben diseñar los algoritmos de construcción en memoria secundaria. Para lograr algoritmos eficientes en memoria secundaria es necesario que los mismos tengan alta localidad de referencia. El algoritmo de construcción de LCSA tiene una muy baja localidad de referencia dado que recorre  $A$  usando la función  $\Psi$ , donde  $\Psi[i] = j$  si  $A[j] = A[i] + 1$ . El algoritmo de construcción del LCSA en memoria secundaria fue propuesto en [6]. Allí se presenta el diseño de dicho algoritmo y el desarrollo de complejidad del mismo, pero sin realizar la implementación y la evaluación empírica del algoritmo. No hay aún resultados experimentales sobre cómo se comporta esta implementación, por lo cual es posible que aún requiera de ajustes para lograr un rendimiento aceptable. En este momento hemos finalizado la implementación del algoritmo de construcción del LCSA en memoria secundaria encontrándonos en la etapa de evaluación empírica del mismo.

### 3.3. String B-Tree

El String B-Tree (SBT) [3] es un índice dinámico para búsquedas de patrones en memoria secundaria. Básicamente consiste en una combinación de dos estructuras: el B-Tree y el Pat-Tree [4]. No es un índice comprimido y su versión estática requiere en espacio de 5 a

6 veces el tamaño del texto.

Sobre el SBT el objetivo principal es lograr una reducción en el espacio utilizado por el mismo manteniendo los costos de búsquedas de la versión original. Para ello, se han diseñado e implementado dos variantes que consisten en modificar la representación de cada nodo del árbol B subyacente. Una de las variantes consiste en usar un Pat-Tree como originalmente proponen los autores para los nodos pero usando representación de paréntesis para el mismo [13]. La otra variante consiste en representar cada nodo con la representación de arreglos propuesta en [14] que ofrece las mismas funcionalidades que un Pat-Tree pero que tienen las características necesarias como para permitir una posterior compresión de los mismos. Nos encontramos en la etapa de evaluación experimental de estas versiones para su posterior ajuste de parámetros.

## 4. Resultados Esperados

Se espera obtener índices en memoria secundaria eficientes, tanto en espacio como en tiempo, para el procesamiento de consultas en bases de datos textuales. Los mismos serán evaluados tanto analíticamente como empíricamente.

## 5. Recursos Humanos

El trabajo desarrollado en esta línea forma parte del desarrollo de un Trabajo Final de la Licenciatura, dos Tesis de Maestría y una Tesis de Doctorado, todas ellas en el ámbito de Ciencias de la Computación en la Universidad Nacional de San Luis.

## Referencias

- [1] D. Arroyuelo and G. Navarro. A lempel-ziv text index on secondary storage. In

- Proc. 18th Annual Symposium on Combinatorial Pattern Matching (CPM)*, LNCS 4580, pages 83–94, 2007.
- [2] D. Clark and I. Munro. Efficient suffix tree on secondary storage. In *Proc. 7th ACM-SIAM Symposium on Discrete Algorithms*, pages 383–391, 1996.
- [3] P. Ferragina and R. Grossi. The string B-tree: a new data structure for string search in external memory and its applications. *Journal of the ACM*, 46(2):236–280, 1999.
- [4] G. H. Gonnet, R. Baeza-Yates, and T. Snider. *New indices for text: PAT trees and PAT arrays*, pages 66–82. Prentice Hall, New Jersey, 1992.
- [5] R. González and G. Navarro. Compressed text indexes with fast locate. In *Proc. 18th Annual Symposium on Combinatorial Pattern Matching (CPM)*, LNCS 4580, pages 216–227, 2007.
- [6] R. González and G. Navarro. A compressed text index on secondary memory. *Journal of Combinatorial Mathematics and Combinatorial Computing*, 71:127–154, 2009.
- [7] N. Herrera and G. Navarro. Árboles de sufijos comprimidos en memoria secundaria. In *Proc. XXXV Latin American Conference on Informatics (CLEI)*, Pelotas, Brazil, 2009.
- [8] N. Jesper Larsson and Alistair Moffat. Offline dictionary-based compression. In *DCC '99: Proceedings of the Conference on Data Compression*, page 296, Washington, DC, USA, 1999. IEEE Computer Society.
- [9] V. Mäkinen. Compact suffix array: a space-efficient full-text index. *Fundam. Inf.*, 56(1,2):191–210, 2002.
- [10] V. Mäkinen and G. Navarro. Succinct suffix arrays based on run-length encoding. *Nordic J. of Computing*, 12(1):40–66, 2005.
- [11] V. Mäkinen, G. Navarro, and K. Sadakane. Advantages of backward searching - efficient secondary memory and distributed implementation of compressed suffix arrays. In *Proc. 15th Annual International Symposium on Algorithms and Computation (ISAAC)*, LNCS 3341, pages 681–692. Springer, 2004.
- [12] U. Manber and G. Myers. Suffix arrays: A new method for on-line string searches. *SIAM Journal of Computing*, 22(5):935–948, 1993.
- [13] J. Ian Munro and Venkatesh Raman. Succinct representation of balanced parentheses and static trees. *SIAM J. Comput.*, 31(3):762–776, 2001.
- [14] J. Na and K. Park. Simple implementation of string b-trees. In Alberto Apostolico and Massimo Melucci, editors, *SPIRE*, volume 3246 of *Lecture Notes in Computer Science*, pages 214–215. Springer, 2004.
- [15] G. Navarro. Indexing text using the ziv-lempel trie. *Journal of Discrete Algorithms (JDA)*, 2(1):87–114, 2004.
- [16] G. Navarro and V. Mäkinen. Compressed full-text indexes. *ACM Computing Surveys*, 39(1):2, 2007.
- [17] K. Sadakane. New text indexing functionalities of the compressed suffix arrays. *J. Algorithms*, 48(2):294–313, 2003.
- [18] P. Weiner. Linear pattern matching algorithm. In *Proc. 14th IEEE Symposium Switching Theory and Automata Theory*, pages 1–11, 1973.

## Minería de Datos utilizando Sistemas Inteligentes

Laura Lanzarini<sup>1</sup>, Waldo Hasperué<sup>2</sup>, Leonardo Corbalán<sup>3</sup>, Sonia Formia<sup>4</sup>,  
César Estrebou<sup>5</sup>, Augusto Villa Monte<sup>6</sup>, Germán Aquino<sup>7</sup>, Marcela Jeréz<sup>8</sup>

Instituto de Investigación en Informática LIDI (III-LIDI)<sup>9</sup>  
Facultad de Informática. UNLP

### CONTEXTO

Esta presentación corresponde al Subproyecto “Sistemas Inteligentes” perteneciente al Proyecto “Procesamiento paralelo y distribuido. Fundamentos y aplicaciones en Sistemas Inteligentes y Tratamiento de imágenes y video” del Instituto de Investigación en Informática LIDI.

### RESUMEN

Esta línea de investigación se centra en el estudio y desarrollo de estrategias adaptativas aplicables a la Minería de Datos.

Se han desarrollado distintas metaheurísticas que permiten obtener reglas y listas de clasificación, capaces de operar sobre datos numéricos y categóricos, a partir de datos estructurados etiquetados.

Se han analizado distintas técnicas no supervisadas para identificar las características más importantes de la deserción universitaria en base a la información de la UNRN.

Se está comenzando a trabajar con técnicas que operan con información no estructurada aplicables a la caracterización y clasificación de documentos.

**Palabras clave:** Minería de Datos, Minería de Textos, Reglas de Clasificación. Redes Neuronales. Hiper-rectángulos. Técnicas de Optimización.

### 1. INTRODUCCIÓN

En el Instituto de Investigación en Informática LIDI se está trabajando, desde hace varios años, en la resolución de problemas pertenecientes al área de Minería de Datos utilizando Sistemas Inteligentes.

Se han desarrollado diferentes técnicas de extracción de conocimiento utilizando diferentes herramientas inteligentes como las redes neuronales, la lógica difusa y las metaheurísticas de algoritmos genéticos y optimización por cúmulo de partículas.

Todas las técnicas desarrolladas trabajan sobre bases de datos estructuradas generando modelos en forma de reglas de clasificación o de asociación. En estos últimos meses se ha comenzado a trabajar con información no estructurada como el análisis semántico para la clasificación de documentos.

A continuación se detallan brevemente los avances realizados últimamente.

#### 1.1. Obtención de reglas a partir de hiperrectángulos

Se ha desarrollado una técnica adaptativa, denominada CLUHR [5], que permite extraer conocimiento de grandes bases de datos a partir de un modelo dinámico capaz de adaptarse a los cambios de la información.

Esta técnica utiliza hiper-rectángulos, una poderosa forma de representación de datos,

<sup>1</sup> Profesor Titular DE. Facultad de Informática. UNLP

<sup>2</sup> Becario Post-doctoral (CONICET) – Jefe de Trabajos Prácticos - Facultad de Informática. UNLP

<sup>3</sup> Profesor Adjunto – Facultad de Informática – UNLP.

<sup>4</sup> Profesor Adjunto UNRN.

<sup>5</sup> Jefe de Trabajos Prácticos SD - Facultad de Informática. UNLP

<sup>6</sup> Becario CIN. Ayudante Diplomado - Facultad de Informática. UNLP

<sup>7</sup> Becario III-LIDI. Adscripto a cátedra - Facultad de Informática. UNLP

<sup>8</sup> Jefe de Trabajos Prácticos DE. Facultad de Ingeniería. UNPSJB.

<sup>9</sup> Calle 50 y 120 - 2do Piso, (1900) La Plata, Argentina, TE/Fax +(54) (221) 422-7707. <http://weblidi.info.unlp.edu.ar>



por su capacidad para describir de manera casi natural el subconjunto de datos al cual representa. Esto se debe a que los límites de cada hiper-rectángulo pueden ser utilizados como cláusulas en las reglas del tipo IF-THEN que resulten del proceso de extracción de conocimiento.

La estrategia propuesta comienza con un conjunto de hiper-rectángulos definidos a partir de los datos iniciales. Luego, en un proceso iterativo, se van eliminando superposiciones de acuerdo a los valores tomados por un conjunto de índices especialmente diseñados para este efecto. Como resultado de este proceso, los hiper-rectángulos cambian de tamaño o se dividen.

Este proceso de optimización continúa hasta minimizar (o anular) el volumen de intersección entre hiper-rectángulos de distintas clases. Finalmente se generan las reglas que resultan de los distintos hiper-rectángulos conseguidos.

La gran desventaja de CLUHR es que solo trabaja con atributos numéricos. Se ha desarrollado una mejora a esta técnica que incluye el tratamiento de atributos nominales. Esta mejora resulta en una técnica más poderosa denominada CLUIN [6].

Ambas técnicas, para llevar a cabo el armado del modelo de datos, deben tomar muchas decisiones. La decisión de que superposición resolver es decidida mediante el cálculo de los índices de superposición. Todas estas decisiones se configuran en forma previa al armado del modelo permitiendo que el algoritmo opere automáticamente.

Sin embargo, ambas técnicas son completamente flexibles permitiendo su ejecución de manera semi-automática o manual. Un experto puede intervenir en el proceso de obtención del modelo decidiendo ciertos aspectos claves del armado del mismo, logrando un resultado más acorde al problema presentado.

Otro aspecto importante de las técnicas desarrolladas es que presentan un comportamiento adaptativo, los nuevos datos que se ingresen al modelo de datos, causan la

modificación de la estructura interna del modelo, evitando que el modelo no deba rehacerse nuevamente utilizando el conjunto de datos completo. Una vez actualizada la estructura interna del modelo se actualiza el conjunto de reglas existentes.

## 1.2. Obtención de reglas a partir de técnicas de optimización

Esta línea de investigación está centrada en la obtención de reglas de clasificación, del tipo IF-THEN, a partir de redes neuronales y técnicas de optimización.

En especial se estudian métodos de clustering y clasificación de patrones que permitan identificar, de una manera no supervisada, aquellos atributos relevantes para el problema. Dichos atributos serán especialmente considerados en el momento de construir el antecedente de la regla.

Este tipo de estrategias fueron utilizadas previamente en [11] para medir la relevancia de los términos más utilizados en un conjunto de e-mails. Sin embargo, las técnicas de agrupamiento no poseen la capacidad de seleccionar atributos. Dado que sólo operan con información numérica pueden utilizarse para obtener medidas de tendencia central para cada grupo a partir del conjunto de datos asociado. Esto último no implica el proceso de selección.

Como forma de identificar cuáles de los atributos son relevantes para la construcción de la regla, se investigan distintas variantes de optimización por cúmulos de partículas. Interesa especialmente el control adecuado de la velocidad ya que se relaciona directamente con la precisión de la respuesta obtenida.

Los resultados de esta investigación han dado lugar a una estrategia adaptativa capaz de generar una lista de reglas de clasificación reducida operando con atributos nominales y numéricos. La misma se basa en la combinación de una red neuronal SOM [7] y una técnica de optimización poblacional. Las reglas obtenidas se caracterizan por su

simplicidad y facilidad de interpretación dado que poseen pocos atributos en su antecedente.

Los resultados de esta investigación han sido publicados en [12]

### 1.3. Minería de Datos en Educación

La aplicación de técnicas de Minería de Datos en el ámbito educativo ha permitido caracterizar a los distintos actores que intervienen en los procesos de enseñanza-aprendizaje [10].

En el III-LIDI, se trabaja en este tema desde 2008. Las investigaciones realizadas han permitido evaluar la pertinencia y calidad del material desarrollado para un curso dado [4] [8]. También se estudiaron técnicas aplicables a la modelización del estudiante en lo referido a su proceso de aprendizaje [1] [9].

Actualmente, uno de los temas que más preocupa a las distintas unidades académicas es la deserción universitaria. Por tal motivo, se han estudiado distintas técnicas de modelización no supervisadas encontrando que las Redes Neuronales ofrecen una buena caracterización de los datos del problema. Se ha trabajado a partir de la información de los alumnos de la UNRN recolectados a través del sistema SIU-Guaraní exceptuando todos los datos sensibles del alumnado (DNI, Apellido y Nombre, nombre de los padres, etc).

Se realizó un preprocesamiento importante de la información que dio lugar a la vista minable. Luego a partir de distintos procesos de clustering basados en redes neuronales se identificaron las características más relevantes del problema llegando a la conclusión que la situación socio-económica del alumno tiene una fuerte incidencia en su permanencia en el ámbito universitario. Esto permitió demostrar también que las técnicas elegidas se adaptan al objetivo planteado.

Los resultados de esta investigación fueron presentados en [3]

### 1.4. Minería de Textos

La Minería de Textos posee los mismos objetivos generales que la Minería de Datos pero opera sobre colecciones de documentos de texto no estructurado. Las tareas que habitualmente se llevan a cabo pueden dividirse básicamente en las siguientes categorías: agrupamiento de documentos, categorización, clasificación y asociaciones de conceptos [2].

Esta línea de investigación tiene su eje central en el estudio y aplicación de distintos métodos de representación de documentos así como de distintas técnicas adaptativas aplicables en la resolución de problemas de agrupamiento y categorización.

Por el momento se ha utilizado el enfoque convencional basado en la generación de un diccionario de palabras y mediante distintas herramientas de visualización se han representado las relaciones más relevantes.

A futuro, interesa representar la relación semántica entre los términos que componen un mismo documento. Esto ayudaría a reforzar su importancia a la hora de buscar similitudes brindando mejores resultados.

## 2. TEMAS DE INVESTIGACIÓN Y DESARROLLO

- Investigación de nuevos índices de superposición de hiper-rectángulos con el objetivo de obtener mejores resultados ante problemas específicos.
- Estudio de técnicas de simplificación de modelos basados en reglas de clasificación.
- Estudio de distintas variantes de PSO capaces de controlar adecuadamente la velocidad con la que se mueven las partículas.
- Análisis de las limitaciones de PSO para operar sobre datos nominales. Identificación de las zonas más prometedoras del espacio de búsqueda utilizando SOM.

- Estudio de distintas técnicas de preprocesamiento aplicables a Minería de Textos.
- Estudio, análisis y comparación de diferentes técnicas de visualización.
- Estudio y desarrollo de métodos para la identificación de los atributos más relevantes de un conjunto de datos.
- Estudio de técnicas de agrupamiento aplicables a información numérica y categórica.

### 3. RESULTADOS OBTENIDOS/ ESPERADOS.

- Desarrollo de una técnica de extracción de conocimiento que opera sobre bases de datos con atributos numéricos y nominales.
- Desarrollo de técnicas capaces de adaptar su estructura interna ante la llegada de nuevos datos.
- Desarrollo e implementación de una representación para PSO de una regla de clasificación que incluya atributos numéricos y nominales.
- Desarrollo e implementación de un nuevo método capaz de obtener una lista de clasificación formada por un número reducido de reglas sencillas que operan tanto sobre atributos numéricos como nominales a partir de la combinación de una red SOM con una metaheurística basada en cúmulo de partículas.
- Desarrollo de una prueba de concepto que arroja información preliminar relevante respecto a la problemática del abandono.
- Descripción del perfil de los estudiantes aportando información útil en relación a su composición socio-económica y su permanencia en el ámbito universitario.
- Desarrollo de una herramienta para caracterización y clasificación de documentos.

### 4. FORMACIÓN DE RECURSOS HUMANOS

Dentro de los temas involucrados en esta línea de investigación se ha finalizado una tesis de doctorado y se están desarrollando actualmente 2 tesis de doctorado, 2 de maestría y al menos 3 tesinas de grado de Licenciatura. También participan en el desarrollo de las tareas becarios y pasantes del III-LIDI.

### 5. REFERENCIAS

- [1] Arona, Huapaya, Lanzarini, Lizarralde. Lógica Difusa aplicada al Modelo del Estudiante de un Sistema Tutorial Inteligente. IV Congreso de Tecnología en Educación y Educación en Tecnología. TE&ET'09. Julio 2009. La Plata. Bs.As
- [2] Berry, M.B., Kogan, J. (editors). Text Mining: Applications and Theory. John Wiley & Sons. 2010. ISBN 978-470-74982-1
- [3] Formia S. Evaluación de técnicas de Extracción de Conocimiento en Bases de Datos y su aplicación a la deserción de alumnos universitarios. Tesis de Especialista en Tecnología Informática aplicada en Educación. Dic 2012.
- [4] Grossi, M.D., Lanzarini, L. Reglas de Predicción aplicables al Diseño de un Curso de Computación. III Congreso de Tecnología en Educación y Educación en Tecnología. TE&ET'08. Mayo 2008. Bahía Blanca.
- [5] Hasperué, W., Lanzarini, L., De Guisti, A. Rule Extraction on Numeric Datasets Using Hyper-rectangles. Computer and Information Science. Vol. 5, No 4, pp. 116 131. 2012. <http://dx.doi.org/10.5539/cis.v5n4p116>
- [6] Hasperué, W., Corbalan, L. CLUIN – A New Method for Extracting Rules for Large Databases. XIII Workshop de Agentes y Sistemas Inteligentes, XVIII Congreso Argentino de Ciencias de la Computación. Bahía Blanca. Argentina. Octubre 2012. Págs. 130-139.

- [7] Kohonen, T. Self-Organizing Maps. 2nd Edition. Springer. ISSN 0720-678X (1997)
- [8] Lanzarini, Denazis, Grossi Estrategias Inteligentes aplicables a un Sistema Educativo. X Workshop de Investigadores en Ciencias de la Computación (WICC 2008), Area Tecnología Informática Aplicada en Educación. Mayo de 2008. La Pampa
- [9] Lanzarini, Huapaya. Diagnóstico Adaptativo del Estudiante en Sistemas Tutoriales Inteligentes. XI Workshop de Investigadores en Ciencias de la Computación (WICC 2009), Area Tecnología Informática Aplicada en Educación. Mayo de 2009. San Juan
- [10] Romero C., Ventura S. Educational data mining: A survey from 1995 to 2005. Expert Systems with Applications, Volume 33, Issue 1, July 2007, Pages 135-146, ISSN 0957-4174.
- [11] Villa Monte, A. Estrebou, C., Lanzarini, L. E-mail processing using data mining techniques. Publicado en el Libro *Computer Science & Technology Series – XVI Argentine Congress of Computer Science Selected Papers*, ISBN 978-950-34-0757-8. EDULP, Argentina , 2011. Págs. 109-120.
- [12] Villa Monte, A., Ronchetti, F., Lanzarini, L., Jeréz, M. Obtención de reglas de clasificación usando SOM+PSO. XIII Workshop de Agentes y Sistemas Inteligentes”. CACIC2012. ISBN: 978987-1648-34-4. Pág. 210-219. Bahía Blanca, Buenos Aires, Argentina, Octubre 2012.



# Sistema de Información Geográfica en la Web para el Problema de Ruteo de Vehículos

de San Pedro M.E., Lasso M., Serón N.,  
Carrizo, A., Montenegro C., Burgos E., Ramos L.

Laboratorio de Tecnologías Emergentes (LabTEM)

Unidad Académica Caleta Olivia

Universidad Nacional de La Patagonia Austral

e-mail: {edesanpedro, mlasso, nseron, acarrizo, cmontenegro}@uaco.unpa.edu.ar;

[edubual@gmail.com](mailto:edubual@gmail.com); [ramosluis\\_1@hotmail.com](mailto:ramosluis_1@hotmail.com)

## Resumen

En este proyecto se abordan dos temáticas desde diferentes ámbitos de investigación y desarrollo; se plantea como objetivo, el estudio de las herramientas tecnológicas para lograr la integración entre el SIG, la información logística del Problema de Ruteo de Vehículos (VRP), los modelos matemáticos y técnicas metaheurísticas de optimización que permita resolver los problemas de ruteo de vehículos.

En la actualidad existen innumerables aplicaciones de software basadas en web, que intercambian información. Existe un alto grado de crecimiento de aplicaciones de software basadas en Web, por sobre las aplicaciones de escritorio.

Los Sistemas de Información Geográfica (GIS) no son la excepción a esta tendencia. Son sistemas de Información para capturar, almacenar, comprobar, integrar, manipular, analizar y presentar eficientemente todas las formas de información georeferenciada.

**Palabras clave:** *Sistema de Información Geográfica, Metaheurísticas, Optimización de recorridos, Problema de Ruteo de Vehículos.*

## Contexto

Esta línea de investigación comienza como una actividad de Vinculación Tecnológica entre la Universidad y la Municipalidad de Caleta Olivia [Serón et al. 2008], y continuó en el marco de la convocatoria a becas de innovación tecnológica que ofrece todos los años la Fundación del Banco Santa Cruz, a las cuales se ha podido acceder en dos oportunidades [de San Pedro et al. 2009], [Serón et al. 2010a], [Serón et al. 2010b].

Con el objetivo de continuar en esta línea, se presentó el Proyecto de Investigación, Sistema de Información Geográfica integrado con un sistema

logístico basado en el Problema de Ruteo de Vehículos, en el marco del programa de Investigación en Ciencia y Tecnología de la Universidad Nacional de la Patagonia Austral; el que se está desarrollando en el Laboratorio de Tecnologías Emergentes (LabTEM), de la Unidad Académica Caleta Olivia, donde el objetivo general es la investigación y el desarrollo en nuevas tecnologías basadas en Sistemas Inteligentes y Software Libre.

Actualmente se está trabajando en el estudio de herramientas que permitan utilizar GIS basados en Web, analizando las diferencias entre Gis Internet y Gis Convencional, cuales son las herramientas Gis Basadas en Web, y los lenguajes de programación para tal fin.

## Introducción

El amplio crecimiento del uso de la Web afecta a los GIS, en el sentido de un mayor requerimiento de Información Geográfica. Esto se debe a que la Información no geográfica que obtenemos a través de la Web, posee una relación implícita o explícita con datos geográficos.

Las aplicaciones GIS basadas en Web, son sistemas de Información geográfica equivalentes en cuanto a prestaciones y posibilidades de un Gis de Escritorio. La información geográfica se encuentra en Servidores Remotos y no en la computadora del usuario.

En Gis Web se accede a los datos a través de algún Web Browser, es decir el Navegador Web se transforma en la principal vía de visualización de información geográfica.

En numerosas ocasiones necesitamos conocer información relacionada a distancias entre sitios, localización de servicios entre otros. Para resolver estos requerimientos los mapas son de una gran

utilidad, convirtiéndose en una herramienta de relevancia en la toma de decisiones.

En la actualidad se acentúa el uso de GIS Web, en el análisis y consultas de grandes volúmenes de información geográfica. Además resuelven las consultas geográficas en lapsos cortos de tiempo.

#### *Herramientas Gis Basadas en Web*

Actualmente varias empresas de GIS ofrecen productos comerciales basados en el Web. A su vez existe un amplio número de Aplicaciones GIS Web de Software Libre.

Algunos productos comerciales que podemos mencionar son los siguientes:

- **ESRI: Arc IMS (Internet Map Server)**

Desarrollado específicamente para distribuir Servicios GIS en la Internet, ArcIMS está diseñado para crear servicios de geoprocesamiento, diseñar páginas Web para los clientes y administrar los recursos realizando balanceo de cargas. ArcIMS opera en un ambiente distribuido compuesto por recursos informáticos repartidos entre los clientes y los servidores que permiten un máximo aprovechamiento de los recursos. Típicamente un cliente envía un requerimiento al servidor, este accede a las bases de datos, genera el análisis/reporte y devuelve la información al cliente en forma de mapas, datos tabulares y gráficos de fácil comprensión.

- **ERDAS: Image Web Server**

Una aplicación especializada de alta velocidad que distribuye eficientemente cantidades masivas de imágenes geoespaciales a miles de usuarios, en un solo servidor. Solucionando problemas de congestión de infraestructura tradicionalmente asociados con el despliegue de grandes cantidades de información de imágenes, los usuarios tienen acceso a la información que necesitan rápidamente.

#### *Algunos Proyectos GIS Web de Software Libre:*

- **Map Server**

Es una aplicación servidor de Código Abierto que te permite publicar imágenes de mapas geográficos y datos vectoriales a través de Internet

- **Alov Map**

Es una solución basada en Java con la que crear aplicaciones cartográficas, libre y sin costo, destinadas a la web.

- **GeoServer**

Es un Servidor Web que permite servir mapas y datos de diferentes formatos para aplicaciones Web,

ya sean clientes Web ligeros, o programas GIS desktop.

- **OpenLayer**

Es un framework JavaScript que permite incluir mapas georeferenciados en cualquier página web.

Aunque estos GIS basados en el Web parecen ser similares se basan en diferentes arquitecturas bases de datos, plataformas, formatos de datos y diferentes metodologías como CGI (Common Gateway Interface), Plug-ins, html extendido y Java.

#### *El problema de un formato estándar de almacenamiento.*

Sin embargo a pesar de las diferencias mencionadas anteriormente, el problema más común que se tiene en el campo de los GIS es la existencia de diferentes formatos para almacenar datos geográficos, lo que dificulta el intercambio de la información geográfica.

Establecer un formato estándar de almacenamiento permitirá un mayor acceso a la información de diferentes GIS sin la necesidad de contar con algún producto en especial, sobre todo en el contexto de Internet.

Esfuerzos de este tipo son llevados a cabo por comités como el Consorcio OPEN GIS (Open Geospatial Intereoperability Specification Consortium, o OGC).

Sin embargo actualmente muchos de los datos de un GIS son almacenados en sistemas de archivos o bases de datos con formatos específicos y solo a través del respectivo GIS es posible consultar y analizar los datos, como consecuencia los usuarios no pueden utilizar un solo producto GIS para tener acceso a todos los datos que deseen, lo que resulta en un alto costo si se adquiere alguno o varios productos GIS.

Otro problema que se tiene es el tiempo que se requiere para conocer y entender algún producto específico, además del tiempo y los recursos que se utilizan para instalar y configurar cada producto.

El problema aumenta en Internet ya que existen datos geográficos distribuidos en todo el mundo, almacenados con diferentes formatos y administrados por diferentes GIS.

#### *Java gana aceptación*

Para resolver el problema mencionado anteriormente un número considerable de Applets de Java y Software relacionado con los GIS han empezado a emerger en la Web recientemente. El rango va desde Applets simples, inmaduras y experimentales hasta aplicaciones complejas.

En este campo, Java ha ganado una rápida aceptación como un Lenguaje de Programación Web, por su independencia de plataforma y programación orientada a objetos. Aunque muchas de las Java Applets, son hechas principalmente en el campo de la investigación, con el objeto de resolver problemas que aparecen durante investigaciones.

Los SIG tienen dos componentes fundamentales: a) Un modelo de datos en el que se almacenan las características de los objetos geográficos, de manera similar a como se almacenan en una base de datos convencional, junto con información posicional (coordenadas) y las relaciones entre los distintos objetos, b) Una colección de funciones que nos permiten interrogar a la base de datos y obtener respuestas, ya sea en base a listados o a imágenes (mapas).

Una característica esencial de los SIG, es que intentan capturar en su modelo, datos de la realidad, y no una imagen determinada de ésta. [Rubio Barroso 1997]

Los datos (en este caso datos espaciales), son uno de los elementos más importante en un SIG, por esto es imprescindible conocer cuales son los tipos de datos característicos en estos sistemas.

Por otro lado, los problemas de ruteo de vehículos o de distribución física de mercancías desde almacenes a clientes aparecen en la literatura científica como Vehicle Routing Problems, más comúnmente como VRP. También se puede encontrar, aunque en menor medida, referencias como Vehicle Scheduling Problems. En términos generales, un problema de rutas de vehículos consiste en determinar las rutas de un conjunto (o flota) de vehículos que deben iniciar un recorrido (y finalizarlo) en los almacenes (o depósitos) para atender la demanda de servicio de un conjunto disperso de clientes sobre una red. En la literatura, algunos autores han intentado clasificar y simplificar la gran variedad de posibles problemas, como por ejemplo los criterios propuestos por [Bodin y Golden, 1981] y [Desrochers et al. 1990]; que intentan reflejar y ordenar las principales características en aspectos como: el almacén, la flota, la demanda, el servicio y el objetivo a alcanzar. Esta clasificación de los problemas, ha facilitado tanto el desarrollo de modelos matemáticos y estrategias de resolución, como la toma de decisiones por parte de las empresas

Uno de los primeros estudios que trataron el problema de ruteo de vehículos se remonta al año 59, en este trabajo Dantzing y Ramser [Dantzing & Ramser, 1959] tratan un problema de despacho con camiones, que surge como una generalización del problema clásico del agente viajero (TSP) en el que un vendedor tiene que visitar una serie de clientes una sola vez, para luego volver al lugar de partida, construyendo una camino hamiltoniano sobre el

grafo constituido por los clientes (vértices) y los caminos posibles entre un cliente y otro (aristas). El VRP tiene por objetivo encontrar las rutas que recorran cada uno de los vehículos (ubicados en un depósito) de manera que se satisfagan los requerimientos de los clientes, las restricciones operativas y se minimice el costo total de transporte.

En la historia reciente del VRP ha habido una evolución constante en la calidad de las metodologías resolutivas utilizadas en este problema, pertenecientes tanto al campo de investigación exacto como al heurístico. De todas formas, dada la dificultad del problema, ningún método exacto conocido es capaz de encontrar el óptimo para instancias que contengan más de 50 clientes [Golden et al. 1998].

Una recopilación de técnicas exactas de solución existentes para los problemas de ruteo de vehículos puede encontrarse en [Laporte 1992]. No obstante, los de gran dimensión resultan imposibles de solucionar en tiempo polinomial, por lo que el VRP es un problema NP-hard [Machado et al. 2000] y [Olivera 2004], donde no es posible alcanzar una solución óptima y dependiendo de las características especiales de clientes, locaciones y producto/servicio, requiere la elaboración de una metodología de solución específica con la cual se aproxime lo mejor posible al óptimo. Debido a estas razones y a la relevancia práctica del VRP se han propuestos varias soluciones a este problema haciendo uso de heurísticas y metaheurísticas. Algunos ejemplos incluyen Tabu Search [Cordeau 1997], Simulated Annealing [Osman 1993], Ant Colony [Bell & McMullen 2004], Algoritmos Evolutivos [Bäker & Ayechev 2003], [Xu et al. 2005], entre otras.

Con el propósito de poder integrar las tecnologías que brindan los sistemas de información geográfica, con diferentes metaheurísticas para resolver problemas de optimización, existe actualmente un grupo de desarrollo en la Universidad Politécnica de Valencia, <http://personales.upv.es/arodrigu> que aborda esta problemática a través de una herramienta informática donde el objetivo está planteado para facilitar la resolución de problemas reales de flotas de vehículos capacitados (CVRP), el cálculo de rutas, y su gestión. Aquí se integran inteligentemente tres elementos: el SIG, la información del sistema logístico (VRP-XML), los modelos matemáticos y técnicas de optimización combinatoria que conjuntamente permiten resolver los problemas de rutas para flotas de vehículos, [Toth et al. 2001].

El éxito en la gestión logística depende de la capacidad de integración (información y sistemas, proveedores y clientes, recursos y decisiones, etc.). Por ello, el proyecto desarrollado en la Universidad de Valencia, ha prestado especial interés a la

integración necesaria para la optimización del transporte, la toma de decisiones y la gestión de flotas.

Para el modelado, resolución y análisis de este tipo de problemas, es necesario gestionar una enorme cantidad de información: datos sobre las características de la flota de vehículos, los planes de ruta, los cargamentos, depósitos y retiros, información geográfica, las restricciones, la función objetivo, etc. Tal y como se explica en [Rodríguez 2006], se trata de una estructura de etiquetas VRP-XML que define los elementos de un documento que facilita el intercambio de datos en el contexto de los VRP. El módulo VRP-XML se enlaza fácilmente con la planificación de recursos empresariales (ERP), compartiendo datos de: clientes, servicios y órdenes de trabajo, información sobre costos de operaciones, disponibilidad e información sobre los recursos logísticos (flota de vehículos), franjas horarias y otro tipo de restricciones. Además este tipo de problemas son dinámicos y cambian en el tiempo, sus datos deben de estar soportados por una estructura flexible, capaz no sólo de atender tal cantidad de información según los actuales requerimientos de la empresa, sino también los futuros del sistema logístico (ampliación del número de clientes, de la flota de vehículos, nuevas restricciones, etc.).

## Líneas de Investigación y desarrollo

Este proyecto tiene como objetivo principal, integrar un Sistema de Información Geográfica con un modelo de optimización de ruteo de vehículos (*Vehicle Routing Problem*) realizando el cálculo de las rutas y su gestión logística.

Este objetivo podrá lograrse desde el abordaje de dos temáticas específicas:

- El estudio del arte sobre herramientas tecnológicas existentes dentro del ámbito de software libre, que permitan la integración de la resolución de un problema real en un Sistema de Información Geográfica y,
- La investigación, desarrollo y aplicación de metaheurísticas para la optimización de recorridos que facilite la resolución de problemas reales de flotas de vehículos capacitados.

## Resultados y Objetivos

En el marco de este proyecto, se busca promover la relación Universidad – Empresa, a través de convenios realizados con empresas dedicadas al rubro de servicios petroleros (para distribución de personal, recorrido de locaciones, etc.), y la empresa Teleservicios Caleta Olivia S.E.

(prestadora de servicios de telecomunicaciones e Internet en Caleta Olivia y Cañadón Seco).

En cuanto a la relación del proyecto con las actividades de Formación de Grado, se propone su vinculación con los espacios curriculares, Programación, Modelos y Simulación, Sistemas Inteligentes Artificiales. Además de promover la generación de espacios donde los alumnos puedan desarrollar sus proyectos finales de la carrera Ingeniería en Sistemas.

## Formación de Recursos Humanos

Un integrante se ha incorporado como becario alumno con el objetivo de seguir expandiendo esta línea en la comunidad de la UACO.

Un integrante, actualmente cursando la Maestría en Informática y Sistemas de la UNPA, ha comenzado a definir su plan de tesis en temas afines a los propuestos en el proyecto.

Dos alumnos de la carrera Ingeniería en Sistemas, están en proceso de definición de su plan de tesis para el desarrollo del Proyecto Final para obtener su titulación, en temáticas que estarían orientadas a las propuestas como líneas de desarrollo en este proyecto.

## Referencias

- [Bäker & Ayechev 2003] Bäker B.M. and Ayechev M.A., A genetic algorithm for the vehicle routing problem. *Computers & Operations Research*, pages 787-800. 2003
- [Bell & McMullen 2004] Bell J. and McMullen P., Ant Colony Optimization techniques for the vehicle routing problem, *Advanced Engineering Informatics*, pages 41-48. 2004
- [Bodin y Golden 1981] Bodin, L., Golden, B. (1981). Classification in Vehicle-Routing and Scheduling, *Networks*, 11(2), pp. 97-108.
- [Cordeau 1997, Cordeau J-F, Gendreau M. and Laporte G. A tabu search heuristic for periodic and multi-depot vehicle routing problems. *Networks*, 30(2): 105-119, 1997.
- [Dantzing & Ramser 1959] Dantzing G.B. and Ramser J.H., "The Truck Dispatching Problem" *Management Science*, Vol. 6, Nro. 1 (Oct., 1959), pp 80-91.
- [de San Pedro et al. 2009a] de San Pedro Maria Eugenia, Serón Natalia, Cristian Montenegro - Sistema de información geográfica aplicado a turismo y patrimonio histórico y cultural - 11° Workshop de Investigadores de Ciencias de la Computación WICC'09, pp.438-441 ISBN 978-



- 950-605-570-7, Universidad Nacional de San Juan, San Juan – Mayo 2009.
- [Desrochers et al. 1990] Desrochers, M., Lenstra, J. K., Savelsbergh, M. W. P. (1990). A Classification Scheme for Vehicle-Routing and Scheduling Problems, *European Journal of Operational Research*, 46(3), pp. 322-332.
- [Golden et al. 1998] Golden B.L., Wasil E.A., Kelly J.P., and Chao I-M. *Fleet Management and Logistics*, chapter The Impact of Metaheuristics on Solving the Vehicle Routing Problem: algorithms, problem sets, and computational results, pages 33-56. Kluwer Academic Publishers, Boston 1998
- [Laporte 1992] Laporte, G., The vehicle routing problem: an overview of exact and approximate algorithms. *European Journal of Operational Research*, 59, 1992, pp. 345.358.
- [Machado et al. 2000] Machado, P., Tavares, J., Pereira, F. and Costa, E., *Vehicle Routing Problem: Doing it the Evolutionary Way*.2000.
- [Olivera 2004] Olivera, A., *Heurísticas para Problemas de Ruteo de Vehículos*. Instituto de Computación, Facultad de Ingeniería. Universidad de la República, Montevideo, Uruguay. 2004.
- [Osman 1993] Osman I. H.. *Metastrategy Simulated Annealing and Tabu Search Algorithms for the vehicle routing problem*. *Ann. Operations Research* 40(1), pages 421-451. 1993.
- [Rodríguez 2006] Rodríguez, A. (2006). *VRP-XML: lenguaje de marcas extensible para los problemas de rutas de vehículos*. X Congreso de Ingeniería de Organización, Valencia.
- [Rubio Barroso 1997] Rubio Barroso “Los Sistemas de Información Geográficos: Origen y perspectivas” *Revista general de información y documentación*, ISSN 1132-1873, Vol. 7, N° 1, 1997, Págs. 93-106
- [Serón et al. 2008] Serón N., Montenegro C., Vidal P., Villagra S., Orozco S., Valdéz J., Díaz F., de San Pedro M. -Implementación de un Sistema de Información Geográfica en Municipios de la Zona Norte de la Provincia de Santa Cruz. X Workshop de Investigadores de Ciencias de la Computación - WICC 2008, La Pampa, Argentina, Mayo 2008, pp. 530-533.
- [Serón et al. 2010a] Serón N., Montenegro C., Valdez J., de San Pedro M.E. – *Sistema de Información Geográfica Aplicado al Turismo y Patrimonio Histórico y Cultural - II Jornadas de Extensión Universitaria UNPA*, Caleta Olivia, 18 de Agosto de 2010.
- [Serón et al. 2010b] Serón N., Montenegro C., Valdez J., de San Pedro M.E. – *Sistema de Información Geográfica Aplicado al Turismo y Patrimonio Histórico y Cultural - III Jornadas RedVITEC*, Mendoza, 25 y 26 de Noviembre de 2010.
- [Toth et al. 2001] Toth, P., Vigo, D. (2001). An overview of vehicle routing problems. In the *Vehicle Routing Problem*. Ed. Society for Industrial and Applied Mathematics. Philadelphia.
- [Xu et al. 2005] Xu Y L, Lim M. H. and Er M. J.. *Investigation on Genetic Representations for Vehicle Routing Problem*, *IEEE International Conference on System, Man and Cybernetics*, pages 3083- 3088 2005

## **CODIFEXA – Colecciones Digitales para Exactas**

Prog. Luis Alberto Olguin; Mag. Raúl Oscar Klenzi

Instituto de Informática – Departamento de Informática

Facultad de Ciencias Exactas, Físicas y Naturales

Universidad Nacional de San Juan

[lolguin@info.unsj.edu.ar](mailto:lolguin@info.unsj.edu.ar); [rauloscarklenzi@gmail.com](mailto:rauloscarklenzi@gmail.com)

### **Resumen**

La construcción de repositorios institucionales (RI) por parte de facultades de universidades argentinas es cada vez más significativa, pero para que este tipo de emprendimientos – generalmente con presupuestos ajustados – tenga sustentabilidad en el tiempo es necesario “mostrar que es útil” para la institución - los docentes en particular - en el sentido que permite aumentar la visibilidad de la producción científica local y por ende el impacto entre sus pares.

El proyecto que aquí describimos, Colecciones Digitales para la Facultad de Exactas (CODIFEXA), es el resultado del trabajo interdisciplinario de informáticos y bibliotecarios cuyo objetivo es sentar las bases para construir un RI para la Facultad de Ciencias Exactas de la Universidad Nacional de San Juan siguiendo la premisa de que en este tipo de emprendimientos “el centro de atención no es la tecnología, sino en las actividades humanas que se sirven de la tecnología”.

### **Contexto**

Este trabajo se enmarca en las actividades previstas en los proyectos “Minería de Datos en la determinación de perfiles de uso y perfiles de usuarios” (Código 21/E889) y “Colecciones Digitales para la Facultad de Exactas” (Código 21/E903), ejecutados por el Instituto de Informática y el Departamento de Informática de la Facultad de Ciencias Exactas, Físicas y Naturales de la Universidad Nacional de San Juan.

Estos proyectos fueron aprobados durante la convocatoria para la ejecución de proyectos bianuales realizada por CICITCA-UNSJ (2011-2012).

### **Introducción**

El concepto de objeto digital está presente en una abundante bibliografía. Testa [1] indica que se trata de un conjunto de “información nacida en soportes físicos que ha sido digitalizada y también documentos nacidos en formato digital”.

El conjunto de estos objetos digitales, generados por una misma institución y organizados de manera eficiente, conforma un repositorio institucional (RI), cuyo objetivo es la mejora en las posibilidades de encontrabilidad, recuperación y lectura por parte de otros.

Comenzar con la implementación de un RI no es tarea trivial. Murray<sup>[2]</sup> indica la importancia de entender que es “a system of people, practices, values and technologies in a particular local environment”.

El “sistema de personas” lo conforman autores, estudiantes, bibliotecarios, lectores y “web surfers”. Las “prácticas” se enmarcan en las acciones de auto-archivo, publicación, difusión, almacenamiento, leer, descargar. Los “valores” son los que se asocian al libre acceso a la información, la comunicación académica, la difusión de la investigación. Por último, las “tecnologías” se asocian al software a utilizar en el RI, los formatos de archivos, metadatos aplicados, la posibilidad de “cosecha” por parte de otros repositorios.

Según cita Tramullas<sup>[3]</sup>, los RI “se van convirtiendo progresivamente en los lugares en los que numerosas organizaciones almacenan y difunden el resultado de sus actividades. La importancia de este tipo de recursos de información comienza a considerarse como estratégica”.

Al igual que en muchos casos en Argentina, se decidió el estudio de software libre para la implementación de RI en lugar del desarrollo de una solución propia. En la actualidad ya no se piensa que estos software son desarrollos específicos para cada unidad sino que por el contrario se sustentan en equipos de desarrollo fuertes, sostenidos por instituciones o proyectos que permiten asegurar el financiamiento necesario para

asegurar la mejora e innovación del producto a lo largo del tiempo. Una de las características destacables de la adopción de software libre es la adaptabilidad, lo que permite aplicar soluciones locales al momento de la implementación.

Uno de estos softwares es Greenstone, incubado dentro del Proyecto Biblioteca Digital de Nueva Zelanda, Universidad de Waikato y distribuido en colaboración con la UNESCO y la FAO con sede en Amberes, Bélgica.

Nuestro proyecto, Colecciones Digitales para Exactas – en adelante CODIFEXA – adopta este software como plataforma para la implementación del repositorio fundamentalmente por sus facilidades de operación, lo que permite que rápidamente los bibliotecarios hagan uso de su interfaz, y porque responde al concepto de “Acceso abierto, Código abierto y Estándares abiertos”.

En este sentido, Corrado<sup>[4]</sup> afirma que las instituciones se beneficiarán trabajando con estos tres componentes en conjunto ya que es posible reducir costos de licencias de software, garantizar el acceso (abierto) - a largo plazo - a las publicaciones generadas por la institución sin las trabas que generalmente se presentan en las publicaciones por suscripción y continuar trabajando con estándares para la descripción de los recursos, lo que garantiza la interoperabilidad, visibilidad y encontrabilidad de la literatura.

Fournier<sup>[5]</sup> al realizar una descripción de metadatos indica que “entre la gran

variedad de formatos de metadatos existentes en Internet, la iniciativa Dublin Core [¿?] es la más citada y aceptada; es un formato producto de un esfuerzo internacional e interdisciplinario con una vida muy intensa y el más influyente en relación con el desarrollo de la teoría del uso de los metadatos para la recuperación de información en la red”.

En nuestra investigación en pos de definir el conjunto de metadatos aplicar en el RI, encontramos que muchos proyectos exitosos aplican Dublin Core, tal es el caso de la Biblioteca Digital Colombiana en la que confluyen 73 repositorios con más de 80.000 documentos digitales.

Así mismo desde el Ministerio de Ciencia, Tecnología e Innovación Productiva de Argentina, se publicó la recomendación de adopción de este modelo de metadatos para participar en el Sistema Nacional de Repositorios.

Actualmente CODIFEXA aplica el conjunto Dublin Core Simplificado, generando los metadatos a partir de los registros catalográficos creados por la biblioteca de la facultad de Exactas. Para esta tarea se diseñaron herramientas de software que permiten extraer la información de los campos de la base de datos CDS/ISIS hacia archivos DC-XML que luego son importados por Greenstone.

El protocolo OAI-PMH es el estándar abierto fundamental para asegurar la exposición, agregación, acceso e interoperabilidad de los contenidos depositados en los repositorios. El

protocolo OAI-PMH es muy sencillo y requiere el uso de otros estándares y protocolos, como el protocolo HTTP (protocolo Web) y el estándar Dublin Core para el formato de los metadatos que permite la interoperabilidad entre los repositorios.

Greenstone, desde su versión 2.84, posee un “OAI data provider” lo que asegura que el contenido del repositorio podrá ser cosechado por otros, aumentando la visibilidad del mismo.

Un RI sin la participación de sus actores principales, los autores de los trabajos, es imposible que logre apoyo institucional en el tiempo. Keefer<sup>[6]</sup> destaca que “los autores tienen mucho por ganar porque la presencia de sus trabajos en los repositorios aumentará la visibilidad y estabilidad, lo que presumiblemente contribuirá a recibir un mayor número de citas”.

Consientes que uno de los principales obstáculos para introducir documentos es el “miedo al plagio”, se conformó un grupo de trabajo en el proyecto que se encarga de las pautas de difusión y el estudio de normativas de otras universidades respecto a la cesión de derechos de publicación.

La difusión de los beneficios de contar con RI en la UNSJ es labor de este grupo, habiéndose elaborado una encuesta, que se encuentra actualmente vigente, para consultar a los docentes/investigadores acerca de sus actitudes frente al depósito de los trabajos por ellos generados. Resultados parciales muestran que un



26% de los docentes de la facultad han generado más de 10 artículos científicos en los últimos años; el área que más investigación muestra es Informática con un 43%; el 71% expresa desconocer el concepto “archivo institucional” y un 18% manifiesta que difunde sus publicaciones en sitios externos (Blog, Wikis, etc.).

El recurso humano de CODIFEXA conforma, como lo define Hernandez Arias<sup>[7]</sup>, un grupo de trabajo en el cual “un conjunto de individuos realizan actividades con la finalidad de lograr un objetivo específico”. Dado que los integrantes del proyecto no comparten el mismo espacio de trabajo, el planteo de trabajo colaborativo se basó en un mix de trabajo presencial y telemático. Con el fin de aplicar este último, se realizó el estudio de plataformas “Site As A Service” habiéndose seleccionado TeamLab como plataforma de actividad en el proyecto. Dentro de este espacio virtual de trabajo se comunican, coordinan, discuten, etc. las tareas asignadas a cada participante.

“No solo se trata de la interconexión en red de la tecnología sino la interconexión en red de los seres humanos mediante la tecnología”. Para nuestro equipo de trabajo, en particular para el grupo de bibliotecarios, aplicar tecnología en el modelo colaborativo ha sido novedoso y ha permitido organizar y optimizar el tiempo productivo.

## **Líneas de investigación y desarrollo**

Dentro del proyecto, que se ajusta a una investigación de informática aplicada, la línea principal es la aplicación de soluciones open source para el diseño e implementación de repositorios institucionales.

Las acciones llevadas adelante abarcan temas referidos a la implementación de software libre en diferentes plataformas, el estudio y aplicación de protocolos de interoperabilidad en repositorios, el manejo de grandes volúmenes de información mediante la aplicación de la API Lucene para recuperación en texto completo.

Un área de especial interés que surge como línea de investigación en el proyecto es la minería web, en particular lo referido a la aplicación de soluciones de código abierto (o gratuito) para obtener métricas de accesibilidad a los contenidos de repositorios. Este trabajo se lleva adelante en forma conjunta con el grupo de investigación del proyecto “Minería de Datos en la determinación de perfiles de uso y perfiles de usuarios” y apunta a formar recursos humanos capaces de interactuar con estas herramientas para aplicarlas en futuras actualizaciones del repositorio en construcción.

También forma parte de las líneas de investigación abordadas por el proyecto la preservación de objetos digitales, tarea que se encara en forma conjunta con las actividades que desarrolla el equipo del proyecto “*Red COES de repositorios de*

*acceso abierto* - PICT-O 2010-132” y tiene como objetivo realizar un relevamiento de las actuales técnicas de preservación aplicadas a nivel internacional a fin de realizar recomendaciones en esta área para ser aplicadas al repositorios en creación.

### **Resultados y Objetivos**

Dentro de los objetivos planteados en CODIFEXA se han alcanzado los referidos a “implementar un repositorio institucional con el objeto de almacenar la producción intelectual generado por la Facultad de Ciencias Exactas de la Universidad Nacional de San Juan”. Actualmente se encuentran publicadas 70 tesis de licenciaturas en Informática, Geología y Geofísica.

El software aplicado es Greenstone versión 2.85 sobre el que se trabajó aplicando HTML5 y CSS3 con el objeto de personalizar la interfaz de usuario. El motor de búsqueda aplicado es Lucene, provisto por Greenstone.

La tabla “crosswalk” de migración de registros desde formato LIBRI-UNSJ hacia XML-DC y el software asociado forma parte de los objetivos alcanzados por el proyecto.

Se pretende alcanzar nuevos objetivos relacionados al Open Access y los autores, a partir de los resultados que se obtengan al completar el estudio de respuestas a la encuesta lanzada a docentes de la FCEF.N.

### **Formación de RRHH**

Durante este periodo integrantes del proyecto han trabajado en temas de investigación conjuntos con los proyectos “Minería de Datos en la Determinación de Patrones de Uso y Perfiles de Usuarios” (cód. 21/ E889) y “Red COES de Repositorios de Acceso Abierto” (PICT-O CIN 2010-Agencia Nacional de Promoción Científica y Tecnológica).

La formación de los integrantes del proyecto se concretó en la participación de cursos acerca de aplicación de metadatos, construcción de colecciones digitales, open Access.

La transferencia al medio se realizó mediante el dictado de charlas, talleres referidos a los temas de investigación abordados.

Actualmente se encuentran en curso la concreción de tesis de grado de alumnos de la Licenciatura en Informática que abordan las temáticas OPAC-Web.

### **Referencias**

[1] TESTA, Patricia. Descripción de Objetos Digitales: METADATOS [en línea]. 2009. Disponible en [http://www.bn.gov.ar/descargas/catalogadores/ponencias/251109\\_09a.pdf](http://www.bn.gov.ar/descargas/catalogadores/ponencias/251109_09a.pdf)

[2] Murray, Annie. What we talk about when we talk about institutional repositories [en línea]. 2008 [fecha de consulta: 5 de marzo 2013]. Disponible en

[http://library.concordia.ca/about/staff/forum/What\\_We\\_Talk\\_About\\_When\\_%20We\\_Talk\\_About.pdf](http://library.concordia.ca/about/staff/forum/What_We_Talk_About_When_%20We_Talk_About.pdf)

[3] TRAMULLAS, Jesús. Software libre para repositorios institucionales: propuestas para un modelo de evaluación de prestaciones. El profesional de la información. 2006, v. 15, nro. 3 [fecha de consulta: 10 de marzo 2013]. Disponible en [http://eprints.rclis.org/9345/1/vol15\\_3.1.pdf](http://eprints.rclis.org/9345/1/vol15_3.1.pdf)

[4] CORRADO, Edward M. The Importance of Open Access, Open Source, and Open Standards for Libraries [en línea]. 2005. [fecha de consulta: 7 de marzo 2013]. Disponible en <http://www.istl.org/05-spring/article2.html>

[5] FOUNIER, Isabel. Descripción de los recursos de información en Internet: formato Dublín Core [en línea]. ACIMED. 2006, v. 14, nro. 4. [fecha de consulta: 9 de marzo 2013]. Disponible en [http://bvs.sld.cu/revistas/aci/vol14\\_4\\_06/aci09406.htm](http://bvs.sld.cu/revistas/aci/vol14_4_06/aci09406.htm)

[6] KEEFER, Alice. Los repositorios digitales universitarios y los autores [en línea]. Anales de documentación. 2007, nro. 10. [fecha de consulta: 11 de marzo 2013]. Disponible en <http://revistas.um.es/analesdoc/article/download/1151/1201>

[7] HERNANDEZ ARIAS, Aymara. La tecnología de Trabajo Colaborativo en el contexto universitario [en línea]. Compendium. [fecha de consulta: 11 de

marzo 2013]. Disponible en <http://www.ucla.edu/ve/dac/compendium/compendium6/Tecnologia%20de%20trabajo%20colaborativo.htm>

## Consolidación de un Modelo para Bases de Datos no Convencionales

Jorge Arroyuelo, Susana Esquivel, Alejandro Grosso, Verónica Ludueña, Nora Reyes  
Dpto. de Informática, Fac. de Cs. Físico-Matemáticas y Naturales, Universidad Nacional de San Luis  
{bjarroyu, esquivel, agrosso, vlud, nreyes}@unsl.edu.ar

Edgar Chávez

Escuela de Ciencias Físico-Matemáticas, Universidad Michoacana de San Nicolás de Hidalgo  
elchavez@umich.mx

Gonzalo Navarro

Departamento de Ciencias de la Computación, Universidad de Chile.

gnavarro@dcc.uchile.cl

### Resumen

Por la evolución de las tecnologías de información y comunicación, han surgido en la actualidad aplicaciones no tradicionales sobre bases de datos que contienen datos no estructurados tales como texto libre, imágenes, audio, video, secuencias de ADN, etc., provenientes de diversas fuentes como revistas, transacciones financieras, fotografías, música, etc., Además estos datos *multimedia* no pueden ser consultados de manera significativa en el sentido clásico, todas las consultas son por objetos similares a uno dado. Estos escenarios requieren modelos más generales, como las *Bases de Datos Métricas*, con una madurez semejante al de las bases de datos tradicionales.

Por otro lado, el desarrollo de memorias más rápidas y de gran capacidad, promovió la aparición de estructuras de datos que tienen en cuenta estas arquitecturas como las *estructuras de datos con I/O eficiente*. Además, en algunos casos los lenguajes de consulta poseen poco poder expresivo para poder expresar todas las consultas consideradas de interés en este modelo de base de datos. Nuestra investigación pretende contribuir a la consolidación de este nuevo modelo de bases de datos.

**Palabras Claves:** bases de datos no convencionales, lenguajes de consulta, índices, expresividad.

### Contexto

En el Proyecto Consolidado 330303 “Tecnologías Avanzadas de Bases de Datos” se encuentra la línea *Bases de Datos no Convencionales*, la cual motiva esta presentación. Este proyecto pertenece a la Univ. Nac. de San Luis y se encuentra dentro del Programa de Incentivos a la Investigación (Código 22/F014). El ámbito de este proyecto ha permitido el estudio y tratamiento de objetos de diversos tipos, útiles en distintos campos de aplicación: sistemas de información geográfica, robótica, visión artificial, compu-

tación móvil, diseño asistido por computadora, motores de búsqueda en internet, computación gráfica, entre otras, y que se relacionan en tales bases de datos. Se consideran como actividades centrales de esta línea el análisis de distintos tipos de bases de datos, la investigación de aspectos empíricos, teóricos y aplicativos derivados de la administración de una base de datos que maneja tipos de datos no convencionales, la expresividad de los lenguajes de consulta, los operadores necesarios para responder consultas de interés, y también las estructuras y operaciones necesarias para resolverlas eficientemente.

El contacto permanente con investigadores de otros países permite nuevas perspectivas en nuestras investigaciones, gracias a la participación de nuestros integrantes en proyectos conjuntos de cooperación internacional con: Universidad de Chile, Universidad de Massey (Nueva Zelanda), Universidad Michoacana de San Nicolás de Hidalgo (México).

### Introducción

La evolución de las nuevas tecnologías dio lugar a diversas aplicaciones en las cuales el modelo clásico de bases de datos no puede aplicarse, ya sea por el tipo de datos que ellas manejan, como por las exigencias de las mismas. Este escenario requiere modelos más generales, además la necesidad de una respuesta rápida y adecuada, y un eficiente uso del espacio disponible, hacen necesaria la existencia de estructuras de datos especializadas que incluyan estos aspectos. Todas estas aplicaciones tienen características comunes, capturadas en el modelo de *espacio métrico*. Formalmente, un espacio métrico consiste de un universo de objetos  $\mathbb{U}$  y una función de distan-



cia definida entre ellos  $d : \mathbb{U} \times \mathbb{U} \mapsto \mathbb{R}^+$  que mide la (di)similitud entre los objetos. En este ámbito las búsquedas exactas carecen de sentido y es importante la elección de este modelo por las *búsquedas por similitud*, más naturales sobre estos tipos de datos.

El trabajo con bases de datos masivas, o con bases de datos que almacenan objetos muy grandes, da lugar a líneas de investigación que, conscientes de estos problemas de costos, diseñan estructuras de datos más eficientes para memorias jerárquicas.

La conocida “*maldición de la dimensionalidad*”, por la cual el desempeño de los índices existentes se deteriora exponencialmente con la dimensión del espacio, afecta tanto a espacios de vectores (representación común de datos multimedia) como a espacios métricos, aunque en éstos últimos no está completamente analizado su efecto sobre los MAMs (*métodos de acceso métricos*). De las numerosas estructuras que existen para búsquedas por similitud en espacios métricos, sólo pocas son eficientes en espacios de alta o mediana dimensión, y la mayoría no admiten dinamismo, ni están diseñadas para conjuntos masivos de datos (en memoria secundaria). Por lo tanto, nos dedicamos a estudiar distintas maneras de optimizarlas.

Además del dinamismo en las estructuras y operaciones de búsqueda complejas, también se están investigando son la obtención de mayor expresividad en los lenguajes utilizados para expresar consultas y caracterizar la clase de consultas computables.

## Líneas de Investigación y Desarrollo

### Bases de Datos Métricas

Tomando como modelo para las bases de datos no convencionales a los espacios métricos, es necesario responder consultas por similitud eficientemente haciendo uso de MAMs. En espacios métricos generales la complejidad usualmente se mide como el número de cálculos de distancias realizados. Por ello, se analizan aquellos MAMs que han mostrado buen desempeño en las búsquedas, para optimizarlos más, considerando la jerarquía de memorias. En general, dada una base de datos  $X \subseteq \mathbb{U}$  y un objeto de consulta  $q \in \mathbb{U}$  las consultas son de dos tipos: por *rango* o de *k-vecinos más cercanos*.

### Métodos de Acceso Métricos

El estudio del *Árbol de Aproximación Espacial* [12], que había mostrado un muy buen desempeño en espacios de mediana a alta dimensión, pero totalmente estático, nos permitió el desarrollo de un nue-

vo índice llamado *Árbol de Aproximación Espacial Dinámico (DSAT)* [13] que permite realizar inserciones y eliminaciones, conservando su buen desempeño en las búsquedas, lo cual es importante porque pocos índices son completamente dinámicos.

El *DSAT* es una estructura que particiona el espacio considerando la proximidad espacial; pero, si el árbol agrupara los elementos muy cercanos entre sí, lograría mejorar las búsquedas al evitar recorrerlo para alcanzarlos. Podemos pensar entonces que construimos un *DSAT*, en el que cada nodo representa un grupo de elementos cercanos (“clusters”) y los relacionamos por su proximidad en el espacio. Cada nodo mantiene el centro del cluster correspondiente, y almacena los  $k$  elementos más cercanos a él; cualquier elemento a mayor distancia del centro que los  $k$  almacenados, forma parte de otro nodo en el árbol [2]. Sin embargo, falta analizar cuán bueno es el agrupamiento o “clustering” que logra esta estructura, lo cual se podría analizar con nuevas estrategias de optimización de funciones a través de heurísticas bioinspiradas, que han mostrado ser útiles en detección de clusters.

Al trabajar sobre base de datos métricas, puede surgir la necesidad de hacer uso de la memoria secundaria. Es posible que la base de datos no pueda almacenarse en memoria principal por ser masiva o porque sus objetos son muy grandes, o que el índice no quepa en memoria principal, o ambas cosas. Por lo tanto, existe la necesidad de diseñar los índices especialmente para memoria secundaria. Así, en [14] se presentaron versiones preliminares del *DSAT* (*DSAT+* y *DSAT\**) especialmente diseñadas para memoria secundaria: índices con buena ocupación de página y eficientes tanto en el número de cálculos de distancia y de operaciones de I/O para cada operación, y se están analizando variantes que mejoren aún más su desempeño. En numerosas aplicaciones es muy importante mantener el total dinamismo de las estructuras, es decir soportar tanto inserciones como eliminaciones de elementos, por lo tanto otro aspecto sobre el que se está trabajando es lograr que las operaciones de inserción y eliminación de elementos en las versiones de memoria secundaria del *DSAT* sean eficientes.

### Join Métricos

El modelo de espacios métricos permite cubrir muchos problemas de búsqueda por similitud, aunque en general se deja fuera de consideración al operador de ensamble o “join” por similitud, otra primitiva importante [5].

De hecho, a pesar de la atención que esta primitiva ha recibido en las bases de datos tradicionales y aún en las multidimensionales, no han habido grandes avances para espacios métricos generales. Nos hemos planteado resolver algunas variantes del problema de join por similitud: (1) *join por rango*: dadas dos bases de datos de un espacio métrico y un radio  $r$ , encontrar todos los pares de objetos (uno desde cada base de datos) a distancia a lo sumo  $r$ , (2) *k-pares más cercanos*: encontrar los  $k$  pares de objetos más cercanos entre sí (uno desde cada base de datos). Para resolver estas operaciones de manera eficiente hemos diseñado un nuevo índice métrico, llamado *Lista de Clusters Gemelos (LTC)* [18], éste se construye sobre ambas bases de datos conjuntamente, en lugar de indexar una o ambas bases de datos independientemente. y permite también resolver las consultas por similitud clásicas sobre cada una de las bases de datos independientemente.

A pesar de que esta estructura ha mostrado ser competitiva y obtener buen desempeño en relación a las alternativas más comunes para resolver las operaciones de join, queda mucho por mejorar para que se vuelva una estructura práctica y mucho más eficiente para trabajar con grandes bases de datos métricas. A la fecha se está analizando la construcción de otra clase de índice basada en “permutantes” para resolver el join aproximado de dos bases de datos métricas; es decir que permita rápida y eficientemente encontrar los pares de elementos más similares entre ambas bases de datos, aunque no los obtenga a todos. Así sería posible extender apropiadamente el álgebra relacional como lenguaje de consulta y diseñar soluciones eficientes para nuevas operaciones, considerando aspectos de memoria secundaria, de concurrencia, de confiabilidad, etc. Algunos de estos problemas ya poseen solución en bases de datos espaciales, pero no en bases de datos métricas.

### Búsqueda aproximada de los *All-k-NN*

Hay muchas otras aplicaciones, tales como la clasificación y aprendizaje automático, donde un nuevo elemento debe ser clasificado de acuerdo a sus vecinos más cercanos, la cuantificación y compresión de imágenes, donde sólo algunos vectores pueden ser representados y los que no deben ser codificados como su punto representable más cercano, la predicción de funciones, en la que desea buscar el comportamiento más similar de una función en el pasado para predecir su comportamiento futuro probable, etc. Todas estas aplicaciones tienen características comunes y necesitan estructuras de datos espe-

cializadas que las incluyan, como las de *espacios métricos*.

Dado que en muchas aplicaciones la evaluación de la función de distancia  $d$  suele ser una operación muy costosa, se usa como medida de complejidad en la mayoría de los casos. Las investigaciones en la actualidad tienden al estudio de algoritmos en espacios métricos generales, donde existen varias técnicas conocidas para resolver el problema de consultas por similitud en un número sublineal de cálculos de distancia, con la condición del preprocesamiento del conjunto de datos.

Si consideramos las búsquedas por similitud vemos que la recuperación de los  $k$ -vecinos más cercanos, es uno de sus primitivos básicos. Este puede definirse como: Sea  $X$  un conjunto de elementos y la función de distancia definida entre ellos  $d$  los  $k$ -NN( $u$ ) son los  $k$  elementos en  $X - \{u\}$  que tengan la menor distancia a  $u$  de acuerdo con la función  $d$ . Una variante de este problema, quizá menos estudiada, es la búsqueda de los  $k$ -vecinos más cercanos de *todos los elementos* de  $X$ , *All-k-NN*, es decir: Sea  $u_i \in X$ , obtener los *All-k-NN* es calcular los  $k$ -vecinos más cercanos para *todos* los  $u_i$  en  $X$ , por supuesto realizando menos de  $n^2$  cálculos de distancia. Así, en el marco de una etapa de investigación previa, se desarrollaron y propusieron soluciones a este problema [17, 16] basadas en la construcción del *Grafo de los k-vecinos más cercanos (kNNG)* para indexar un espacio métrico requiriendo una cantidad moderada de memoria y la utilización del mismo en la resolución de las consultas por similitud en espacios métricos generales. El *kNNG* es un grafo dirigido ponderado que conecta cada elemento del espacio métrico mediante un conjunto de arcos cuyos pesos se calculan de acuerdo a la métrica del espacio en cuestión. El desempeño en las búsquedas por similitud de esta propuesta es superior al obtenido utilizando las técnicas clásicas basadas en pivotes.

Por otro lado, el compromiso de tratar de realizar la menor cantidad de cálculos de distancias posibles durante una búsqueda, ha llevado a investigar un enfoque *aproximado* eficiente para resolver estas consultas por similitud. Este enfoque consiste en permitir una relajación en la precisión de la consulta con el fin de obtener una aceleración en la complejidad del tiempo de consulta [19, 6, 15]. El objetivo de la búsqueda por similitud aproximada es reducir significativamente los tiempos de búsqueda al permitir algunos errores en el resultado de la consulta. Adicionalmente a la consulta se especifica un parámetro

$\varepsilon$  de precisión para controlar cuán lejos queremos el resultado de la consulta del resultado correcto. Un comportamiento razonable para este tipo de algoritmo es acercarse asintóticamente a la respuesta correcta como  $\varepsilon$  se acerca a cero. Por lo tanto, el éxito de una técnica de aproximación se basa en la resolución del compromiso calidad/tiempo [3]. Esta alternativa a la búsqueda por similitud “exacta” se llama *búsqueda de similitud aproximada* [3], y abarca algoritmos aproximados y probabilísticos.

## Lenguajes de Consulta

La relación existente entre lógica y teoría de bases de datos es muy estrecha y natural, ya que es posible pensar en una base de datos simplemente como una estructura finita, y utilizar las lógicas para expresar consultas sobre éstas. Esto les da una posición central como modelo computacional para el análisis del poder expresivo de los lenguajes de consultas que nos permiten obtener información de una base de datos, siendo relevante como marco teórico para el estudio de las bases de datos.

La mayoría de los lenguajes de consulta sobre bases de datos es equivalente, en su poder expresivo, a *FO* (First-Order logic). El principal problema es que la expresividad de *FO* no es lo suficientemente poderosa, porque no alcanza para reflejar ciertas consultas. Esto ha llevado a la búsqueda de una mayor expresividad por medio de diferentes mecanismos de extensión sobre *FO* utilizados como herramientas de construcción de lógicas más poderosas. Uno de ellos gracias a incorporar cuantificadores que no pueden ser expresados en *FO*, como *clausura transitiva* y *punto fijo*, entre otros, los que han sido ampliamente estudiados. La idea de agregar cuantificadores es generalizada mediante la noción de *cuantificadores generalizados de Lindström*[8]. Aún así, estas lógicas todavía resultan incompletas, por lo que se analizan lógicas de orden superior, *SO* (Second-Order Logic), y algunos de sus fragmentos que han demostrado poseer propiedades interesantes sobre las estructuras finitas. Un resultado importante de R. Fagin fue la caracterización del fragmento existencial  $SO\exists$ [7]. Allí se establece que las propiedades de las estructuras finitas que son definidas por sentencias existenciales de segundo orden coinciden con las propiedades de la clase de complejidad NP, lo cual fue extendido por Stockmeyer [20], estableciendo una relación cercana entre la lógica *SO* y la jerarquía de tiempo polinomial (*PH*).

Actualmente existen muchos resultados igualando la expresividad lógica a la complejidad compu-

tacional, pero requieren estructuras ordenadas [9], [10]. Estas relaciones entre la complejidad computacional (cantidad de recursos necesarios para resolver un problema sobre algún modelo de máquina computacional) y la complejidad descriptiva (el orden de la lógica que se necesita para describir el problema), han llevado a que los resultados obtenidos en alguno de estos campos sea transferido de manera inmediata al otro.

En uno de nuestros trabajos de investigación se ha introducido la definición de una restricción de *SO*, que consiste en limitar las relaciones que pueden tomar los cuantificadores de *SO*, considerando a la lógica como uno de los lenguajes de consulta a base de datos. El tipo de relaciones a los que estos cuantificadores pueden referirse son relaciones cerradas bajo *FO-type*. Esta lógica (*SOF*) intenta lograr una lógica de mayor poder expresivo que la definida por Dawar ( $SO^w$ ) en la que los cuantificadores sólo pueden tomar relaciones cerradas bajo *FO-k* tipos. Se demostró que nuestra lógica incluye estrictamente la de Dawar [4]. Se ha podido definir una nueva clase de complejidad descriptiva (*NPF*), que caracteriza el fragmento existencial de nuestra lógica gracias a modificar de las máquinas relacionales.

En otro de nuestros trabajos se estudia el impacto del aumento del orden de las variables en las lógicas. Se continúa con el estudio del poder expresivo de las lógicas *HO* (High-Order logic) y en particular de los fragmentos de la lógica *VO* (Variable-Order logic) definida en [11], que nace debido a que ninguna de las lógicas de orden superior cubre la clase completa de consultas computables (*CQ*)[1], es decir que no son completas. Además, si consideramos la unión de todas las lógicas de orden superior, es decir  $\bigcup_{i \geq 2} HO^i$  ( $HO^i$  representa la lógica de orden  $i$ ), tampoco obtenemos una lógica completa. De aquí, se define *VO* permitiendo el uso de variables de orden variable, mediante el uso de cuantificadores de orden. Las restricciones más importantes estudiadas sobre *VO* son sobre: la cantidad de alternaciones de cuantificadores, la aridez de las variables de orden variable, los valores que pueden asignarse a las variables de orden en función del tamaño del dominio, el rango de cuantificadores y la cantidad de variables, de valuación y de orden.

## Resultados y Objetivos

Como trabajo futuro de esta línea de investigación se consideran varios aspectos relacionados al diseño de estructuras de datos que, consciente de la jerar-



quía de memorias y de las características particulares de los datos a ser indexados, saquen el mejor partido haciéndolas eficientes en espacio y en tiempo.

Respecto de los lenguajes de consulta se continuará analizando la expresividad de distintas extensiones de FO y posibles restricciones de SO, para lograr caracterizar la clase de las consultas computables sobre bases de datos no convencionales.

En el caso de bases de datos métricas, se intentará que los índices se adapten mejor al espacio métrico particular considerado, por la determinación de su dimensión intrínseca, y también al nivel de la jerarquía de memorias donde se almacenará. Estos estudios sobre espacios métricos y sobre algunas estructuras de datos particulares permitirán no sólo mejorar el desempeño de las mismas sino también aplicar, eventualmente, muchos de los resultados que se obtengan a otros MAMs.

## Actividades de Formación

Dentro de esta línea de investigación se están formando alumnos y docentes-investigadores de acuerdo al siguiente detalle:

**Doctorado en Cs. de la Computación:** dos integrantes de la línea se encuentran desarrollando su tesis sobre la expresividad de la lógica como lenguaje de consulta. Otro integrante desarrolla su tesis sobre bases de datos métricas.

**Maestría en Cs. de la Computación:** un investigador de la línea está desarrollando su tesis en bases de datos métricas sobre búsqueda por similitud aproximada.

Además, se están dirigiendo actualmente tres trabajos finales de alumnos de la Licenciatura en Cs. de la Computación, prontos a finalizar.

## Referencias

- [1] D. Harel A. K. Chandra. Computable queries for relational data bases. *J. of Computer and System Sciences*, 21(2):156–178, 1980.
- [2] M. Barroso, N. Reyes, and R. Paredes. Enlarging nodes to improve dynamic spatial approximation trees. In *Procs. of the 3rd Int. Conf. on Similarity Search and Applications*, 41–48. ACM Press, 2010.
- [3] P. Ciaccia and M. Patella. Approximate and probabilistic methods. *SIGSPATIAL Special*, 2(2):16–19, 2010.
- [4] A. Dawar. A restricted second order logic for finite structures. *Information and Computation*, 143:154–174, 1998.
- [5] V. Dohnal, C. Gennaro, P. Savino, and P. Zezula. Similarity join in metric spaces. In *Proc. 25th European Conf. on IR Research*, LNCS 2634, 452–467, 2003.
- [6] R. Baeza-Yates E. Chávez, G. Navarro and J. Marroquín. Searching in metric spaces. *ACM Computing Surveys*, 33(3):273–321, 2001.
- [7] R. Fagin. Generalized first-order spectra and polynomial-time recognizable sets. *Complexity of Computation*, 7:43–73, 1974.
- [8] J. Flum. H. Ebbinghaus. Finite model theory, second edition. *Springer*, 1999.
- [9] N. Immerman. Descriptive and computational complexity. *Computational Complexity Theory*, 38:75–91, 1989.
- [10] N. Immerman. Descriptive complexity. *Springer*, 1998.
- [11] J. M. Turull Torres L. Hella. Computing queries with higher-order logics. *Theoretical Computer Science*, 355:197–214, 2006.
- [12] G. Navarro. Searching in metric spaces by spatial approximation. *The Very Large Databases Journal (VLDBJ)*, 11(1):28–46, 2002.
- [13] G. Navarro and N. Reyes. Dynamic spatial approximation trees. *J. of Experimental Algorithms*, 12:1–68, 2008.
- [14] G. Navarro and N. Reyes. Dynamic spatial approximation trees for massive data. In Tomás Skopal and Pavel Zezula, editors, *SISAP*, 81–88. IEEE Computer Society, 2009.
- [15] V. Dohnal P. Zezula, G. Amato and M. Batko. Similarity search: The metric space approach. *Advances in Database Systems*, 32.
- [16] R. Paredes. *Graphs for Metric Space Searching*. PhD thesis, University of Chile, 2008.
- [17] R. Paredes, E. Chávez, K. Figueroa, and G. Navarro. Practical construction of  $k$ -nearest neighbor graphs in metric spaces. In *Proc. 5th Workshop on Efficient and Experimental Algorithms*, LNCS 4007, 85–97, 2006.
- [18] R. Paredes and N. Reyes. Solving similarity joins and range queries in metric spaces with the list of twin clusters. *J. of Discrete Algorithms*, 7(1):18–35, 2009.
- [19] H. Samet. Foundations of multidimensional and metric data structures. *Morgan Kaufmann*, 2006.
- [20] L. Stockmeyer. The polynomial-time hierarchy. *Theoret. Comput. Sci.*, 3:1–22, 1976.



## Recuperación Eficiente de Datos en Bases de Datos Multimedia

Luis Britos, María E. Di Gennaro, Jacqueline Fernández, Veronica Gil-Costa, Fernando Kasián, Verónica Ludueña, Marcela Printista, Nora Reyes, Patricia Roggero

LIDIC, Dpto. de Informática, Fac. de Cs. Físico Matemáticas y Naturales, Universidad Nacional de San Luis  
{lebritos, mdigena, jmfer, gvcosta, fkasian, vlud, mprinti, nreyes, proggero}@unsl.edu.ar

Edgar Chávez

Escuela de Cs. Físico-Matemáticas, Universidad Michoacana de San Nicolás de Hidalgo  
elchavez@umich.mx

Claudia Deco

Facultad de Ciencias Exactas, Ingeniería y Agrimensura, Universidad Nacional de Rosario  
deco@fceia.unr.edu.ar

### Resumen

Los sistemas de información necesitan realizar de manera eficiente la búsqueda de los datos a través de estructuras de almacenamiento. Dicha búsqueda consiste en obtener los elementos que cumplen una serie de condiciones. En general, es tan difícil para los usuarios que intentan recuperar información multimedia poder especificar claramente sus intereses, o las condiciones a cumplir, a través de una consulta bien definida, como para los diseñadores del sistema decidir qué características de los objetos multimedia pueden ser relevantes. Dada una consulta, el objetivo de un sistema de recuperación de información es obtener lo que podría ser útil o relevante para el usuario, en general usando un índice especialmente diseñado para responder eficientemente la consulta.

Así, nuestra línea de investigación tiene como principal objetivo desarrollar herramientas eficientes para la recuperación de información multimedia. Se investigan nuevas técnicas que soporten la interacción con el usuario, nuevas estructuras de datos (índices) capaces de manipular eficientemente datos multimedia y que permitan manejar grandes volúmenes de este tipo de datos.

**Palabras Claves:** Recuperación de Información, Bases de Datos Multimedia, Indexación, Paralelismo.

### Contexto

Esta línea de investigación se encuentra enmarcada dentro del Proyecto Consolidado 30310 de la Universidad Nacional de San Luis y en el Programa de Incentivos (código 22/F034): “Nuevas Tecnologías para el Tratamiento Integral de Datos Multimedia”, dentro de la línea “Recuperación de Datos e Información Multimedia”, desarrollada en el ámbito del Laboratorio de Investigación y Desarrollo en Inteligencia Computacional (LIDIC) de la UNSL.

En este marco se pretende avanzar en la integra-

ción de las investigaciones sobre adquisición, procesamiento y análisis de datos no estructurados y su aplicación en dominios no convencionales. Como principal aporte de esta propuesta se pretende incorporar información no estructurada en los procesos de toma de decisiones y resolución de problemas que quedan sin considerar en los enfoques clásicos.

Dentro de este contexto nuestra línea se dedica, principalmente, al diseño de índices eficientes que sirvan de apoyo a sistemas de recuperación de información orientados a datos no estructurados, en particular datos multimedia. Se espera así contribuir a estos sistemas obteniendo índices más eficientes para memorias jerárquicas, dinámicos, con I/O eficiente y capaces de manejar grandes volúmenes de datos. Se propone analizar las estructuras de datos existentes, proponer optimizaciones o diseñar nuevas estructuras, para manipular y recuperar algunos de los tipos de datos no estructurados que aparecen en entornos multimedia, considerando el uso de técnicas de computación de alto desempeño en los mismos, con el objetivo de lograr una recuperación eficiente.

### 1. Introducción y Motivación

Los sistemas de computación hacen uso intensivo de información estructurada, es decir datos elementales o estructuras, generadas con un formato específico. Una característica principal en estos casos, es que la estructura o formato de esta información puede ser fácilmente interpretada y directamente utilizada por un programa de computadora. Pero el hecho de restringirse al uso de este tipo de información conduce, muchas veces, a representar una visión parcial del problema y dejar fuera información

que podría ser importante para la resolución efectiva del mismo. En este contexto gran parte de la información que se requiere para la toma de decisiones y la resolución de problemas de índole general proviene de información no estructurada.

En general, para responder eficientemente consultas para recuperación de información sobre bases de datos multimedia se utilizan diferentes métodos de acceso o índices [9], principalmente por el volumen de datos con el que se trabaja. Algunos poseen características que los hacen indicados para aplicaciones reales: eficientes, dinámicos, escalables y resistentes a la *maldición de la dimensión*.

Un enfoque prometedor para sistemas de recuperación usando búsqueda por similitud es una búsqueda basada en contenidos, la cual usa el dato multimedia mismo. Para calcular la similitud entre dos objetos multimedia, se debe definir una función de distancia. Dicha función mide la similitud, o más bien la disimilitud, entre dos objetos. En muchos casos para modelar la similitud de objetos multimedia se transforman los objetos en puntos de un espacio vectorial, el cual es un tipo particular de espacio métrico. Cada objeto es representado por un vector de características o descriptor, generalmente de alta dimensionalidad. Sobre espacios vectoriales existen numerosas funciones de distancia. El tipo de aplicación, las características a explotar o la dimensionalidad son aspectos a considerar para definir la mejor función de distancia a utilizar.

El concepto de búsqueda por similitud se puede definir a partir del de espacios métricos, que da un marco formal independiente del dominio de aplicación. Un espacio métrico está compuesto por un universo  $\mathcal{U}$  de objetos y una función de distancia  $d : \mathcal{U} \times \mathcal{U} \rightarrow \mathbb{R}^+$ , que satisface las propiedades que la hacen una métrica. Las consultas por similitud, sobre una *base de datos*  $\mathcal{S} \subseteq \mathcal{U}$ , son usualmente de dos tipos: *Búsqueda por rango* y *Búsqueda de los  $k$  vecinos más cercanos*.

En el caso de los espacios métricos, la función de similitud (distancia) mide el mínimo esfuerzo (costo) necesario para transformar un objeto en otro. Dependiendo de los tipos de datos multimedia reales la función de similitud puede ser muy compleja. En particular, para poder ahorrar cálculos de distancia es importante que la distancia satisfaga la propiedad triangular. Si la base de datos  $\mathcal{S}$  posee  $n$  objetos, las consultas pueden ser respondidas llevando a cabo  $n$  evaluaciones de distancia. Sin embargo, en la mayoría de las aplicaciones las distancias son costosas

de computar (por ej.: comparación de huellas digitales), por lo que la búsqueda secuencial no sirve para problemas de tamaño medio o grande, que son los tamaños más habituales de las bases de datos multimedia. Así debemos preprocesar la base de datos, construyendo un índice, para responder a las consultas con la menor cantidad de cálculos de distancia. Además, es probable que la base de datos, el índice o ambos no puedan almacenarse en memoria principal con lo cual se debe considerar minimizar también el número de operaciones de E/S, tener siempre presente la jerarquía de memorias y tratar de lograr mayor eficiencia a través de técnicas paralelas.

En suma, esta propuesta se enfoca en mejorar las herramientas de recuperación desarrollando nuevas técnicas que soporten la interacción con el usuario, diseñando estructuras de datos (índices), capaces de manipular eficientemente grandes volúmenes de datos multimedia y facilitando la realización de operaciones, de modo de acercarse al nivel de desarrollo que poseen las bases de datos tradicionales.

## 2. Líneas de Investigación

Se pretende investigar sobre distintos aspectos de los sistemas de recuperación de información multimedia: diseñar nuevos índices, definir representaciones que reflejen características de interés de los objetos y manejar distintas operaciones sobre estos tipos de bases de datos, considerando trabajar eficientemente sobre grandes volúmenes de datos.

### Diseño de Índices

Un catálogo importante de índices para espacios métricos aparece en [9]. La mayoría usan la desigualdad triangular para evitar el análisis secuencial de la base de datos. La distancia entre la consulta  $q$  y los objetos de la base de datos puede ser estimada calculando de antemano algunas distancias a objetos distinguidos llamados *pivotes* y sin calcular las distancias reales desde  $q$  a los objetos de la base de datos durante una búsqueda. Otra técnica común es indexar a través de una partición del espacio en regiones denominadas *particiones compactas*.

Existen dos posibles situaciones por el tipo de base de datos con la que se va a trabajar, que determinan una característica importante que debe tener el índice que la manipulará: los objetos de la base de datos se conocen de antemano y por lo tanto el índice se creará de una sola vez y se realizarán consultas sobre él (índices estáticos); o no se conocen los objetos de antemano y por lo tanto el índice se debe ir

creando a medida que arriban los elementos y preferentemente de manera incremental (índices dinámicos). Las estructuras estáticas se benefician desde el conocimiento de la base de datos seleccionando los mejores puntos de referencia para una estructura de datos determinada, lo cual no es posible en las estructuras de datos dinámicas donde tanto los objetos como las consultas arriban al azar.

Cuando se trabaja sobre grandes volúmenes de datos una manera de lograr eficiencia en las operaciones sobre los índices es aplicando técnicas de computación de alto desempeño y en otros casos mediante la adaptación o diseño de las estructuras para que sean concientes de la jerarquía de memorias, minimizando no sólo la cantidad de cálculos de la función de distancia, sino también el número de operaciones de E/S. Además, el estudio detallado de qué hace que un índice sea intrínsecamente más eficiente ayuda a mejorar los costos de las operaciones.

### Índices Estáticos

En este caso, al conocer de antemano los elementos a indexar, es posible elegir con más información cómo hacerlo de manera tal que las búsquedas sean eficientes. Sin embargo, hay ejemplos como el del *Árbol de Aproximación Espacial*, SAT, que por ser una estructura estática debería ser más eficiente que la versión dinámica, el DSAT [13], y no lo es. En estos casos ha sido posible investigar alguno de los motivos por los que la versión dinámica, usando menos información, proporciona búsquedas más rápidas. En nuestras investigaciones hemos detectado que una condición clave para mejorar la performance de SAT es modificar la estrategia de selección de vecinos, es por ello que se ha estado trabajando en diferentes heurísticas, como la de utilizar un orden de inserción arbitrario de los vecinos o hasta elegirlos de manera totalmente contraria a lo que la versión original lo hacía y se están consiguiendo en este caso resultados muy interesantes [6]. El análisis detallado del comportamiento de esta estructura nos ha llevado a demostrar que el SAT se beneficia significativamente al mejorar la distribución de los vecinos de un nodo y por lo tanto los hiperplanos y radios de cobertura resultantes, tan importantes para podar nodos o subárboles durante las búsquedas, son mejores. El DiSAT, ha demostrado no sólo superar a la versión original de la estructura, sino también a los mejores representantes del estado del arte, como es la *Lista de Clusters* [8], y además ser escalable y ser resistente a la maldición de la dimensión [7].

### Índices Dinámicos

Aquí el interés está en mejorar el desempeño de índices dinámicos jerárquicos (árboles), que es el caso de algunos de los índices para espacios métricos. Estos índices dinámicos, en general, se construyen incrementalmente vía inserciones. De tal manera, la raíz del árbol es el primer objeto que llega, y esto se repite recursivamente en cada nivel del árbol.

Como ya mencionamos, el DiSAT es un índice para espacios métricos muy eficiente, considerando el número de cálculos de distancias realizados tanto en construcción como en búsquedas. La desventaja del DiSAT es que no admite inserciones ni eliminaciones y no es posible construirlo incrementalmente. Sin embargo, una opción para transformarlo en una estructura dinámica es mediante la aplicación de la técnica de Bentley y Saxe [2] que permite lograr dinamismo a partir de una estructura estática cuando la búsqueda sobre ella cumple con ser un problema que se puede descomponer en partes independientes (descomponible). Por lo tanto, se está desarrollando una versión dinámica del DiSAT, que además pueda resolver eficientemente las búsquedas de los  $k$ -vecinos más cercanos de cualquier objeto, búsqueda que no cumple con ser un problema descomponible.

También en esta línea se ha propuesto una técnica donde el “buffering” logra un buen compromiso entre una estructura estática, construida con toda la información necesaria y una dinámica con conocimiento local de los datos. Entonces, en lugar de elegir al primer elemento como la raíz, se demora la selección hasta que hayan arribado suficientes elementos para estar en condiciones de realizar dicha selección, y de esta manera se toma una decisión en base a más información. Dado que las consultas arriban a un ritmo desconocido, para mantener el dinamismo es necesario contar con un índice que responda a las consultas con mejor desempeño que la técnica de fuerza bruta. Esta técnica provee un marco adecuado para diseñar estructuras de datos dinámicas estables. Por lo tanto, un “buffer” en todos los niveles de una estructura jerárquica podría ser útil en estrategias de ruteo para guiar las búsquedas, lo cual resulta un área promisoría de investigación [10].

En muchos casos los volúmenes de información con los que se debe trabajar (millones de imágenes en la Web), hacen necesario que los índices sean almacenados en memoria secundaria. En este caso, para hacerlos eficientes, no sólo se debe considerar que durante las búsquedas se realice el menor número de cálculos de distancia sino también, dado el costo

de las operaciones sobre disco, se efectúe la menor cantidad posible de operaciones de E/S. Por ello, en esta línea nos hemos dedicado a diseñar índices especialmente adaptados para trabajar en memoria secundaria, logrando un buen desempeño de los mismos, principalmente en las búsquedas.

Hemos diseñado e implementado las siguientes estructuras *DSACL\*-tree* y el *DSACL+-tree* [4], las cuáles son optimizaciones para memoria secundaria de la estructura propuesta en [1] y demostraron ser competitivas frente a otras de las estructuras conocidas tales como el *M-tree* y *DSA\*-tree* y *DSA+-tree* [13]. Además, existen nuevas propuestas en evaluación que prometen ser aún más adecuadas para memoria secundaria. Por otro lado, nos proponemos optimizarlas todavía más gracias a la aplicación de técnicas de computación de alto desempeño, aplicando y comparando distintas estrategias de paralelización con el fin de determinar la más adecuada.

### Índices para Sistemas Paralelos

El desarrollo de nuevas tecnologías permite combinar las facilidades provistas por los sistemas de memoria compartida y los sistemas de memoria distribuida (cluster de nodos o procesadores). Es decir, permiten acelerar las operaciones de cómputo al contar con una mayor cantidad de cores (núcleos) y aumentar la capacidad de almacenamiento. Esta combinación de diferentes jerarquías o niveles de hardware, permite ejecutar  $P$  procesos en diferentes nodos conectados mediante una red de intercomunicación y dentro de cada proceso ejecutar  $PT$  threads (hilos) en paralelo, cada uno en un “core” diferente utilizando un esquema en el cual todos los threads tienen acceso a la misma memoria principal. El desafío en este tipo de arquitecturas es reducir el tiempo de ejecución de los programas, lo cual para algunas aplicaciones como las búsquedas de texto y búsqueda multimedia sobre espacios métricos puede ser aún más difícil de lograr, debido a las competencias por los recursos causadas por el acceso concurrente de los threads a los datos compartidos.

Para afrontar este desafío exitosamente, es necesario estudiar y diseñar diferentes técnicas de particionado de índices sobre el sistema de memoria distribuido y diseñar estrategias de asignación de trabajo a los diferentes threads del sistema de memoria compartida. Estas técnicas deben explotar las características inherentes de los índices para maximizar el cómputo paralelo en forma balanceada y reducir los costos de comunicación y sincronización.

En particular, el trabajo presentado en [12] evalúa

diferentes métricas como la eficiencia, *throughput* y *speedup* de diferentes técnicas de particionado sobre un índice métrico denominado *Sparse Spatial Selection* (SSS) [3]. Por otro lado, en el trabajo presentado en [14] se utilizan herramientas de evaluación de performance para evaluar la eficiencia en el uso de jerarquías de memorias caché de un sistema multi-core sobre el índice de Arreglos de Sufijos.

### Consultas sobre Bases de Datos Multimedia

Aunque las operaciones más comunes sobre bases de datos multimedia son las búsquedas por rango o de  $k$ -vecinos más cercanos, existen otras operaciones de interés como las distintas variantes del *join* por similitud. Para estas operaciones se consideran dos bases de datos  $A$  y  $B$ , ambas subconjuntos del mismo universo del espacio métrico  $\mathcal{U}$ . El resultado de cualquier operación de *join* por similitud entre  $A$  y  $B$  obtiene el conjunto de pares formados por un objeto de  $A$  y otro de  $B$ , tales que entre ellos se satisface el predicado de similitud considerado. Las variantes más conocidas del *join* por similitud son: el *join* por rango, el *join* de  $k$ -vecinos más cercanos y el *join* de vecino más cercano; entre otras.

Formalmente, dadas  $A, B \subseteq \mathcal{U}$ , se define el *join por similitud* entre  $A$  y  $B$  ( $A \bowtie_{\Phi} B$ ) como el conjunto de todos los pares  $(x, y)$ , donde  $x \in A$  e  $y \in B$ ; es decir,  $(x, y) \in A \times B$ , tal que entre  $x$  e  $y$  se satisface el criterio de similitud considerado  $\Phi$ . De acuerdo al criterio de similitud el *join* puede llamarse: *Join por rango* o *Join de  $k$ -vecinos más cercanos*.

Existen dos situaciones distintas sobre las que se puede trabajar, para resolver el *join* por similitud: que ambas bases de datos se encuentren indexadas, por separado; o que ambas bases de datos se indexen conjuntamente, con un índice diseñado para el *join*. Como calcular cualquiera de las variantes del *join* por similitud de manera exacta es muy costoso [15], vale la pena analizar posibilidades de obtener más rápidamente una respuesta aproximada al *join*, logrando una respuesta rápida y de buena calidad.

*PostgreSQL* es el primer sistema de base de datos que permite realizar consultas por similitud sobre algunos atributos, particularmente indexación para búsquedas de  $k$ -vecinos más cercanos (KNN-GiST indexes). Estos índices pueden ser usados sobre texto, comparación de ubicación geoespacial, etc.. Sin embargo, los índices K-NN GiST proveen plantillas para índices con estructura de árbol balanceado (B-tree, R-tree), aunque el balance no siempre es bueno para los índices que se utilizan en búsquedas por si-



militud [5]. Además, no se dispone de este tipo de consultas para todo tipo de datos métricos. Así, es importante proveer un DBMS para bases de datos métricas que maneje todos los posibles datos métricos y las operaciones de interés sobre ellos [11].

### 3. Resultados

Se ha comprobado experimentalmente que las estrategias de “buffering” mejoran el desempeño en un índice dinámico [10]. Se seleccionó el Árbol de Aproximación Espacial Dinámico (DSA-tree) [13] y se obtuvo una mejora sistemática en los costos de las consultas. En particular, se verificó que el DSAT es mejor que su versión estática [13], por dejar como “vecinos” a objetos alejados, permitiendo así avanzar en la exploración espacial a “pasos más grandes”, obteniendo así el DiSAT [7].

Se implementaron dos versiones: DSACL\*-tree y DSACL+-tree, que trabajan con grandes volúmenes de datos, por haber sido diseñadas para memoria secundaria y que mostraron ser competitivas contra otras estructuras diseñadas para tal fin [4]. Se espera lograr una implementación paralela eficiente de estos índices. Se ha logrado mayor eficiencia en otros índices [12, 14] gracias al uso de técnicas paralelas.

### 4. Formación de Recursos

Para contribuir al desarrollo de sistemas de recuperación de información multimedia, se están capacitando los siguientes investigadores:

**Tesis de Doctorado en Cs. de la Computación:** un integrante se encuentra definiendo su plan de doctorado sobre diseño y optimización de índices, para aplicaciones de minería de datos multimedia.

**Tesis de Maestría en Cs. de la Computación:** una sobre índices dinámicos para búsqueda eficiente en memoria principal, dos sobre índices dinámicos eficientes para datos masivos: una que aplica técnicas de computación de alto desempeño (con beca de posgrado-UNSL) y la otra sólo aplicando optimizaciones considerando la jerarquía de memorias, y una sobre un DBMS para bases de datos métricas.

### Referencias

- [1] M. Barroso, N. Reyes, and R. Paredes. Enlarging nodes to improve dynamic spatial approximation trees. In *Proc. of the 3rd SISAP*, pages 41–48. ACM Press, 2010.
- [2] J. Bentley and J. Saxe. Decomposable searching problems i: Static-to-dynamic transformation. *J. Algorithms*, 1(4):301–358, 1980.
- [3] N. R. Brisaboa, A. Fariña, O. Pedreira, and N. Reyes. Similarity search using sparse pivots for efficient multimedia information retrieval. In *Proc. of the ISM*, 881–888. IEEE, 2006.
- [4] L. Britos, A. M. Printista, and N. Reyes. Dynamic spatial approximation trees with clusters for secondary memory. *XVI CACIC Selected Papers*, 205–215. Editorial UNLP, 2011.
- [5] E. Chávez, V. Ludueña, and N. Reyes. Revisiting the VP-forest: Unbalance to improve the performance. In *Proc. de las JCC08*, 26, 2008.
- [6] E. Chávez, V. Ludueña, N. Reyes, and P. Roggero. Reaching near neighbors with far and random proxies. In *CCE, 8th Int. Conf. on*, 1–8, oct. 2011.
- [7] E. Chávez, V. Ludueña, N. Reyes, and P. Roggero. Faster proximity searching with the distal sat. Manuscrito IEEE TPAMI-2012-12-0969. 2012.
- [8] E. Chávez and G. Navarro. A compact space decomposition for effective metric indexing. *Pattern Recognition Letters*, 26(9):1363–1376, 2005.
- [9] E. Chávez, G. Navarro, R. Baeza-Yates, and J. Marroquín. Searching in metric spaces. *ACM*, 33(3):273–321, sep 2001.
- [10] E. Chávez, N. Reyes, and P. Roggero. Delayed insertion strategies in dynamic metric indexes. In *Procs. of the SCCC*, 34–42. IEEE Computer Society, 2009.
- [11] F. Kasián and N. Reyes. Búsquedas por similitud en PostgreSQL. In *XVIII CACIC*, pages 1098–1107, 2012.
- [12] V. Mancini, F. Bustos, V. Gil-Costa, and A.M. Printista. Data partitioning evaluation for multimedia systems in hybrid environments. In *P2P, Parallel, Grid, Cloud and Internet Computing, 7th Int. Conf. on*, 321–326, 2012.
- [13] G. Navarro and N. Reyes. Dynamic spatial approximation trees. *J. of Exp. Algorithmics*, 12:1–68, 2008.
- [14] C. Ochoa, V. Gil-Costa, and M. Printista. Suffix array performance analysis for multi-core platforms. In *4th Int. SuperComputing Conf. in Mexico*, 2013. Por aparecer.
- [15] R. Paredes and N. Reyes. Solving similarity joins and range queries in metric spaces with the list of twin clusters. *JDA*, 7:18–35, 2009.

## Aplicando estrategias y tecnologías de Inteligencia de Negocio en sistemas de gestión académica

Francisco Javier Díaz, M. Alejandra Osorio, A. Paola Amadeo, Dalila L. Romero  
Laboratorio de Investigación en Nuevas Tecnologías Informáticas.  
Facultad de Informática. Universidad Nacional de La Plata  
Calle 50 esq. 120, 2do Piso  
+54 221 4223528  
jdiaz@unlp.edu.ar, {aosorio, pamadeo, dalila\_r}@cespi.unlp.edu.ar

### Resumen

El siguiente artículo presenta una línea de investigación que tomo identidad propia a partir del año 2011, como resultado de actividades previas relacionadas con la aplicación de herramientas y técnicas de inteligencia de negocio a datos almacenados en los sistemas académicos de gestión universitaria, que se utilizan para la operativa diaria en las unidades académicas. Todos ellos basados en software libre o con acceso. El cruce con la información relevada de la actividad de los estudiantes en las redes sociales abre un campo de investigación muy interesante, para analizarla en conjunto con su desempeño académico, su comportamiento en las Plataformas Virtuales de Aprendizaje, en el uso de la biblioteca, etc.

Este análisis permitirá analizar el perfil del estudiante de la Facultad de Informática de la UNLP, carreras y planes de estudio implementados, la actividad de los docentes en relación al desenvolvimiento de las materias que dictan, permitiendo realizar acciones para resolver distintas situaciones identificadas a partir de datos confiables y fiables. Esta línea de investigación ha promovido la propuesta de dictado de una materia optativa relacionada con el tema a partir del año 2011, con amplia aceptación por parte de los estudiantes y una tesina de Licenciatura.

**Palabras clave:** Inteligencia de Negocios, Sistemas de Gestión Académica, Redes Sociales

### Contexto

El Laboratorio de Investigación en nuevas Tecnologías Informáticas de la Facultad de Informática de la Universidad Nacional de La Plata incluye dentro de sus líneas de investigación el área de Análisis de la Información. Esta área intenta dar respuesta a la creciente demanda de analistas de información que permitan generar conocimiento a partir de los sistemas de gestión operativa que se utilizan hoy en día en toda organización, con distintos fines.

En particular, en la Facultad de Informática se utiliza el sistema SIU Guaraní desde el año 2005, gestionando la información académica de más de 4.151 alumnos en el año 2011, más de 700 ingresantes y graduados según el Anuario Estadístico de la UNLP [1]. El proyecto de Data Warehouse del Ministerio de Educación de la Nación [2] provee un sistema de análisis de información de sistemas académicos, basado inicialmente en O3 [3] y en Pentaho BI [4], una solución open source para inteligencia de negocios.

Asimismo se utiliza el sistema de código abierto Moodle [5] para la gestión de cursos virtuales que complementan la actividad presencial. Este sistema está integrado con el sistema SIU Guaraní a través de una interfaz común, que facilita la gestión de las inscripciones a trabajos prácticos entre ambos sistemas. Este desarrollo fue propuesto inicialmente por el Consorcio SIU del Ministerio de Educación de la Nación [6], y la integración con Moodle 2 fue desarrollada por el equipo de implementación de Sistemas

Académicos del Centro Superior para el Procesamiento de la Información [7] de la UNLP. El sistema Moodle se utiliza en la Facultad desde el año 2005, incluyendo más de 310 cursos y más de 14000 usuarios. Puede consultarse a través de <http://catedras.info.unlp.edu.ar>, <http://cursos.linti.unlp.edu.ar> y <http://postgrado.linti.unlp.edu.ar>. La información almacenada en este sistema involucra entregas de tareas, participación en los foros, acceso a los diferentes recursos y demás registros los cuales, al cruzarlos con el desenvolvimiento académico pueden resultar en datos significativos para la toma de decisiones de los directivos de la Facultad.

Por otra parte, para la gestión de la biblioteca se utiliza el sistema de software libre Meran [8], desarrollado por el equipo de desarrollo del CeSPI a partir de Koha [9], el sistema de SL para la gestión de bibliotecas implementado en distintas facultades de la UNLP a partir del año 2003. Este sistema permite gestionar los procesos bibliotecarios y los servicios a los usuarios, como estantes virtuales para las cátedras, la posibilidad de votar un libro o dar recomendaciones. Todo esto integrado también al sistema SIU Guaraní.

Finalmente, la Facultad de Informática de la UNLP está haciendo uso también de las redes sociales para poder llegar a los distintos sectores de la sociedad que están vinculados con ella y de esta manera realizar una mejor difusión de las actividades que se llevan a cabo. En particular dispone de un perfil en Facebook,

<http://www.facebook.com/facultad.informatic.a.unlp> y distintas cuentas de difusión académica como @InformaticaUNLP con más de 1100 seguidores, @infounlp con más de 500 seguidores entre otras. Además ciertas cátedras utilizan estos canales para comunicarse con sus estudiantes, como la cátedra de Algoritmos y Estructuras de Datos @ayed\_fi, Introducción a los Sistemas Operativos @iso\_info\_unlp con más de 100 seguidores, Sistemas Operativos, Tecnologías Aplicadas a Business Intelligence

@catedraBI con más de 20 seguidores, entre otros.

Como se puede observar, se cuentan con distintas fuentes de información relacionadas con el mismo sujeto, el estudiante, que consolidadas en único repositorio permitirá realizar análisis transversales, que sirvan de insumo a grupos interdisciplinarios con distintos perfiles.

## **Introducción**

Business Intelligence, es la habilidad para transformar los datos en información, y la información en conocimiento, de forma que se pueda optimizar el proceso de toma de decisiones en los negocios. Asociándolo directamente con las tecnologías de la información, podemos definir Business Intelligence como el conjunto de metodologías, aplicaciones y tecnologías que permiten reunir, depurar y transformar datos de los sistemas transaccionales e información desestructurada (interna y externa a la organización) en información estructurada, para su explotación directa (reporting, análisis OLTP / OLAP, alertas...) o para su análisis y conversión en conocimiento, dando así soporte a la toma de decisiones sobre el negocio[10]. Por su parte, un Data Warehouse se constituye como una parte fundamental de BI que surge a partir de la necesidad de las organizaciones de contar con información homogénea y confiable, en forma global e integrada. Facilitan la construcción y mantenimiento de datos estructurados que permiten realizar análisis de los datos en un entorno diferente al operacional. Este análisis promueve la transformación de los datos en información y la información en conocimiento.[11]

Asimismo, la tendencia en el avance de la tecnología que ha abierto las puertas hacia un nuevo enfoque de entendimiento y toma de decisiones, la cual es utilizada para describir enormes cantidades de datos (estructurados, no estructurados y semi estructurados) que tomaría demasiado tiempo y sería muy costoso cargarlos a un base de datos relacional para su análisis. De tal manera que,

el concepto de Big Data aplica para toda aquella información que no puede ser procesada o analizada utilizando procesos o herramientas tradicionales. Esta tecnología se encuentra en pleno desarrollo, encontrando soluciones open source que es necesario estudiar e investigar en forma sistemática para obtener resultados comparativos que sean de utilidad. Podemos mencionar aquí herramientas que realizan analytics en memoria, por ejemplo el streaming processing que realizan empresas como LinkedIn, Groupon a través de aplicaciones como Storm[12] y Kafk[13]; o Drill[14] y Dramel[15] para la exploración de datos. D3[16] es otra aplicación muy poderosa para crear tableros interactivos en forma rápida y eficaz visualmente.

Analizar y aplicar técnicas y estrategias, evaluar distintas herramientas de software libre disponibles en el mercado, relacionadas con BI y Big Data es la propuesta principal de este trabajo.

Asimismo, el estudio de las mismas requieren conocimientos previos de bases de datos básico. Es interesante entonces trabajarlos también con estudiantes de grados avanzados de la carrera, que hacen uso de buscadores de internet, de libros en línea, de las redes sociales como herramienta motivador y que permita reflejar emociones, compartir y aplicar técnicas del aprendizaje en red, reflexionando asimismo sobre su propio aprendizaje. El conectivismo como teoría de aprendizaje[17][18] y la aplicación de técnicas de learnign analytics[19] se constituyen como últimas tendencias en educación[20], sobre las que es necesario reflexionar y hacer experiencias que reflejen la forma de aprender de nuestros estudiantes.

### ***Líneas de Investigación y Desarrollo***

La tarea de investigación que estamos llevando a cabo comprenden varias líneas de trabajo todas ellas en el área de Inteligencia de Negocios y en el aprendizaje de estas temáticas.

Por un lado, aplicar distintas metodologías para la constitución de un Data Warehouse que soporte las distintas estrategias y técnicas para obtener conocimiento a partir de los datos almacenados en distintos sistemas. Que contemple la información de todas las fuentes de información mencionadas previamente. La construcción de un DW involucra una serie de actividades relacionadas con distintas técnicas para la Extracción, Transformación y Carga soportadas por distintas herramientas como Kettle de Pentaho BI, Spago BI [21], entre otras. La naturaleza del negocio determina los requerimientos de los usuarios que en este caso están relacionados con rendimiento académico, comportamiento en entornos virtuales y redes sociales.

Respecto al primer punto, rendimiento académico, se está trabajando en los cubos de análisis de información provistos por el Ministerio de Educación de la Nación, así como también en el diseño e implementación de un cubo de desgramiento por cohorte que permitan abordar la problemática de deserción en la Facultad y se puedan definir distintas políticas o estrategias para trabajarla.

Durante el año 2011 y 2012 se trabajó con la Facultad de Ingeniería de la UNLP, intentando dar respuesta a los requerimientos referidos al análisis del rendimiento de determinadas materias, cohortes, docentes y años académicos. El trabajo realizado permitió a la Facultad realizar un análisis pormenorizado de sus alumnos a lo largo de la carrera y trabajar en conjunto con el MEN para lanzar el Plan Estratégico de Ingeniería 2012 - 2016 [22]

El trabajo conjunto con la Facultad de Ciencias Económicas ha permitido implementar programas como de caracterización de aspirantes y del seguimiento del graduado, para el diagnóstico y comprensión de las características de los alumnos de esta Facultad. [23]

Por otro lado, la aplicación de técnicas de minería de datos permitirá identificar los diferentes perfiles de los estudiantes y egresados, ayudando a comprender mejor su comportamiento en las plataformas virtuales y las redes sociales, para implementar



propuestas educativas que los asistan para completar sus estudios, con el objetivo de disminuir la deserción en las carreras de Informática [24]. El grado de aprovechamiento de la información minada por parte del usuario final depende en gran medida de una correcta visualización y una interfaz amigable de interacción por este motivo se prevé el desarrollo de aplicaciones ad hoc para facilitar las consultas de los usuarios finales. Distintas herramientas de minería brindan librerías que se integran con entornos de desarrollo para simplificar esta tarea.

La aplicación de técnicas y el análisis y uso de herramientas de Big Data se constituye como otra línea de investigación por sí misma. Durante el año 2012 en la materia Tecnologías Aplicadas para Business Intelligence se introdujo el tema y los alumnos investigaron sobre las distintas herramientas, profundizando sobre una de ellas. El trabajo realizado se puede consultar en <http://www.facebook.com/catedrabi.unlp>. Poder extender el análisis a datos de toda la universidad y facilitar la integración las bases de datos abiertas de organismos públicos permitirá aplicar las técnicas y herramientas analizadas sobre esta temática. Analizar el comportamiento de los estudiantes en las redes sociales constituirá otro caso de estudio.

El uso de redes sociales para la enseñanza de TI con alumnos avanzados permite realizar primeras experiencias en la aplicación de nuevos paradigmas educativos, como el conectivismo y el aprendizaje en red [25].

## **Resultados y Objetivos**

Entre los objetivos planteados podemos mencionar:

- Brindar herramientas que favorezcan la toma de decisiones gerenciales trabajando para la mejora continua que se vean reflejados en cambios permanentes para la institución.

- Transformar toda la información contenida en los sistemas operacionales, sobre los cuales se tiene acceso, de una manera rápida, consistente, íntegra y confiable, colaborando con ayudar a los directivos para propiciar mejoras en las ofertas educativas y los servicios ofrecidos por la institución.
- Analizar herramientas para aplicar técnicas de data analytics y Big data sobre las redes sociales, como Netvizz [26], Google Analytics y Gephi [27]
- Innovar en el dictado de materias de años avanzados de la carrera acorde a los nuevos paradigmas educativos y formas de comunicarse de la sociedad.

Entre los resultados obtenidos al momento podemos mencionar que se ha afianzado el dictado de la materia optativa “Tecnologías Aplicadas a BI” dado que la matrícula va en aumento y se incorporaron dos colaboradores en este último año. Asimismo la tasa de asistencia es alta lo cual es significativo en alumnos que son activos laboralmente en un 80%.

Se han realizado convenios de capacitación sobre la temática con organismos públicos provinciales, como la Dirección Provincial de Sistemas de Información y Tecnología de la provincia de Buenos Aires.

Se trabaja en conjunto con el Ministerio de Educación de la Nación en el desarrollo de nuevos cursos e implementación sobre sistemas de SL.

## **Formación de Recursos Humanos**

La capacitación permanente en temas relacionados con la temática ha permitido afianzar el grupo de trabajo de 2 profesores, 1 ayudante y 2 colaboradores, que en conjunto aprenden y colaboran en la experimentación y

construcción de casos de estudio. El trabajo conjunto con el personal del Centro Superior para el Procesamiento facilita y promueve la implementación de las técnicas y estrategias propuestas.

Se encuentran en desarrollo una tesina de Licenciatura en Informática denominada Aplicación de técnicas y estrategias de Inteligencia de Negocio (BI) para analizar/integrar información de los alumnos de la Facultad de Informática de la UNLP. Y se trabaja con distintos alumnos interesados en realizar su tesina sobre la temática.

## Referencias

- [1] Anuario Estadístico UNLP 2012: Indicadores Comparados. <http://www.unlp.edu.ar/indicadores> (Ultima visita Marzo 2013)
- [2] <http://www.siu.edu.ar/>
- [3] <http://www.ideasoft.biz/>
- [4] <http://www.pentaho.com/>
- [5] <http://www.moodle.org/>
- [6] Consorcio SIU <http://.siu.edu.ar> (Ultima visita Marzo 2013)
- [7] <http://www.cespi.unlp.edu.ar/>
- [8] Meran UNLP <http://www.meran.unlp.edu.ar/> (Ultima visita Marzo 2013)
- [9] <http://www.koha.org/>
- [10] <http://www.bi-argentina.com.ar/que-es-business-intelligence/>
- [11] Kimball 1992
- [12] storm-project.net
- [13] <http://kafka.apache.org/design.html>  
<http://www.ibm.com/developerworks/ssa/local/im/que-es-big-data/index.html>
- [14] <http://gigaom.com/cloud/for-fast-interactive-hadoop-queries-drill-may-be-the-answer/>
- [15] [http://static.googleusercontent.com/external\\_content/untrusted\\_dlcp/research.google.com/en/us/pubs/archive/36632.pdf](http://static.googleusercontent.com/external_content/untrusted_dlcp/research.google.com/en/us/pubs/archive/36632.pdf)
- [16] <http://d3js.org>
- [17] <http://www.elearnspace.org/Articles/connectivism.htm>
- [18] <http://aprenderapensar.net/2012/12/10/el-conectivismo-como-forma-de-aprendizaje/>
- [19] <http://www.solaresearch.org/>
- [20] <http://blogcued.blogspot.com.ar/2011/10/es-el-conectivismo-una-teoria-lo-es-del.html>
- [21] [www.spagobi.org](http://www.spagobi.org)
- [22] [http://www.ing.unlp.edu.ar/institucional/difusion/2012/ingreso\\_retencion](http://www.ing.unlp.edu.ar/institucional/difusion/2012/ingreso_retencion)

- [23] [http://www.econo.unlp.edu.ar/caracterizacion\\_aspirantes](http://www.econo.unlp.edu.ar/caracterizacion_aspirantes)
- [24] Tesina de Licenciatura en Informática de A.C. Kruzylko Ostojic, Claudia Yanina en desarrollo
- [25] [www.fernandosantamaria.com](http://www.fernandosantamaria.com)
- [26] <https://lab.digitalmethods.net/~brieder/facebook/netvizz/>  
[fb\\_source=bookmark\\_apps&ref=bookmarks&count=0&fb\\_bmpos=3\\_0&state=cddffb9cc8ddcabfa791ca4a2df113f5&code=AQB5eo6CLldv\\_Dj0BSAUwApJDFhwSQBw1kwIrnoZD0pNqxwoy0lu05Ys8uPDqw8GefvHtGADhDeaKHYAGC9bnb8HTEeabHW6wbWNqIsFzLMhUARwN3ZFwd6yOgOGZURPM3PN5Qjz-T9oFnuADbE9IHVwVRzDUvIMK620GY4qFX9pgdutky0qRjqibaMj7-rGX59w8VVk2wbaIY79xpuEHUG#\\_=\\_](https://www.facebook.com/fernandosantamaria/)
- [27] <http://www.whatsnew.com/2011/09/29/gephi-impresionante-plataforma-de-visualizacion-de-datos-de-codigo-abierto/>

# Transformaciones geométricas para facilitar la identificación de objetos en imágenes digitales

Jorge Kamlofsky, María Lorena Bergamini

CAETI – Facultad de Tecnología Informática. Universidad Abierta Interamericana

Av. Montes de Oca 745. Ciudad de Buenos Aires

(+54 11) 43015323; 43015240 ; 43015248

[Jorge.Kamlofsky@uai.edu.ar](mailto:Jorge.Kamlofsky@uai.edu.ar); [Maria.Bergamini@uai.edu.ar](mailto:Maria.Bergamini@uai.edu.ar)

## Resumen

La visión artificial por reconocimiento de patrones es un concepto que permite dar visión de mayor calidad, ver en tiempo real y seguir objetos. Dependiendo de su posición, orientación o tamaño, un objeto puede generar millones de imágenes diferentes, lo que dificulta su identificación.

Las características topológicas de un objeto pueden estudiarse por medio de un análisis de los bordes. Convexidad, concavidad, curvatura, puntos extremos pueden detectarse siguiendo los puntos frontera del objeto, y en su conjunto permiten identificar el mismo.

Estos patrones del borde tienen relación con su ubicación dentro del borde del objeto, y también con la inclinación y escala del mismo. Dotando a una imagen digital de una topología y aprovechando las propiedades invariantes, se estudian estrategias de reconocimiento de formas.

## Palabras claves

Topología digital, visión artificial, invariantes geométricos, reconocimiento de patrones.

## Contexto

El proyecto presentado comenzó en agosto de 2011. Se desarrolla en el Centro de Altos Estudios en Tecnología Informática (CAETI),

dependiente de la Facultad de Tecnología Informática de la Universidad Abierta Interamericana (UAI). Los proyectos de investigación que se desarrollan en el CAETI se organizan en 5 líneas temáticas identificadas como prioritarias. El proyecto aquí presentado se enmarca en la línea prioritaria de Algoritmos y Software. Esta línea se orienta al diseño de software atendiendo el creciente aumento en la complejidad de las tareas a resolver, manteniendo calidad, tiempos y costos controlados. Con mejores herramientas, se alcanza mayor complejidad del software que se construye y son más demandantes los requerimientos que debe cumplir. Es esperable que la tendencia en cuanto al ritmo de crecimiento de la complejidad del software y la dependencia que las empresas, el estado y la sociedad en general tienen sobre los servicios provistos por software continuará en las próximas décadas.

En definitiva, la investigación, transferencia, innovación y capacitación en el área de algoritmos y software es fundamental para poder acompañar las necesidades de producción, operación y mantenimiento de software en los próximos años.

## Introducción

El reconocimiento de patrones específicos para la identificación de un objeto dentro de

una imagen hace que el proceso de visión artificial sea mucho más eficiente, permitiendo incrementar la calidad de visión, trabajar en tiempo real o bien reducir los requisitos tecnológicos para crear un sistema de visión. El desarrollo de esta tecnología abre nuevas ventanas en los procesos de automatización: identificación y conteo de vehículos, seguimiento de objetos en movimiento, detección de fuego sin utilización de sensores de temperatura, diferenciación de formas humanas y animales en alarmas perimetrales, lectura de chapas patentes, dactiloscopia, diagnóstico médico por imágenes, etc.

Dado que las imágenes digitales, sobre todo en secuencias de tiempo real, involucran una alta cantidad de información digital, no es razonable aplicar métodos basados en análisis pixel a pixel, que requerirían tiempos computacionales incompatibles con el tiempo real. Es fundamental el diseño de algoritmos eficientes que puedan realizar tareas de reconocimiento de objetos manejando sólo la información necesaria de la imagen. Tratar a ciertos píxeles que conforman un objeto dentro de una imagen como un conjunto, y detectar sus propiedades topológicas es un enfoque simple que ha demostrado ser altamente efectivo cuando se pretende interpretar automáticamente el contenido de una imagen.

Las aplicaciones de visión artificial se sostienen en el concepto matemático de topología digital (Eckhardt, Latecki, 2003). La topología digital ha demostrado ser un marco teórico adecuado para el análisis de objetos dentro de imágenes digitales, desde su concepción a inicios de la década de 1970. En este contexto se aplica el *Teorema de curvas de Jordan* en el plano digital, que afirma que toda curva simple cerrada contenida en un conjunto conexo separa al conjunto en dos subconjuntos: su interior o agujero y su complemento o fondo. Este teorema requiere las nociones topológicas de vecindad, adyacencia, camino, conectividad, arco y curva. El teorema de curvas de Jordan permite separar un objeto del resto de la imagen digital en la que se encuentra inmerso. Los trabajos en el área de topología digital y geometría digital de Eckhardt y Latecki (1994) y Latecki

et al. (2000) han marcado una clara guía en esta línea.

En la práctica, se aplica un algoritmo para demarcar la curva que representa el borde o frontera de un objeto dentro de una imagen digital. El algoritmo permite localizar los puntos de un conjunto conexo que limitan con su complemento y los del complemento que limitan con el conjunto, conformando así la frontera del mismo (Kamlofsky, 2011).

Gross y Latecki (1995) proponen un conjunto de patrones de frontera que no se deforman en el proceso de digitalización, y demuestran que, bajo ciertas condiciones, son los únicos patrones locales que pueden ocurrir. Esto permite estudiar propiedades geométricas del objeto digitalizado (convexidad, curvatura) y detectar ruido.

Un concepto clave en el estudio geométrico de imágenes digitales es el de curvatura. Se han sugerido varias generalizaciones de la definición de curvatura proveniente de geometría diferencial a la geometría digital. En particular, en Liu et al. (2008) se plantea medir la curvatura en imágenes digitales basándose en una función altura medida con una rotación de ejes. Este nuevo concepto coincide con la definición estándar de curvatura en curvas suaves cuando el número de alturas es suficientemente alto, y coincide con el concepto de ángulo de giro en una curva poligonal.

La tarea siguiente consiste en definir qué tipo de objeto tiene a esa curva por borde, analizando las propiedades geométricas del mismo. Éstas se pueden comparar con ciertos patrones previamente recopilados, identificando de este modo al objeto en cuestión.

Pero estos patrones están relacionados con la posición, ángulo de inclinación y tamaño del objeto. Pueden tenerse patrones almacenados en memoria, en diferentes posiciones, tamaños o inclinaciones, o bien, más lógica y eficientemente, puede guardárselos en una única posición, tamaño e inclinación.

La tecnología de visión por reconocimiento de patrones utiliza el concepto de *voting* para decidir acerca de a qué objeto corresponde la imagen analizada (van Erp et al., 2002).



Mediante voting se compara la imagen con patrones y se decide con cuál se identifica, midiendo la similitud entre el objeto y los patrones candidatos.

Para esta comparación, se aplican transformaciones geométricas en los objetos.

Wu et al. (2010) proponen un modelo de base activa para representar clases de objetos, junto con un algoritmo de aprendizaje y reconocimiento de patrones. Sugieren representar el borde de una clase de objetos por una base de elementos en determinada posición y orientación. En la etapa de aprendizaje, los elementos de la base son seleccionados y aplicados a plantillas de aprendizaje, perturbando levemente su posición y orientación, ajustando estos parámetros a la imagen. Finalmente, se propone un esquema de detección y reconocimiento.

La identificación de formas en general se basa en una medida de similaridad con los patrones existentes en una base de datos. Muchos investigadores han propuesto medidas de similaridad, o distancias entre formas, que tienen en cuenta la distancia interna (Ling y Jacobs, 2007), partes visualmente significativas (Latecki y Lakämper, 2000), entre otros.

Enfoques actuales proponen técnicas de aprendizaje de similaridad, para mejorar la performance de reconocimiento. Éstos se basan en propagación en vecindades lineales – LNP- (Wang et al. 2006), aprendizaje multi-visual (Brefeld y Buscher, 2005), modelos generativos (Lawrence, Jordan, 2004), entre otros.

En otra línea de investigación, se estudian transformaciones de imágenes digitales que preservan las propiedades topológicas (conexidad, agujeros, etc). Estas transformaciones abarcan afinamiento mediante la eliminación de puntos simples, o conjuntos simples, segmentación basado en atlas, etc. (Faisan et al., 2008; Passat et al., 2009)

## **Líneas de Investigación y Desarrollo**

Dos estrategias para identificar y seguir objetos en movimiento son usadas: análisis de los bordes del objeto y análisis de sus esqueletos. En ambos casos, sólo se analiza una pequeña cantidad de información (píxeles), pero ese conjunto de puntos obtenidos mantiene propiedades topológicas respecto del conjunto original.

Los puntos del objeto que son adyacentes al complemento constituyen el borde del objeto. Dicho borde puede verse como una curva digital simple cerrada. Este subconjunto es de gran interés en el análisis de objetos dentro de imágenes, ya que muchas características (convexidad, dimensiones, agujeros, etc.) pueden estudiarse por medio de ese subconjunto.

Luego de recorrer la frontera de un objeto en una imagen digital, la tarea siguiente consiste en identificar qué tipo de objeto tiene a esa curva por borde, analizando las propiedades geométricas del mismo. Esta identificación puede llevarse a cabo por comparación con ciertos patrones previamente recopilados y estandarizados. Dado que cada patrón se guarda en una determinada posición y escala, se requiere modificar la imagen analizada (o su frontera) mediante transformaciones geométricas, de modo que mantenga las propiedades de su geometría intrínseca (conexión, agujeros, convexidad, concavidad, puntos de curvatura máxima y mínima, puntos de inflexión, etc).

Se pretende profundizar el estudio del comportamiento de curvas digitales mediante transformaciones geométricas, el efecto de éstas sobre las características topológicas y geométricas, a fin de determinar cuáles son factibles de ser aplicadas. Los resultados teóricos obtenidos fundamentarán un algoritmo de reconocimiento de imágenes, que será diseñado e implementado para tal fin.

El proceso de digitalización, y el sistema de adquisición de la imagen generan ruido, que tiene una presencia notoria en las fronteras de los objetos donde aparecen formas de dientes de serrucho. Esto dificulta su estudio, ya que el análisis de bordes por geometría diferencial requiere el análisis de longitud, rectitud, curvatura, convexidad, características que se

ven afectadas por las irregularidades del borde, producto del proceso de digitalización y adquisición de imagen. Es necesario eludir la influencia del mencionado ruido, a fin de, por un lado, reconocer eficientemente el objeto, y por otro, optimizar el uso de memoria, y no ocuparla con datos superfluos.

## Resultados y objetivos

El proyecto pretende avanzar en el desarrollo teórico de topología digital, específicamente estudiando procesos de discretización y su efecto sobre las características geométricas del objeto, y cómo estas características se mantienen o no en un proceso de transformación geométrica (rotación, escala, perspectiva). Y desde el punto de vista práctico, enfocando hacia las aplicaciones concretas, se pretende diseñar e implementar un algoritmo que, dado un objeto en una imagen digital, lo transforme en un objeto equivalente (aprovechando propiedades invariantes de la topología) a fin de facilitar la identificación del objeto mediante comparación con patrones.

A fin de suavizar y simplificar la curva de borde, se diseñó un algoritmo para aproximar una curva digital por un polígono, con tolerancia de error variable.

Así, el borde se reduce ahora a un conjunto de puntos que son los vértices del polígono aproximante, lo que brinda una enorme ganancia en capacidad de procesamiento.

Con tolerancias muy pequeñas, se generan polígonos de muchos vértices y una mejor aproximación. En el caso límite de tolerancia cero, los vértices del polígono son casi todos los puntos de la curva. Con tolerancia más grandes se obtienen polígonos de menos lados pero una aproximación menos ajustada.

Como preprocesamiento de este algoritmo, se aplica un método para disminuir la cantidad de puntos de la curva, sin perder información. Este preprocesamiento consiste en detectar sucesiones de puntos de borde que están sobre una recta (conjunto de ancho cero) y eliminar los todos esos puntos, excepto los extremos. Luego, con la tolerancia deseada, se elimina puntos que están casi alineados.

Teniendo un polígono que aproxima una curva digital, se puede aproximar la curvatura en cada parte de la curva mediante los ángulos de giro en los vértices del polígono. La sucesión de ángulos de giro muestra un patrón específico de la forma analizada, y es invariante a la escala y rotación. Además, ciertas características que se pueden detectar con dicho patrón son invariantes frente a movimientos de perspectivas (puntos de inflexión, puntos de curvatura máxima, bordes rectos, etc.). El patrón normalizado luego se compara con patrones de objetos conocidos, se mide la similaridad, y se reconoce cuál es el objeto en la imagen.

## Formación de recursos humanos

Este proyecto bianual se desarrolla bajo la dirección de la Dra. María Lorena Bergamini. Actualmente el equipo se completa con el investigador Licenciado Jorge Kamlofsky y Luciana Klacherian (estudiante de Licenciatura en Matemática - UAI).

La tesis de grado de Kamlofsky con la que accedió al título de Licenciado en Matemática, y dirigida por la Dra. Samira Abdel Masih y el Ing. Néstor Balich, sirvió de investigación alentadora del proyecto aquí presentado. Luciana Klacherian, con la dirección de la Dra. Bergamini, se encuentra desarrollando su tesis de Licenciatura, en la que estudia caracterizaciones de rectas digitales y segmentos de rectas digitales. Se estima que en mayo de este año finalizará dicha tesis.

## Referencias

1. U. Brefeld, C. Buscher, T. Scheffer. Multiview Discriminative Sequential Learning, Proc. European Conf. Machine Learning, 2005.
2. U. Eckhardt, L. Latecki. Topologies for the digital spaces  $Z^2$  and  $Z^3$ . Computer Vision Image Understanding, 90, pp. 295-312, 2003.
3. U. Eckhardt, L. Latecki. Digital Topology. In Current Topics in Pattern Recognition Research, Research Trends,

- Council of Scientific Information, Vilayil Gardens, Trivandrum, India, 1994.
4. S. Faisan, N. Passat, V. Noblet, R. Chabrier, C. Meyer. Topology Preserving Warping of Binary Images: Application to Atlas-Based Skull Segmentation. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2008 Lecture Notes in Computer Science*, 5241, pp. 211-218, 2008.
  5. A. Gross, L. Latecki. Digital geometric invariance and shape representation, IEEE, 1995.
  6. J. Kamlofsky. Topología digital: base para la visión artificial. Tesis de grado. Facultad de Tecnología Informática. UAI. Buenos Aires, 2011.
  7. J. Kamlofsky, M. L. Bergamini. Aproximación de formas de objetos digitales por polígonos, VII Jornadas Argentinas de Robótica, 2012.
  8. L. J. Latecki, R.-R. Ghadially, R. Lakaemper, U. Eckhardt. Continuity of the discrete curve evolution. *Journal of Electronic Imaging* 9 (3), pp. 317-326, 2000.
  9. L. J. Latecki, R. Lakämper. Shape Similarity Measure Based on Correspondence of Visual Parts, *IEEE Trans. Pattern Analysis and Machine Intelligence* 22(10), pp. 1185-1190, 2000.
  10. N. D. Lawrence, M.I. Jordan. Semi-Supervised Learning via Gaussian Processes, *Advances in Neural Information Processing Systems*, MIT Press, 2004
  11. H. Ling, D. Jacobs. Shape Classification Using the Inner-Distance, *IEEE Trans. Pattern Analysis and Machine Intelligence* 29(2), pp. 286-299, 2007.
  12. H. R. Liu , L. J. Latecki, W. Liu. A Unified Curvature Definition for Regular, Polygonal, and Digital Planar Curves. *Int J Comput Vis* 80, pp 104–124, 2008.
  13. N. Passat, M. Couprie, L. Mazo, G. Bertrand. Topology-Preserving Thinning in 2-D Pseudomanifolds. *Discrete Geometry for Computer Imagery Lecture Notes in Computer Science*, 5810, pp. 217-228, 2009.
  14. M. van Erp, L. Vuurpijl, L. Schomaker. An overview and comparison of voting methods for pattern recognition. *Proceedings of the Eighth International Workshop on Frontiers in Handwriting Recognition (IWFHR'02)*, pp 195-200, 2002.
  15. F. Wang, J. Wang, C. Zhang, H. Shen. Semi-Supervised Classification Using Linear Neighborhood Propagation, *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition*, 2006.
  16. Y. N. Wu, Z. Si, H. Gong, S. C. Zhu. Learning active basis model for object detection and recognition, *Int J Comput Vis* 90, pp 198–235, 2010.

## Ontologías y Semántica en el Proceso de Visualización

Martín L. Larrea, Sebastián Escarza, Dana K. Urribarri y Silvia M. Castro  
 Laboratorio de Investigación y Desarrollo en Visualización y Computación Gráfica (VyGLab)  
 Departamento de Ciencias e Ingeniería de la Computación  
 Universidad Nacional del Sur  
 Bahía Blanca, Argentina  
 {mll, se, dku, smc}@cs.uns.edu.ar

### Resumen

La Visualización como área se encuentra en un estado en el cual existe un consenso parcial acerca de sus principales conceptos. Sin embargo, aún no se ha alcanzado un estado de madurez suficiente en el área y el establecimiento de un vocabulario formal común sigue siendo un objetivo pendiente.

Esta línea de investigación tiene como objetivo crear un modelo de visualización que considere la semántica de los datos, del contexto y de las etapas del “Modelo Unificado de Visualización” (MUV) para poder asistir al usuario en el seteo de los parámetros de la visualización logrando así una mejor representación visual. Este nuevo modelo, que se presenta como una extensión del MUV, deberá formalizar el modelo de referencia mediante una Ontología de Visualización, incorporar una Ontología de Datos e incorporar una capa de inferencia que le permita derivar nueva información, a partir de la ya conocida.

**Palabras clave:** *Ontología, Representación Formal, Visualización basada en semántica, Visualización.*

### Contexto

Este trabajo continúa la línea presentada en [LEUC12] y se lleva a cabo en el Laboratorio de Investigación y Desarrollo en Visualización y Computación Gráfica (VyGLab) del Departamento de Ciencias e Ingeniería de la Computación de la Universidad Nacional del Sur.

La línea de Investigación presentada está inserta en el proyecto y “Representaciones Visuales e Interacciones para el Análisis Visual de Grandes Conjuntos de Datos”

(24/N020), dirigido por la Doctora Silvia Castro y , financiado por la Secretaría General de Ciencia y Tecnología de la Universidad Nacional del Sur; y acreditado por la Universidad Nacional del Sur, Bahía Blanca.

### Introducción

Actualmente existe un gran número de modelos de referencia en Visualización [CM97, Chi00, Shn04, BN07] que, de diversas maneras, han identificado y delineado los principales componentes y procesos que sufren los datos para ser visualizados. En particular, en el seno del VyGLab hemos estado desarrollando el Modelo Unificado de Visualización (MUV) [MCFE03], un modelo que constituye un marco conceptual de referencia en términos del cual ubicar los procesos y estados intermedios de los datos y definir las interacciones explícitamente.

Si bien estos esfuerzos por compilar, caracterizar y clasificar los aspectos más relevantes del área han sido y siguen siendo fructíferos, aún no existe un consenso definitivo que permita consolidar una teoría de base en Visualización.

La incorporación de una semántica bien definida en el proceso de Visualización ya ha sido identificada como una necesidad [DBD\*05, VO04, CEH\*09] en el sentido que permitiría formalizar dicho proceso, estableciendo un vocabulario común que permita a los usuarios plantear sus requerimientos, y a los diseñadores de herramientas de visualización, expresar las transformaciones de los datos desde su obtención hasta la construcción de la vista, las operaciones que deben proveerse y las



formas de interacción posibles entre los usuarios y la visualización

Para lograr una formalización consensuada del proceso de Visualización, es preciso describir axiomáticamente el marco de referencia provisto por los modelos mencionados anteriormente. En este contexto, las ontologías surgen como una herramienta natural para estos fines.

Un ejemplo de los beneficios de la formalización del proceso de visualización se encuentra en la formalización de los tipos de datos, es decir una Ontología de Datos. Contar con una ontología que clasifique los conjuntos de datos a visualizar es una guía que asiste a la hora de elegir la técnica de visualización apropiada para determinado conjunto de datos. En la literatura es posible encontrar taxonomías, una versión menos formal de la ontología, de datos pero no son presentadas desde un punto de vista de visualización.

Por otro lado, el desafío de una visualización es encontrar una metáfora visual que permita entender y percibir en forma efectiva un conjunto de datos. Para este propósito es útil contar con cuantificaciones de diferentes aspectos de cada técnica de visualización. Una visualización debe proveer también un conjunto de interacciones a partir de las cuales el usuario explorará el conjunto de datos con una mínima carga cognitiva. La tecnología computacional actual permite la exploración de grandes conjuntos de información. Por un lado, esta situación es extremadamente útil pero la creciente cantidad de información genera una sobrecarga cognitiva. Mientras que el poder computacional ha crecido en forma exponencial, la habilidad para interactuar con dichos datos solo se ha incrementado en forma lineal.

Hoy en día, una gran variedad de usuarios acceden, extraen y muestran información que está distribuida sobre diferentes puntos, con diferentes tipos, formas y contenidos. En muchos casos, el usuario debe tener un control activo sobre el proceso de visualización pero, aún en

este caso, es difícil obtener una visualización efectiva. Es común que la información que se desea representar no tenga una manifestación visual obvia, ante esta situación el proceso de mapeo del conjunto de datos a la vista puede llegar a ser no trivial [CMT01].

## Líneas de Investigación y Desarrollo

Este trabajo presenta tres ejes de investigación las cuales se entrelazan y tienen varios puntos en común. Tales ejes son:

1. “Formalización del proceso de Visualización mediante el uso de Ontologías”.
2. “Caracterización de Conjuntos de Datos”.
3. “Caracterización de Técnicas de Visualización”.
4. “Visualización Basada en Semántica”.

## Formalización del Proceso de Visualización mediante el uso de Ontologías

Este eje es una continuación directa de la línea de investigación presentada en ediciones anteriores de este Workshop [ECM10], en la cual se plantea la necesidad de definir una Ontología de Visualización y de brindar soporte a dicha especificación formal a través de una plataforma de software que facilite la integración de los diversos componentes.

Este trabajo integra el uso de ontologías en Visualización. Las ontologías, si bien se originaron hace tiempo en el área de Representación del Conocimiento, han cobrado nueva relevancia con el surgimiento de proyectos como la Web Semántica. A raíz de ello, numerosos estándares han sido definidos. Especificaciones como RDF, RDF Schema y OWL (el Lenguaje de Ontologías de la Web) permiten definir vocabularios estructurados

que modelan jerarquías de herencia entre conceptos, relaciones, restricciones y reglas, que permiten un tratamiento más riguroso del conocimiento al poseer una semántica formal asociada.

En cuanto a la integración de semántica en visualización, aún no se cuenta con abordajes sistemáticos al problema y se identifican dos enfoques principales. Por un lado, se utilizan representaciones formales para mejorar la integración, consulta y descripción de los datos del usuario como una forma de enriquecer la visualización, y por el otro, trabajos más relacionados con el que proponemos utilizan descripciones semánticas, aunque en principio algo limitadas, del proceso de visualización en sí mismo para ayudar al usuario en la definición y configuración del mismo y en la selección de una técnica de visualización apropiada.

### **Caracterización de Conjuntos de Datos**

Este eje se centra en el desarrollo de una clasificación de los conjuntos de datos orientada a la visualización. Esta clasificación debe brindar suficiente información sobre cuáles son las características con las que debe contar la técnica de visualización que se emplee para visualizar cada categoría de datos.

El desafío es encontrar métricas que, no sólo permitan evaluar en forma lo más sencilla posible cada uno de los aspectos importantes a tener cuenta, sino que también permitan una clasificación conveniente de los datos.

### **Caracterización de Técnicas de Visualización**

Si bien a la hora de visualizar conjuntos de datos pequeños la mayoría de las técnicas resultan adecuadas, la situación cambia cuando se quieren visualizar grandes conjuntos de datos: una mala elección de la técnica a utilizar o de sus atributos y el resultado obtenido puede no

ser satisfactorio. Una posible solución a este problema es contar con alguna herramienta que permita predecir qué técnica es más "conveniente" usar para obtener información de un determinado conjunto de datos.

En este contexto se buscan desarrollar métricas particulares para cada técnica de visualización con el objetivo de caracterizar numéricamente algún aspecto del resultado final de la aplicación de dicha técnica sobre un determinado conjunto de datos.

### **Visualización Basada en Semántica**

Debido a que el objetivo de una visualización es lograr una representación que ayude al usuario a interpretar un conjunto de datos y comunicar su significado, es importante controlar el mapeo de las dimensiones físicas a las preceptuales; un usuario inexperto podría utilizar un mapeo incorrecto afectando negativamente la visualización resultante. Una estrategia para mejorar esta situación es guiar al usuario en la selección de los diferentes parámetros involucrados en la visualización a través de información semántica.

### **Resultados y Objetivos**

A nivel de la primera línea de investigación presentada, la transformación de los objetivos en resultados sigue las etapas delineadas en [ECM10].

Al momento de esta publicación se cuenta con una definición parcial de la Ontología de Visualización. Concretamente, la arquitectura general de la ontología ya ha sido definida así como los principales elementos de la misma.

Adicionalmente, también se ha definido un conjunto de conceptos destinados a caracterizar el espacio de datos del usuario y la representación visual, y se continúa trabajando en la definición de los conceptos ontológicos necesarios para describir el proceso de visualización como una red de etapas interconectadas, así como en las

definiciones necesarias para especificar las interacciones entre el usuario y la visualización.

Sobre la segunda línea presentada, hasta el momento se han evaluado cuales son las posibles métricas que se empelarán para categorizar los datos. En este sentido, se seleccionaron varias métricas globales sobre grafos, métricas sobre tablas de información y diferentes medidas de dispersión.

La tercera línea se desprende del trabajo realizado en la línea anterior y persigue el desarrollo de métricas específicas para cada técnica de visualización. Hasta el momento se han analizado métricas particulares para Scatterplots.

Finalmente, sobre la tercera línea de investigación, se ha logrado incluir la semántica de los datos en el mapeo de datos a colores en una representación visual [LMC10]. Gracias a este trabajo, esta etapa del proceso de visualización ya no requiere de la participación del usuario; las decisiones de qué color usar para cada dato se determinan a partir de un razonador semántico.

## Formación de Recursos

### Humanos

En lo concerniente a la formación de recursos humanos se detallan las tesis concluidas y en desarrollo relacionadas con la línea de investigación presentada:

#### Tesis Concluidas

- Tesis Doctoral. Martín L. Larrea. Tema: *Visualización Basada en Semántica*. Dirección: Dra. Silvia Castro.

#### Tesis en Desarrollo

- Tesis Doctoral. Sebastián Escarza. Tema: *Ontologías de Visualización*. Dirección: Dra. Silvia Castro.

- Tesis Doctoral. Dana Urribarri. Tema: *Escalabilidad Visual*. Dirección: Dra. Silvia Castro.
- Tesis Doctoral. María Luján Ganuza. Tema: *Servicios Web en Visualización de Información*. Dirección: Dra. Silvia Castro.

## Referencias

[BN07] K. Brodlie and N. M. Noor. Visualization Notations, Models and Taxonomies. pages 207–212, Bangor, United Kingdom, 2007. Eurographics Association.

[CEH\*09] M. Chen, D. Ebert, H. Hagen, R. S. Laramée, R. van Liere, K.-L. Ma, W. Ribarsky, G. Scheuermann, and D. Silver. Data, information, and knowledge in visualization. *IEEE Comput. Graph. Appl.*, 29(1):12–19, 2009.

[CM97] S. K. Card and J. Mackinlay. The structure of the information visualization design space. In *INFOVIS '97: Proceedings of the 1997 IEEE Symposium on Information Visualization (InfoVis '97)*, page 92, Washington, DC, USA, 1997. IEEE Computer Society.

[CMT01] Carpendale, M. S. T. 2001. Considering Visual Variables as a Basis for Information Visualization. Technical Report. University of Calgary, Department of Computer Science.

[Chi00] E. H. Chi. A taxonomy of visualization techniques using the data state reference model. In *Proceedings of the IEEE Symposium on Information Visualization (InfoVis'00)*, page 6975. IEEE Computer Society Press, 2000.

[DBD\*05] D. J. Duke, K. W. Brodlie, D. A. Duce, and I. Herman. Do you see what I mean? *IEEE. Computer Graphics and Applications*, 25(3):6–9, 2005

[ECM07] S. Escarza, S. Castro, S. Martig. Ontologías de Visualización. IX Workshop de Investigadores en Ciencias de la Computación (WICC2007) pp. 275-278. Universidad Nacional de la Patagonia San Juan Bosco. 3 y 4 de Mayo de 2007. Trelew. Chubut. Argentina. ISBN 978-950-763-075-0.

[ECM08] S. Escarza, S. Castro, S. Martig. Desafíos en el camino desde el Modelo Unificado de Visualización hasta la construcción de visualizaciones. X Workshop de Investigadores en Ciencias de la Computación (WICC2008) pp. 297-300. Universidad Nacional de la Pampa. 5 y 6 de Mayo de 2008. Gral. Pico. La Pampa. Argentina. ISBN 978-950-863-101-5.

[ECM10] S. Escarza, S. Castro, S. Martig. Formalización del Proceso de Visualización mediante el uso de Ontologías. XII Workshop de Investigadores en Ciencias de la Computación (WICC 2010) pp. 261-265. Universidad Nacional de la Patagonia San Juan Bosco. 5 y 6 de mayo de 2010. El Calafate. Santa Cruz. Argentina. ISBN: 978-950-34-0652-6.

[LEUC12] Martín L. Larrea, Sebastián Escarza, Dana K. Urribarri, Sergio R. Martig y Silvia M. Castro. Ontologías y Semántica en el Proceso de Visualización. XIV Workshop de Investigadores en Ciencias de la Computación (WICC 2012), pp. 302-306. Universidad Nacional de Misiones, Posadas, Misiones (abril 2012). ISBN 978-950-766-082-5

[LMC10] M. Larrea, S. Martig and S. Castro. Semantics-based Color Assignment in Visualization. Journal of Computer Science & Technology. Vol. 10 - No. 1 - April 2010 - ISSN 1666-6038.

[LMC10\*] M. Larrea, S. Martig and S. Castro. Formalización del Proceso de Visualización Basada en Semántica. XII Workshop de Investigadores en Ciencias de

la Computación (WICC 2010) pp. 270-274. Universidad Nacional de la Patagonia San Juan Bosco. 5 y 6 de mayo de 2010. El Calafate. Santa Cruz. Argentina. ISBN: 978-950-34-0652-6.

[MCFE03] Martig S., Castro S., Fillottrani P., Estévez E., Un Modelo Unificado de Visualización, Proceedings, pp. 881-892, 9º Congreso Argentino de Ciencias de la Computación. 6 al 10 de Octubre de 2003. La Plata. Argentina.

[Shn04] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. IEEE Symposium on Visual Languages, 0:336, 1996.

[UCM10] D. K. Urribarri, S. Castro, S. Martig. Caracterización de Conjuntos de Datos. XII Workshop de Investigadores en Ciencias de la Computación (WICC 2010) pp. 266-269. Universidad Nacional de la Patagonia San Juan Bosco. 5 y 6 de mayo de 2010. El Calafate. Santa Cruz. Argentina. ISBN: 978-950-34-0652-6.

[VO04] Reporte del Visualisation Ontology Workshop realizado en Nacional e-Science Centre el 7 y 8 de Abril de 2004. <http://www.nesc.ac.uk/esi/events/393/>



## TICs y Difusión del Patrimonio Cultural

Martín L. Larrea, Damián Flores Choque  
Laboratorio de Investigación y Desarrollo en Visualización y Computación Gráfica (VyGLab)  
Departamento de Ciencias e Ingeniería de la Computación  
{mll, dfc}@cs.uns.edu.ar

Andrés Pinassi, Yanel Martín Varisto, Astrid Bjerg, Patricia Ercolani  
Departamento de Geografía y Turismo  
[andres.pinassi@uns.edu.ar](mailto:andres.pinassi@uns.edu.ar), [yanel.martin@uns.edu.ar](mailto:yanel.martin@uns.edu.ar), [laberjg@hotmail.com](mailto:laberjg@hotmail.com), [ercolani@uns.edu.ar](mailto:ercolani@uns.edu.ar)

Universidad Nacional del Sur  
Bahía Blanca, Argentina

### Resumen

La ciudad de Bahía Blanca cuenta con diferentes áreas de valor patrimonial, dentro de las cuales se destaca su casco fundacional. Dada su importancia en el proceso de desarrollo y consolidación del centro urbano, resulta relevante establecer estrategias que contribuyan a la valorización de sus componentes patrimoniales.

En este contexto, las nuevas tecnologías aplicadas al turismo, permiten un mejor acceso y mayor acercamiento a estos bienes culturales, convirtiéndose en herramientas elementales para su difusión. La Realidad Aumentada y Virtual se visualizan como dos nuevas tecnologías importantes para el tipo de difusión planteado, ya que posibilitan un acercamiento a los bienes patrimoniales, de forma atractiva y diferente a los medios convencionales de impulsión ya conocidos. Se ha planificado, a partir de esto, un trabajo articulado entre el Departamento de Geografía y Turismo y el Departamento de Ciencias e Ingeniería de la Computación, ambos pertenecientes a la Universidad Nacional del Sur, y la Dirección General de Asuntos Turísticos y Relaciones Internacionales, de la Municipalidad de Bahía Blanca. Este trabajo tiene como objetivo contribuir a la difusión del patrimonio arquitectónico y urbano del área fundacional de Bahía Blanca, mediante la implementación de las nuevas tecnologías aplicadas al turismo, como la Realidad Aumentada y Virtual.

**Palabras clave:** TICs, Patrimonio Cultural, Realidad Virtual, Realidad Aumentada, Computación Gráfica.

### Contexto

El trabajo se lleva a cabo en conjunto, entre el Laboratorio de Investigación y Desarrollo en Visualización y Computación Gráfica (VyGLab) del Departamento de Ciencias e Ingeniería de la Computación de la Universidad Nacional del Sur y el Departamento de Geografía y Turismo de la misma universidad.

### Introducción

Las *Tecnologías de la Información y la Comunicación* (TICs) constituyen herramientas que admiten manipular la información, permitiendo su tratamiento y transmisión, principalmente a través de la informática, Internet y telecomunicaciones. Los efectos de las nuevas tecnologías tienen repercusiones en distintos ámbitos de la vida humana, como ser la economía, el empleo, la forma de producción, la educación, entre otros.

La verdadera transformación tecnológica implica disponer las nuevas tecnologías como recurso, pero más aun, en la posibilidad de gestionarlas y aplicarlas, lo que da como resultado diferentes realidades territoriales.

En efecto, la revolución tecnológica consiste en “una revolución supermoderna

que no encuentra su poder en la permanencia sino en la potencialidad de ser descubierta” [1]. La relación entre las TICs y la ciudad se ha reflexionado desde el siglo XX, principalmente cuando comenzó a difundirse ampliamente el uso de Internet. Desde la década de los 80's, su uso se encuentra en constante expansión por la masificación de usuarios y por la diversidad de su aplicación. Particularmente en Argentina, a partir de los 90's este proceso se acentúa por la difusión masiva de Internet.

En la actualidad, el patrimonio se enfrenta a un contexto cambiante y complejo, basado en preceptos a escala global, que exigen una adecuación a las exigencias planteadas. En este marco, entra en juego la gestión en el patrimonio cultural, tratando de repensar iniciativas de puesta en valor, de acuerdo a las demandas del macroentorno.

Este proceso de planificación y asignación de recursos, destinado a la consecución de objetivos concretos, adquiere características disímiles a las de otros tiempos, planteándose un gran desafío: preservar el patrimonio a través de su uso activo. *“Superando las concepciones tradicionales que limitaban el cuidado o tutela del patrimonio al estudio y la conservación, nuestra época ha redescubierto las posibilidades de una gestión integral del patrimonio que plantea, además del reto de la conservación, encontrar los mejores usos para nuestro patrimonio histórico común, sin menoscabo de su preservación ni su valoración social.”* [2]

A los fines de la investigación, interesa resaltar una determinada forma de gestión, que posibilita el conocimiento del patrimonio a todos los actores intervinientes en la iniciativa de planificación y, a su vez, contribuye a informar e interiorizar al público objetivo al que se orienta dicha estrategia. Se hace referencia aquí, a la difusión del patrimonio propiamente dicha.

Las TICs y su injerencia en el campo del turismo y el patrimonio, se transforman

en los medios necesarios para poder difundir el valor histórico, cultural y social que albergan estas obras y manifestaciones contenedoras. De acuerdo al tipo de iniciativa, al componente patrimonial con el que se esté trabajando, al público objetivo que se quiere persuadir y al tipo de herramienta utilizada para la comunicación, dependerán los fundamentos concretos sobre los que se apoye la propuesta, priorizando en algunos casos ciertos principios en detrimento de otros, pero siempre teniendo presente la idea de hacer conocer y apropiarse estos bienes, como elementos portadores de la identidad de un pueblo.

## Realidad Virtual y Aumentada

La *Realidad Virtual* (RV) es una aplicación de la tecnología computacional cuyo objetivo es el de generar representaciones visuales que simulan mundos reales o ficticios. Es clave en RV lograr que el usuario tenga la sensación de estar presente en el mundo virtual. Las aplicaciones de RV en las actividades de la vida cotidiana son muchas y diversas. Podemos destacar la reconstrucción de la herencia cultural, la medicina y la simulación de multitudes, entre muchas otras.

La RV puede ser de dos tipos: inmersiva y no inmersiva. Los métodos inmersivos de RV están asociados a ambientes tridimensionales creados por computadora, los cuales se manipulan a través de dispositivos especiales diseñados para capturar los movimientos del cuerpo humano. En este caso, se busca que el usuario sólo vea el mundo virtual, aislándolo totalmente del mundo real. En el caso no inmersivo, también utiliza ambientes tridimensionales creados por computadora pero no se aísla al usuario del mundo real.

El término *Realidad Aumentada* (RA) se utiliza para describir la combinación, en tiempo real, de imágenes creadas por computadora con imágenes del mundo real,

creando así una realidad mixta. La integración de imágenes virtuales e imágenes reales es la principal diferencia entre RV y RA, puesto que esta última no sustituye la realidad física sino que la complementa o aumenta, agregando más información.

Un sistema de RA posee tres características básicas que lo identifican como tal: la combinación de elementos reales y virtuales en un entorno real; la capacidad de ser interactivos en tiempo real, y que estén correctamente registrados en 3D. Este último ítem se refiere a que los elementos reales y virtuales deben estar correctamente alineados, de manera que parezca que ambos se encuentren en un único entorno.

### **Área de Estudio: Casco Fundacional de Bahía Blanca**

El trabajo de investigación aborda el estudio del área fundacional de la ciudad de Bahía Blanca, la cual está comprendida dentro de su centro histórico. Este último se encuentra enmarcado por el límite establecido entre las calles: Mitre y Soler hacia el Noreste; Belgrano y Donado hacia el Sudeste; Brown y Vieytes hacia el Sudoeste; Rondeau y Rodríguez hacia el Noroeste. Se destacan así, quince cuadrículas de diferentes dimensiones, incluyendo el sitio sobre el que se encuentra la Plaza Rivadavia.

El centro histórico se caracteriza por albergar a la zona del fuerte que dio origen a la ciudad: la "*Fortaleza Protectora Argentina*". Ocupaba, con sus edificios complementarios, cuatro manzanas enmarcadas hoy por el perímetro de calle Estomba, Chiclana, O'Higgins, Brown, Vieytes y Moreno.

### **Líneas de Investigación y Desarrollo**

La propuesta se enmarca en un plan de comunicación a largo plazo, implementando las TICs a la difusión

turística del patrimonio cultural de Bahía Blanca. Como primera aproximación, se presentan tres ejes de investigación, sobre los que se trabajan en la actualidad. Cada uno de ellos se ubica dentro de un programa de mayor alcance a largo plazo. A continuación se describen cada uno de ellos. Tales ejes son:

1. "Difusión del patrimonio cultural a través de Realidad Virtual no Inmersiva".
2. "Difusión del Patrimonio Cultural utilizando Realidad Aumentada".
3. "Difusión del Patrimonio Cultural a través de la Reconstrucción Virtual de Obras".

#### **Difusión del patrimonio cultural a través de Realidad Virtual no Inmersiva**

Este eje busca crear un recorrido virtual de toda la ciudad de Bahía Blanca. Este recorrido incluirá tanto las calles de la ciudad como el interior de los edificios más significativos.

#### **Difusión del Patrimonio Cultural utilizando Realidad Aumentada**

Este eje busca poder masificar e implementar los códigos QR [4] en todos los edificios históricos de la ciudad y poder utilizarlos, no sólo para brindar información en forma de texto, sino con la posibilidad de agregar contenido multimedia como imágenes, videos, direcciones de Internet con información relevante, etc.

#### **Difusión del Patrimonio Cultural a través de la Reconstrucción Virtual de Obras**

El objetivo de este eje es reconstruir, a través de la realidad virtual, aquellas obras edilicias históricas de la ciudad que ya no existen o no pueden ser accedidas por el público general.

### **Resultados y Objetivos**

Sobre los tres ejes presentados se ha comenzado a generar resultados, además de continuar trabajando para alcanzar los objetivos establecidos.

### **Difusión del patrimonio cultural a través de Realidad Virtual no Inmersiva**

Durante el año 2011 se desarrolló para la Biblioteca Popular Bernardino Rivadavia [3], de la ciudad de Bahía Blanca, un recorrido virtual por sus salas.

Este recorrido permite al usuario, vía una navegador web, poder moverse por las diferentes salas disponibles en la biblioteca. Junto con la reproducción virtual de las salas de superpone información relevante para el visitante.

### **Difusión del Patrimonio Cultural utilizando Realidad Aumentada**

Sobre este eje se ha comenzado a trabajar en el uso de realidad aumentada sobre dispositivos móviles, en particular sobre el sistema operativo Android. Un primer objetivo es utilizar la Realidad Aumentada en conjunto con el sistema de transporte público de la ciudad para obtener un sistema móvil que les permita a los usuarios acceder a información asociada al transporte local.

### **Difusión del Patrimonio Cultural a través de la Reconstrucción Virtual de Obras**

La reconstrucción Virtual de la "Fortaleza Protectora Argentina", fuerte que da origen a la ciudad de Bahía Blanca, se conforma como uno de los proyectos integrantes dentro de los ejes de investigación presentados.

La inexistencia actual de esta construcción, y las evidencias asociadas, se presentan como los fundamentos para desarrollar una visualización in situ, que posibilite recrear a la obra con todas sus características y elementos representativos.

La reconstrucción de la fortaleza se realizó a partir de información histórica

consultada en ilustraciones y referencias históricas. Utilizando programas de diseño 3D fue posible reconstruir, al detalle, esta edificación. Como producto final, se desarrolló una aplicación que puede ser utilizada por el público general para poder recorrer la fortaleza. Dicha aplicación estará disponible libremente en la Web durante el corriente año.

## **Formación de Recursos Humanos**

En lo concerniente a la formación de recursos humanos se detallan las tesis concluidas y en desarrollo relacionadas con la línea de investigación presentada:

### **Tesis Concluidas**

- ▣ Tesis Doctoral. Martín L. Larrea. Tema: *Visualización Basada en Semántica*. Dirección: Dra. Silvia Castro.
- ▣ Tesis Maestría. Andrés Pinassi. Tema: *Turismo y Patrimonio. El marketing turístico del patrimonio cultural como alternativa de desarrollo del centro histórico de Bahía Blanca: una propuesta de gestión*. Dirección: Dra. Patricia Ercolani.
- ▣ Tesis de Grado. Cotroneo, Alejandro y Ibañez Bondzio. Tema: *Visualización de Caminos dentro de la Universidad Nacional del Sur*. Dirección: Lic. Sergio Martig y Dr. Martín Larrea.
- ▣ Tesis de Grado. Victoria Ganuza. Tema: *Sistema de Información para el Transporte Público de Bahía Blanca*. Dirección: Dr. Martín Larrea.
- ▣ Tesis de Grado. Augusto Montagna y Gabriel Pachiana. Tema: *Modelado 3D de Sitios Históricos*. Dirección: Dra. Silvia Castro y Dr. Martín Larrea.

### **Tesis en Desarrollo**

- ▣ Tesis Maestría. Damian Flores Choque. Tema: *Realidad Aumentada en*



Visualización. Dirección: Dra. Silvia Castro.

## Referencias

[1] ARTOPOULOS, Alejandro. "El futuro llega hace rato...: usos alternativos de la informática descentralizada en espacios urbanos". En FINQUELIEVICH, Susana y SCHIAVO, Ester (comp.). "La Ciudad y sus Tics. Tecnologías de Información y Comunicación". Bernal: Universidad Nacional de Quilmes. 1998. Pp. 53-65.

[2] BALLART HERNÁNDEZ, Josep y TRESSERRAS, Jordi. "Gestión del patrimonio cultural". Barcelona: Ariel. 2007. 3ª ed. 238 pp.

[3] Biblioteca Popular Bernardino Rivadavia – Bahía Blanca. <http://www.abr.org.ar> – Recorrido Virtual: <http://www.abr.org.ar/oldsite/virtual/index.html>

[4] Toshihiko Wakahara and Noriyasu Yamamoto. 2011. Image Processing of 2-Dimensional Barcode. In *Proceedings of the 2011 14th International Conference on Network-Based Information Systems (NBIS '11)*. IEEE Computer Society, Washington, DC, USA

## Visualización en Ciencias Geológicas

Luján Ganuza<sup>1,3</sup>, Martín Larrea<sup>1</sup>, Silvia Castro<sup>1</sup>, Ernesto Bjerg<sup>2,3</sup> y Gabriela Ferracutti<sup>2,3</sup>

<sup>1</sup>Laboratorio de Investigación y Desarrollo en Visualización y Computación Gráfica (VyGLab)

Departamento de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur.

<sup>2</sup>INGEOSUR y Departamento de Geología, Universidad Nacional del Sur.

<sup>3</sup>Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)

Bahía Blanca, Argentina

{mlg, mll, smc}@cs.uns.edu.ar, ebjerg@ingeosur-conicet.gob.ar, gferrac@uns.edu.ar

### Resumen

Los Conjuntos de datos originados desde las Ciencias Naturales, y en particular desde las Ciencias Geológicas, son cada vez más grandes, involucrando la aplicación de una gran variedad de herramientas de visualización para su análisis y exploración; tal es el caso de conjuntos de datos topográficos, proyecciones cartográficas, datos geofísicos, etc., que requieren un soporte visual adecuado para su exploración.

Esta línea de investigación propone el estudio e implementación de sistemas de visualización interactivos de datos geológicos, que provean un soporte adecuado para la exploración eficiente de los datos.

**Palabras clave:** *Visualización de datos geológicos, Representación de composiciones minerales, Conteo de Puntos.*

### Contexto

Este trabajo continúa la línea presentada en [GLMC\*12] y es llevado a cabo en conjunto por investigadores del Laboratorio de Investigación y Desarrollo en Visualización y Computación Gráfica (VyGLab) del Departamento de Ciencias e Ingeniería de la Computación de la Universidad Nacional del Sur (UNS) e investigadores integrantes del INGEOSUR (UNS-CONICET) y del Departamento de Geología de la UNS.

La línea de Investigación presentada está inserta en el proyecto “Representaciones

Visuales e Interacciones para el Análisis Visual de Grandes Conjuntos de Datos” (24/N020), dirigido por la Doctora Silvia Castro. El proyecto es financiado por la Secretaría General de Ciencia y Tecnología de la Universidad Nacional del Sur; y acreditados por la Universidad Nacional del Sur, Bahía Blanca.

### Introducción

En los distintos campos de aplicación referidos a las ciencias, el crecimiento vertiginoso de la cantidad de información genera volúmenes de datos cada vez más grandes y difíciles de comprender y analizar sin un soporte visual. Es en estos casos donde de la visualización contribuye significativamente a la exploración y entendimiento de estos conjuntos de datos, siempre y cuando se cuente con un soporte adecuado [C93] [K03] [M98] [NS79].

El campo de las Ciencias Geológicas no escapa a esta tendencia. Los Conjuntos de datos originados desde las Ciencias Naturales, y en particular desde las Ciencias Geológicas, son cada vez más grandes, involucrando la aplicación de una gran variedad de herramientas de visualización para su análisis y exploración; tal es el caso de conjuntos de datos topográficos, proyecciones cartográficas, datos geofísicos, etc., que requieren un soporte visual adecuado para su exploración.

Esta línea de investigación propone el estudio e implementación de sistemas de visualización interactivos de datos geológicos, que provean un soporte adecuado para la exploración, análisis y síntesis eficiente de los datos.

## Líneas de Investigación y Desarrollo

En el contexto de esta línea de investigación se están desarrollando los siguientes trabajos en paralelo:

1. Visualización de composiciones minerales utilizando el Prisma de los Espinelos.
2. Análisis visual automático de muestras geológicas para el conteo de puntos.

### Visualización de composiciones minerales utilizando el Prisma de los Espinelos

En este trabajo, nos enfocamos en la visualización de un conjunto de datos geológicos en particular, el conjunto de minerales que integran el grupo de los Espinelos. Este grupo de minerales resulta un excelente candidato a ser explorado y visualizado ya que presenta una gran variedad en su composición que está relacionada con su génesis, proveyendo información vital en lo referido al ambiente tectónico de las rocas presentes en determinada área en el contexto de la tectónica global [R94] [BR01]. Esta última característica es muy importante, ya que un reto considerable en el campo de las Ciencias Geológicas consiste en caracterizar una región geológica en término de su composición tectónica, convirtiendo el análisis de Espinelos en una tarea de vital importancia [L91] [CD97].

Con frecuencia, las muestras correspondientes a los Espinelos se diagraman en un ambiente prismático al que llamaremos Prisma de Espinelos. En este prisma se representan cada uno de los análisis químicos. Los análisis que corresponden a Espinelos provenientes de un determinado ambiente geológico se agrupan en un patrón de referencia de dicho ambiente.

El objetivo de este trabajo de investigación es desarrollar una herramienta de Visualización que permita visualizar y explorar conjuntos de datos correspondientes a composiciones minerales pertenecientes al grupo de los espinelos. Se pretende que esta herramienta sea interactiva, ayudando al usuario a entender fácilmente la composición mineralógica representada en la vista tanto para su identificación como para su comparación con otras muestras o conjuntos de datos.

### Análisis Visual de Muestras Geológicas para Conteo de Puntos

Una técnica utilizada para determinar las características de una muestra geológica es el conteo de puntos. El conteo de puntos es un método que se utiliza para determinar los componentes de una muestra y qué porcentaje de ella representan cada una de estas componentes.

Para poder aplicar esta técnica, el usuario debe realizar observaciones sobre la muestra a intervalos regulares; i.e. se superpone sobre la muestra una grilla regular donde cada intersección es un punto a ser analizado. El usuario debe observar cada punto y decidir qué mineral se encuentra en dicho lugar. Al contar la cantidad de puntos encontrados de cada mineral se puede calcular que porcentajes representan estos valores del total de puntos contados. La cantidad de puntos que normalmente se deben contar para una muestra varían entre 5000 y 8000. Este valor es establecido por un experto antes de iniciar el proceso de conteo de puntos.

Habitualmente esta técnica se aplica utilizando un microscopio junto con equipos especiales; estos dispositivos son extremadamente caros y el análisis completo de una muestra utilizando esta técnica puede llevar hasta 8 horas.

El objetivo de este trabajo de investigación es desarrollar una herramienta visual que asista y facilite el conteo de puntos sobre muestras geológicas.

## Resultados y Objetivos

### Visualización de composiciones minerales utilizando el Prisma de los Espinelos

A partir de este trabajo de investigación se ha desarrollado una aplicación de Visualización de Datos Geológicos llamada SpinelViz [G\*12] [GCM09]. La aplicación consiste en una vista 3D que permite al usuario observar, explorar e interactuar con varios conjuntos de datos simultáneamente en el Prisma de Espinelos. La aplicación provee la capacidad de proyectar los conjuntos de datos sobre las diferentes caras del Prisma en 2D e integrar estas proyecciones en el espacio 3D, ayudando al usuario a entender fácilmente la composición mineralógica representada en la vista.

La necesidad de lograr una visualización interactiva nos impulsa a poner énfasis en el estudio y el análisis de qué interacciones resultan importantes en función del campo de aplicación. En función de lo expuesto se está trabajando en la definición de un esquema de interacciones válido a partir de las interacciones específicas del campo de aplicación.

Habitualmente resulta de mucha utilidad utilizar este tipo de aplicaciones desde dispositivos móviles, con de capacidad de procesamiento y/o capacidad física limitada para almacenar una base de datos geológicos completa. En función de esto, se pretende desarrollar una aplicación Distribuida de Visualización de minerales utilizando el *Prisma de Espinelos*. Para esto se tendrán en cuenta diferentes modelos de distribución de carga y transferencia de datos.

### Análisis Visual de Muestras Geológicas para Conteo de Puntos

A partir de este trabajo de investigación, se desarrolló la aplicación

Rock.AR [LMC10]; este software permite realizar el conteo de puntos en forma semiautomática, asistiendo al usuario en la creación de la grilla y en el conteo mismo.

Las evaluaciones de usuarios realizadas sobre la aplicación demostraron que el tiempo de trabajo necesario para realizar el conteo de puntos se redujo de 8 horas, utilizando un microscopio, a 1 hora utilizando Rock.AR. También se observó un aumento en la satisfacción del usuario al aplicar la técnica. Actualmente se está trabajando en el desarrollo de técnicas que permitan realizar el conteo de manera automática.

## Formación de Recursos Humanos

En lo concerniente a la formación de recursos humanos se detallan las tesis concluidas y en desarrollo relacionadas con la línea de investigación presentada:

### Tesis Concluidas

- ▣ Tesis Doctoral. Martín L. Larrea. Tema: *Visualización Basada en Semántica*. Fecha de finalización: 8 de abril de 2011. Dirección: Dra. Silvia Castro.

### Tesis en Desarrollo

- ▣ Tesis Doctoral. María Luján Ganuza. Tema: *Servicios Web en Visualización*. Dirección: Dra. Silvia Castro.

## Referencias

[BR01] Barnes, S. J.; Roeder, P.L. 2001. The Range of Spinel Compositions in Terrestrial Mafic and Ultramafic Rocks. *Journal of Petrology*, vol. 42, number 12, pp: 2279-2302

[C93] Cleveland, W. S., 1993. *Visualizing Data*. Hobart Press. New Jersey, United States of America.



- [CD97] Castro, S. M.; Delrieux, C.; Larrea, M.; Silvetti, A.; 1997. Low-cost volume visualization. Proceedings International Congress on Imaging Science, Systems and Technology. CISST'97, pp. 489-493. Nevada, EEUU.
- [G\*12] Ganuza, M. L., et al., SpinelViz: An interactive 3D application for visualizing spinel group minerals. *Computers & Geosciences*, Volume 48, November 2012, Pages 50–56
- [GCM09] Ganuza, M. L.; Castro, S. M.; Martig, S. R.; Ferracutti, G.; Bjerg, E.; 2009. Mineral Compositions Visualization Implementig the Spinel Prism. Congreso Argentino de Ciencias de la Computación Proceedings. CACIC 2009, ISBN 978-897-24068-4-1, pp. 576-585. Jujuy, Argentina.
- [GLMC\*12] Luján Ganuza, Martín L. Larrea, Sergio R. Martig, Silvia M. Castro, Ernesto Bjerg y Gabriela Ferracutti. Visualización en Ciencias Geológicas. XIV Workshop de Investigadores en Ciencias de la Computación (WICC 2012), pp. 307–310. Universidad Nacional de Misiones, Posadas, Misiones (abril 2012). ISBN 978-950-766-082-5
- [K03] Koutek, M., 2003. Scientific Visualization in Virtual Reality: *Interaction Techniques and Application Development*. Computer Graphics & CAD/CAM group, Faculty of Information Technology and Systems (ITS), Delft University of Technology (TU Delft).
- [LMC10] Larrea, M., Martig, S., Castro, S., Aliani, P., Bjerg, E. 2010. Rock.AR – A Point Counting Application for Petrographic Thin Sections”. 26<sup>th</sup> String Conference on Computer Graphics, Budmerice, Slovakia.
- [L91] Lindsley, D. H. Ed. 1991. Oxide Minerals: petrologic and magnetic significance. Mineralogical Society of America, Department of Geological Sciences, Virginia Polytechnic Institute and State University, Blacksburg, Virginia. vol. 25, p. 509.
- [M98] McCormick, B. H., 1998. Visualization in scientific computing. SIGBIO News. ACM, Vol.10, pp. 15-21.
- [NS79] Nielson, G. M.; Shriver, B.; Rosenblum, Lawrence. 1979. Visualization in Scientific Computing. IEEE Computer Society Press. United States of America.
- [R94] Roeder, P. L. 1994. Chromite: from the fiery rain of chondrules to the Kilauea Iki lava lake. *Canadian Mineralogist* 32, pp. 729-746.

# Reconocimiento de Acciones en Videos de Tenis usando Flujo Óptico y CRF

José Francisco Manera<sup>†‡</sup>, Jonathan Vainstein<sup>†‡</sup>, Claudio Delrieux<sup>†</sup>, y Ana Maguitman<sup>‡</sup>

josemanera@gmail.com, jjvainstein@gmail.com, cad@uns.edu.ar, agm@cs.uns.edu.ar

<sup>†</sup> Laboratorio de Ciencias de las Imágenes (IIIE - CONICET)

Departamento de Ingeniería Eléctrica y de Computadoras (DIEC)

<sup>‡</sup> Grupo de Investigación en Administración de Conocimiento y Recuperación de Información - LIDIA

Departamento de Ciencias e Ingeniería de la Computación (DCIC)

Universidad Nacional del Sur (UNS)

Av. Alem 1253, (B8000CBP), Bahía Blanca, Argentina

Tel: (0291) 459-5135 / Fax: (0291) 459-5136

## Resumen

El objetivo del Reconocimiento de Acciones (*Action Recognition*) es el análisis e interpretación automatizados de eventos particulares en secuencias de video. Esta área está siendo ampliamente investigada en diferentes dominios tales como videos de seguridad, interacción humano-computadora, monitoreo de pacientes y recuperación de video, entre otros, dadas las importantes aplicaciones que pueden desarrollarse, y la proliferación de cámaras y videos de seguridad y monitoreo en la actualidad. El objetivo de este proyecto es la identificación automática de acciones en secuencia de videos, utilizando *Conditional Random Fields* (CRFs). Como caso de estudio se utilizan videos de partidos de tenis para la identificación de golpes. Se abordan tres desafíos, el *tracking*, la representación del movimiento del jugador y el reconocimiento de acciones.

**Palabras clave:** Reconocimiento de Acciones, Tracking, Conditional Random Fields, Flujo Óptico

## 1. Contexto

El presente proyecto se da en el marco de la colaboración conjunta desarrollada por el Laboratorio de Ciencias de las Imágenes (IIIE-CONICET

<http://www.imaglabs.org>) y el Grupo de Investigación en Administración de Conocimiento y Recuperación de Información (<http://ir.cs.uns.edu.ar>), pertenecientes a la Universidad Nacional del Sur. Esta línea de investigación se lleva a cabo dentro del ámbito del LCI, y está asociada a los siguientes proyectos de investigación:

- Procesamiento inteligente de imágenes. PGI 24/K047 (SECyT-UNS). Director: Claudio Delrieux.
- PICT Start-Up 2442/2010 (ANPCyT). Director Claudio Delrieux.
- Soporte inteligente para el acceso a información Contextualizada en entornos centralizados y distribuidos. PIP: 11220090100863 (CONICET). Director: Ana G. Maguitman.

## 2. Introducción

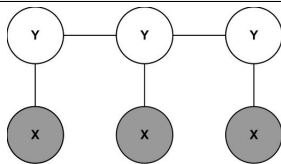
En los últimos años se ha producido un gran crecimiento en la disponibilidad de información multimedia, lo que ha motivado en gran medida el desarrollo del reconocimiento de acciones en video dentro del área de visión por computadora. En función del gran volumen de información se hace necesario desarrollar sistemas automáticos o semi-automáticos que permitan el etiquetado de acciones en video

con diferentes aplicaciones, como por ejemplo la detección de acciones sospechosas en cámaras de vigilancia [17], detección de incidentes de tránsito [9] y búsqueda de acciones en videos deportivos [21].

Este trabajo aborda la temática del análisis de secuencias de video de partidos de tenis. Este tipo de análisis ha sido ya estudiado por diversos autores. En [11] se desarrolla un sistema de anotación automático de acciones en partidos de tenis basado en transiciones de siluetas. En [5] se utiliza un descriptor de movimiento basado en el flujo óptico espacio-temporal junto con un clasificador de vecino más cercano para llevar a cabo la categorización de acciones. En [20] se utiliza el flujo óptico como descriptor de los movimientos de los jugadores y emplea Support Vector Machines para entrenar al clasificador en donde la información del flujo óptico es utilizada como dato de entrada del mismo.

Los CRFs son un modelo probabilístico discriminativo para el etiquetado de secuencias. Este modelo condiciona las probabilidades a la secuencia de observaciones, lo cual evita computar las probabilidades para cada posible observación de la secuencia. En lugar de depender de las probabilidades conjuntas  $P(X, Y)$ , los CRFs especifican la probabilidad de secuencias de etiquetas posibles dada la observación  $P(Y|X)$ . Un modelo gráfico típico de CRF es ilustrado en la Fig. 1, donde  $X$  e  $Y$  refieren a las observaciones y las secuencias de etiquetas respectivamente.

**Figura 1** Modelo gráfico de un CRF. Los nodos rotulados con X corresponden a observaciones y los rotulados con Y a etiquetas.



Los CRFs han sido aplicados a una variedad de dominios, tales como procesamiento de lenguaje natural [15, 16, 10, 4], bioinformática [14, 18] y visión por computadora. En esta última área algunos autores han utilizado CRFs para el etiquetado de imágenes [7] y para el reconocimiento de objetos. En [13] se utilizan CRFs para determinar partes características de un objeto. En particular, para el caso de reconocimientos de acciones en secuencias de video hay diversos trabajos sobre detección de

acciones en video deportivos [8, 20].

El objetivo de este trabajo consiste en diseñar e implementar un sistema de identificación de acciones, que utilice procesamiento de video y reconocimiento basado en CRF. Como caso de entrenamiento, se propone entrenar al sistema para clasificar clips de videos de tenis, según el tipo de golpe llevado a cabo en cada video.

### 3. Líneas de investigación y desarrollo

La línea de investigación se centra en el diseño e implementación de un sistema de reconocimiento de acciones en videos de tenis, más concretamente se avanzará en el estudio de técnicas de procesamiento de video y de CRFs con el objetivo de desarrollar y mejorar nuevas técnicas en ambas áreas. Existen varios desafíos a ser abordados en el desarrollo de esta línea de investigación: tracking, extracción de *features*, clasificación y reconocimiento de patrones.

Se han abordado dos desafíos complementarios: la representación del movimiento del jugador y el reconocimiento de acciones. Para el primero se propone la utilización del flujo óptico [6] para modelar los patrones de movimiento del jugador en el campo de juego. Para el reconocimiento de acciones se propone el uso de CRFs utilizando la información obtenida a partir del flujo óptico de los sucesivos *frames* como atributos de entrada del clasificador.

### 4. Propuesta y Metodología

Como en todo proceso de clasificación supervisada, se pueden describir claramente dos grandes etapas. La primera consiste en la construcción de un modelo de clasificación a partir del entrenamiento realizado con un conjunto de muestras clasificadas de forma supervisada, y una segunda etapa en la cual se lleva a cabo la clasificación de una nueva instancia, empleando para ello el modelo previamente obtenido.

#### 4.1. Entrenamiento

La fase de entrenamiento se compone de un pipeline cuyas etapas son: tracking, extracción de *features* y construcción del clasificador. Los videos utili-

zados como entrada para esta etapa corresponden a capturas de video de dominio público realizadas con una cámara oblicua. Estos videos fueron previamente clasificados en forma supervisada en dos clases: *drives* de derecha y *drives* de izquierda. Para esta etapa se utiliza la biblioteca OpenCV.

La primera etapa tiene como objetivo llevar a cabo el tracking del jugador que se encuentra de espaldas a la cámara, seleccionándolo de manera supervisada en el frame 0 del video. A partir de la selección de esta región de interés se genera un modelo del jugador que se compone de dos histogramas. El primero consiste de los valores de luminosidad correspondientes a los pixeles de la ropa del jugador. Este histograma se calcula a partir de la imagen obtenida al aplicar una máscara que elimina los pixels que no corresponden a la ropa del jugador. El segundo histograma se obtiene a partir de la imagen que resulta de aplicarle una máscara que elimina los pixels que no corresponden a la piel del jugador. Para este histograma se utiliza el canal Hue del espacio cromático HSV [19]. En el frame 1 se toma la misma posición de la región de interés correspondiente al frame 0, y para cada uno de los histogramas antes descritos se aplica el siguiente algoritmo [1]:

1. Para cada pixel de la imagen se toma su valor y se busca el recipiente (*bin*) correspondiente en el histograma.
2. Se toma el valor asociado al bin seleccionado.
3. Se almacena el valor del bin en una nueva imagen.

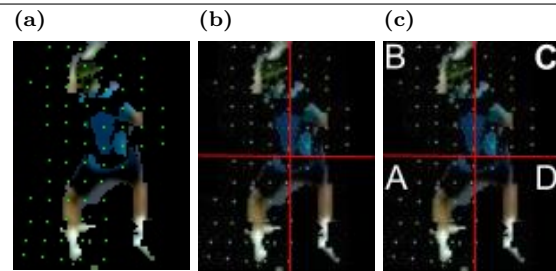
Los valores almacenados en cada imagen de salida representan la probabilidad de que un pixel en la imagen de entrada pertenezca a la zona de interés representada por el histograma usado (en este caso piel y color de la ropa respectivamente).

Luego se suman las imágenes obtenidas y su resultado -junto a la región de interés del frame anterior- son utilizadas como entradas para el algoritmo de Meanshift [3]. Como resultado de dicho algoritmo se obtiene una nueva región de interés que corresponde a la posición del jugador en el frame actual. Este proceso se repite para todos los frames del video.

Luego, para describir de manera robusta y discriminativa cada una de las clases de golpes se utiliza

el flujo óptico (Fig. 2a), el cual es calculado utilizando el algoritmo Gunnar Farnebäck [6]. La matriz de desplazamientos obtenida se divide en cuatro regiones (Fig. 2b) para luego generar una etiqueta para cada parte que representa la variación del flujo óptico en cada región (Fig. 2c). Este último paso tiene como objetivo discretizar los datos del flujo óptico a fin de flexibilizar la coincidencia de atributos en las etapas de construcción y validación del clasificador.

Figura 2 Flujo Óptico



Finalmente, para llevar a cabo la construcción del clasificador que nos permita realizar la tarea de reconocer los dos tipos de golpes propuestos en este trabajo se utiliza Conditional Random Fields. Este clasificador es entrenado utilizando el conjunto de entrenamiento obtenido en los pasos anteriores. Para las etapas de entrenamiento y clasificación se utiliza la herramienta CRFSuite [12]. La entrada a dicha herramienta está compuesta por un archivo de texto plano, en el cual cada clip de video usado para la etapa de entrenamiento es representado por tantas líneas como frames tenga el mismo. A su vez, cada línea está compuesta por 5 columnas. La primera corresponde a la etiqueta del frame (es decir, a qué tipo de golpe corresponde el frame), y las restantes 4 columnas corresponden a la representación del flujo óptico dado en este trabajo.

## 4.2. Clasificación y Validación

Esta etapa consiste en la clasificación de un video de entrada en una de las clases definidas en el proceso de entrenamiento. El video de entrada es procesado empleando el tracking y la extracción de features desarrollados anteriormente. Estos característicos son formateados de forma adecuada para ser utilizados como entrada al clasificador, resultando como salida la clase a la cual pertenece el video de entrada.



Para determinar la efectividad de los métodos propuestos se utilizan las medidas de precisión, recall y F1 [2].

## 5. Resultados y Objetivos

### 5.1. Resultados Preliminares

Se llevó a cabo una primera prueba que consistió en entrenar un clasificador utilizando un conjunto de 20 clips de video (10 clips representando drives de izquierda, y 10 clips de drives de derecha). Luego se validó dicho clasificador utilizando 22 clips de video (11 clips de cada tipo de golpe). La matriz de confusión se encuentra en la Tabla 1. A partir de los datos de la matriz de confusión se pueden inferir los valores de precisión: 86.36%, recall: 85.57% y F1: 85.3%.

		Clase predicha	
		DI	DD
Clase real	DI	0.73	0.27
	DD	0	1

Tabla 1: Matriz de confusión. DI: Drive izquierda. DD: Drive derecha.

### 5.2. Trabajo Futuro

Como trabajo futuro se espera avanzar y mejorar en los siguientes aspectos: tracking, extracción de features para el modelado de las acciones, representación de los datos de entrada para CRFs, implementación propia de CRFs y reconocer una mayor cantidad de golpes.

## 6. Formación de Recursos Humanos

El equipo de trabajo de esta línea de investigación se encuentra integrado por dos becarios de posgrado que cuentan con una beca interna del Conicet, y los respectivos directores. Por otra parte se cuenta con la colaboración de otros becarios de posgrado del LCI y una vinculación con un grupo de trabajo sobre minería de datos compuesto por doctorandos e investigadores formados.

Como parte de las actividades asociadas al proyecto se realizan cursos de grado y postgrado en Procesamiento de Imágenes, Minería de Datos y Aprendizaje Supervisado.

## Referencias

- [1] Opencv: Backprojection.
- [2] Precisión, recall y f1, 2013 (accessed March 9, 2013).
- [3] Dorin Comaniciu, Visvanathan Ramesh, and Peter Meer. Kernel-based object tracking. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(5):564–575, may 2003.
- [4] Aron Culotta, Ron Bekkerman, and Andrew McCallum. Extracting social networks and contact information from email and the web. In *In Proceedings of CEAS-1*, 2004.
- [5] Alexei A. Efros, Alexander C. Berg, Greg Mori, and Jitendra Malik. Recognizing action at a distance. In *Proceedings of the Ninth IEEE International Conference on Computer Vision - Volume 2*, ICCV03, pages 726–, Washington, DC, USA, 2003. IEEE Computer Society.
- [6] Gunnar Farnebäck. Two-frame motion estimation based on polynomial expansion. In Josef Bigun and Tomas Gustavsson, editors, *Image Analysis*, volume 2749 of *Lecture Notes in Computer Science*, pages 363–370. Springer Berlin Heidelberg, 2003.
- [7] Xuming He, Richard S. Zemel, and Miguel Á. Carreira-Perpiñán. Multiscale conditional random fields for image labeling. In *Proceedings of the 2004 IEEE computer society conference on Computer vision and pattern recognition*, CVPR'04, pages 695–703, Washington, DC, USA, 2004. IEEE Computer Society.
- [8] Nisha Jain, Santanu Chaudhury, Sumantra Dutta Roy, Prasenjit Mukherjee, Krishanu Seal, and Kumar Talluri. A novel learning-based framework for detecting interesting events in soccer videos. In *Proceedings of the 2008 Sixth Indian Conference on Computer Vision, Graphics & Image Processing*, ICVGIP08, pages 119–125, Washington, DC, USA, 2008. IEEE Computer Society.
- [9] Shunsuke Kamijo, Yasuyuki Matsushita, Katsushi Ikeuchi, and Masao Sakauchi. Incident detection at intersections utilizing hidden markov model, 1999.

- [10] Andrew McCallum. Efficiently inducing features of conditional random fields. In *Proceedings of the Nineteenth Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-03)*, pages 403–410, San Francisco, CA, 2003. Morgan Kaufmann.
- [11] Hisashi Miyamori and Shu-ichi Iisaku. Video annotation for content-based retrieval using human behavior analysis and domain knowledge. In *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition 2000, FG00*, pages 320–, Washington, DC, USA, 2000. IEEE Computer Society.
- [12] Naoaki Okazaki. Crfsuite: a fast implementation of conditional random fields (crfs), 2007.
- [13] Ariadna Quattoni, Michael Collins, and Trevor Darrell. Conditional random fields for object recognition. In *In NIPS*, pages 1097–1104. MIT Press, 2004.
- [14] Kengo Sato and Yasubumi Sakakibara. Rna secondary structural alignment with conditional random fields. In *ECCB/JBI*, page 242, 2005.
- [15] Fei Sha and Fernando Pereira. Shallow parsing with conditional random fields. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - Volume 1, NAACL '03*, pages 134–141, Stroudsburg, PA, USA, 2003. Association for Computational Linguistics.
- [16] Charles Sutton, Andrew McCallum, and Khashayar Rohanimanesh. Dynamic conditional random fields: Factorized probabilistic models for labeling and segmenting sequence data. *Journal of Machine Learning Research*, 8:693–723, March 2007.
- [17] M. Takahashi, M. Naemura, M. Fujii, and S. Satoh. Human action recognition in crowded surveillance video sequences by using features taken from key-point trajectories. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on*, pages 9–16, June.
- [18] Peter Weigle Yan Liu, Jaime Carbonell and Vanathi Gopalakrishnan. Protein fold recognition using segmentation conditional random fields (scrfs). *Journal of Computational Biology*, 13 (2) 394–406, 2006.
- [19] Benjamin D. Zaitz, Boaz J. Super, and Francis K. H. Quek. Comparison of five color models in skin pixel classification. In *In ICCV 99 Intl Workshop on*, pages 58–63, 1999.
- [20] Guangyu Zhu, Changsheng Xu, Wen Gao, and Qingming Huang. Action recognition in broadcast tennis video using optical flow and support vector machine. In *Proceedings of the 2006 international conference on Computer Vision in Human-Computer Interaction, ECCV06*, pages 89–98, Berlin, Heidelberg, 2006. Springer-Verlag.
- [21] Guangyu Zhu, Changsheng Xu, and Qingming Huang. Player action recognition in broadcast tennis video with applications to semantic analysis of sports game. In *in Proc. ACM Multimedia, 2006*, pages 431–440, 2006.

# Simulación de personajes conversacionales virtuales dentro de un entorno de Realidad Virtual

Y. Alvarado, M. Zúñiga, F. Pahud, J. Fernández, R. Guerrero

*LIDIC- Universidad Nacional de San Luis*

*Ejército de los Andes 950*

*Tel: 0266 4420823, San Luis, Argentina*

*{yalvarado, mezuniga, ferpah, jmfer, rag}@unsl.edu.ar*

## Resumen

En la actualidad, la popularización y expansión de las tecnologías de la comunicación y la información han alterado las formas tradicionales de comunicación de las personas convirtiendo a las tecnologías en canales casi obligados del proceso. Particularmente, las situaciones de adquisición y transmisión de conocimiento han obligado a desarrollar nuevas formas de visualización e interacción que faciliten la transmisión de información. La generación de entornos de interacción que provean nuevos contextos de intercambio y comunicación de información es uno de los fines básicos del área de Realidad Virtual.

La virtualidad consiste en hacer calzar un mundo virtual en un mundo real. Uno de los desafíos más importantes radica en preservar las vías naturales de comunicación, interacción y entendimiento de las personas proveyendo a las computadoras de una visualización e interacción lo más humana posible.

Proveer a las computadoras las capacidades de comunicación de los humanos involucra dotarlas entre, otras cosas, de conocimiento permitiendo automatizar el intercambio de información y otorgando comportamiento y razonamiento en función del contexto.

Esta propuesta de trabajo establece los lineamientos a seguir para la exploración de las sinergias existentes entre las nuevas tecnologías asociadas a la simulación de personajes conversacionales virtuales dentro de un entorno de Realidad Virtual inmersiva.

**Palabras Claves:** Realidad Virtual, Computación Gráfica, Interfaces Humano-Computadoras, Ontologías, Semántica.

## Contexto

La propuesta de trabajo se lleva a cabo dentro de la línea de Investigación "Procesamiento de Información Multimedia" del proyecto "Nuevas Tecnologías para un tratamiento integral de Datos Multimedia". Este proyecto es desarrollado en el ámbito del Laboratorio de Investigación y Desarrollo en Inteligencia Computacional (LIDIC) de la Universidad Nacional de San Luis.

## 1. Introducción

El constante desarrollo tecnológico ha ocasionado la aparición de nuevas tecnologías emergentes que generan un gran volumen de información y conocimiento, obligando a desarrollar nuevas estrategias para llevar a cabo las actividades tradicionales de adquisición y transmisión del conocimiento y modificando la

manera en que el ser humano ha pasado a comunicarse con sus semejantes; convirtiendo a la información y al conocimiento en primordiales para el avance social [1, 2].

Del conjunto de las nuevas tecnologías emergentes hay dos que se destacan como revoluciones tecnológicas importantes. La primera de ellas es la red de redes: INTERNET, que de ser un mero apéndice de la telefonía de cobre se ha transformado en una infraestructura básica para la Sociedad del Conocimiento y motor de la Economía mundial, de tal manera que forma ya parte de la vida de cualquier persona de una manera tan natural, que es como si siempre hubiera estado allí. La segunda revolución tecnológica la ha propiciado la Realidad Virtual, la percepción en 3D de entornos simulados que permiten trasladar al usuario a mundos de ensueño y le posibilitan viajar a través del tiempo al pasado y al futuro [2].

Las aplicaciones de Realidad Virtual (RV) sumergen al usuario en un ambiente generado por computadora que simula la realidad mediante el uso de dispositivos interactivos que envían y reciben información (gafas, cascos, guantes, trajes). La “telepresencia” o ilusión de “estar allí” es controlada por sensores que capturan las acciones del usuario y ajustan en forma acorde lo que es visualizado en pantalla en tiempo real. De esta manera el usuario puede realizar recorridos virtuales de ambientes 3D simulados, al mismo tiempo que experimenta la sensación de tocar, capturar y manipular los objetos que está visualizando [3].

Una importante área de aplicación de los sistemas de RV ha sido siempre el entrenamiento en actividades de alto riesgo de la vida real, creando un ambiente que simule la situación peligrosa de la vida real deseada con el fin de adquirir conocimiento y experiencia sin necesidad de estar inmerso en dicho ambiente peligroso. La sensación provista por las simulaciones es tal que proporcionan un entrenamiento similar o casi similar a la práctica con sistemas reales reduciendo costo, tiempo y riesgo asociado (simuladores de vuelo, de manejo de trenes) [4]. El desafío de tales sistemas se haya en dominar los elementos de la simulación, desde las

leyes físicas que modelan un evento, hasta su visualización gráfica, todo ello a través de la resolución mediante un programa informático. Sin embargo, los avances logrados en la simulación de entornos digitales, en la implementación de Interfaces Humano-Computadoras y el concepto de Realidad Aumentada [3, 20], permiten en la actualidad el desarrollo de sistemas de control inmersivo y en tiempo real no solamente para entrenamiento de actividades de la vida real sino también para tareas de investigación donde es necesario satisfacer demandas de visualización e interacción en modelos de la realidad que manejan grandes cantidades de información [5, 6].

La virtualidad establece una nueva forma de relación entre el uso de las coordenadas de espacio y de tiempo, superando las barreras espacio-temporales y configurando un entorno en el que la información y la comunicación se muestran accesibles desde perspectivas hasta ahora desconocidas, al menos en cuanto a su volumen y posibilidades. En conclusión, la Realidad Virtual permite la generación de entornos de interacción que facilitan nuevos contextos de intercambio y comunicación de información.

Las personas, las organizaciones y los programas de computadoras deben comunicarse, aunque sus necesidades y experiencias establecen diferentes puntos de vistas. Esta divergencia es natural y valiosa, no obstante conduce a problemas de comunicación, interacción y entendimiento. La interacción inteligente humano-computadora es un campo de investigación emergente que tiene por objetivo proveer al hombre de vías naturales de comunicación con un computadora, a modo de convertirlas en herramientas de ayuda. El supuesto colectivo establece que para que una computadora sea capaz de interactuar con un humano ésta debe tener las capacidades de comunicación de los humanos.

Los personajes conversacionales virtuales son interfaces gráficas capaces de utilizar modos de comunicación verbal y no verbal para interactuar con los usuarios de ambientes generados por computadoras. Estos personajes suelen



ser una simple cara animada la cual reproduce expresiones faciales simples o, en ocasiones asociadas a la síntesis del habla, visualizaciones donde se utilizan métodos de sincronización de labios y representaciones gráficas tridimensionales sofisticadas con movimientos complejos del cuerpo, expresiones faciales y emocionales. En este contexto, los presentadores virtuales son los más usualmente utilizados. Existen en la literatura muchos ejemplos de personajes virtuales con apariencia humana que realizan presentaciones: el reportero virtual del clima Noma and Badler [7], el reportero de TV de Thalmann and Kalra [8], el avatar 3D de Baus et al [9], el presentador virtual para reuniones virtuales de Nijholt et al [10], entre otros [11, 12, 13, 14]. Todos ellos son una opción prometedora para el desarrollo de interfaces ya que se basan en el estilo de comunicación e interacción con el cual los humanos están bastante familiarizados.

No obstante, la más importante de las capacidades de comunicación de los humanos es el conocimiento, el cual ya es por sí mismo una parte crucial de la inteligencia humana y por consiguiente un área de investigación prioritaria. [15]. La ontología y la semántica han surgido como un tópico de investigación fundamental en el área de los sistemas de información para la representación del conocimiento; haciendo explícito el significado e intercambio de información y el logro de interoperatividad [16]. Las ontologías y las semánticas han sido utilizadas como soporte de una gran variedad de tareas en diferentes campos tales como la Semántica Web, e-Business, integración de información, minería de datos, diseño de sistemas, etc. [17]. En consecuencia, su uso puede extenderse más allá de los sistemas de información y puede aplicarse para proveer conocimiento semántico a un personaje conversacional virtual. En un sistema de estas características las ontologías brindarían la base para la configuración dinámica del sistema, la automatización del intercambio de información y, el razonamiento en función al comportamiento y el contexto, permitiendo crear aplicaciones dinámicas y concientes del entorno.

## 2. Líneas de Investigación y Desarrollo

En función de lo anteriormente expresado, un personaje conversacional virtual debería poder brindar una interfaz que permita la comunicación con el usuario mediante el uso de diferentes lenguajes de comunicación (verbales y no verbales). Para ello la problemática se resume en el hecho de dominar los elementos de una simulación, desde las leyes físicas que modelan una acción, hasta su visualización gráfica, pasando por la resolución mediante un programa informático, en tiempo real, y dentro de un escenario virtual inmersivo del tipo CAVE (Computer Automatic Virtual Environment). Claramente, se pueden definir tres grandes líneas de investigación a seguir:

- *La visualización gráfica tridimensional del personaje.* Simular la apariencia externa de los personajes virtuales intentando hacerlos indistinguibles de los humanos es un desafío técnico constante el cual que se va consiguiendo a medida que evolucionan no solamente las técnicas de visualización gráfica tridimensional, sino también la técnicas de simulación de los movimientos del cuerpo, las expresiones faciales y emocionales [3, 4].
- *La comunicación verbal y no verbal.* La creación de personajes no se limita solamente a la apariencia externa en hiperrealismo o en movimientos y expresiones. Es necesario potenciar los canales de entrada y salida de información mediante el enriquecimiento de los medios de comunicación verbal (lenguaje escrito, lenguaje sonoro) y no verbal (lenguaje de señas, lenguaje sonoro). Esto último implica desde la simple incorporación de sonido asociado a hechos o eventos, pasando por la animación de manos para la síntesis del lenguaje de señas, hasta la sincronización de labios para la síntesis del habla [18, 19].
- *La incorporación de comportamien-*

*to/razonamiento*. Si bien la apariencia y capacidad de interacción de los personajes son de relevancia al momento de la comunicación de información, es de gran importancia también el poder otorgarles la potestad de elaborar información y reaccionar en forma autónoma. Una versión simplificada del concepto hace referencia a *Agentes Virtuales Autónomos* los cuales poseen la capacidad de desenvolverse por sí mismos dentro del ambiente en que se encuentran inmersos, actuando y reaccionando a su entorno. No obstante se pretende que dichos agentes exhiban conductas más complejas basadas en estados emocionales y como resultado de un proceso de razonamiento en concordancia con su situación dentro del entorno; tal como se esperaría de un humano real. Estos personajes reciben el nombre de *Humanos Virtuales Autónomos* [15, 16].

Si bien el trabajo se centrará en el desarrollo de un personaje conversacional, la tecnología y modelos a desarrollar podrán adaptarse perfectamente a aplicaciones en ambientes inteligentes tales como guía de turismo para visitantes de museos, o aplicaciones de domótica, entre otras.

### 3. Resultados obtenidos / esperados

El grupo de trabajo, además de pertenecer a un proyecto de investigación de la Universidad Nacional de San Luis, se encuentra desarrollando tareas dentro del marco de un Proyecto ALFA III de la Comunidad Europea, denominado GAVIOTA (Grupos Académicos para la Visualización Orientada por Tecnologías Apropriadas), en el que participa la UNSL en conjunto con otras universidades de América Latina y Europa.

En función de ello, se ha desarrollado una investigación para conocer el estado del arte a nivel mundial y principales enfoques, métodos

y técnicas existentes en relación con el comportamiento autónomo de las entidades de un escenario virtual con el propósito de incorporar a futuro nuevos conceptos a los ya existentes, tanto matemáticos como basados en la física.

Actualmente se están realizando desarrollos con el objeto de simular un sistema de RV para la prevención de accidentes de tránsito urbano con escenarios experimentales de la ciudad de Concepción (Chile). Este trabajo se ha abordado en forma conjunta con la Universidad de Bio Bio (Chile).

### 4. Formación de Recursos Humanos

Los trabajos preliminares de estudio del arte han permitido la realización de trabajos de fin de carrera de la Licenciatura en Ciencias de la Computación, así como también la definición de un trabajo de tesis de Maestría en Ciencias de la Computación. Asimismo se ha obtenido una beca de finalización de carrera otorgada por la Secretaría de Ciencia y Técnica de la Fac. de Cs. Físico Matemáticas y Naturales de la UNSL.

### Referencias

- [1] Bowman D., Kruijff E., LaViola J., and Poupyrev I. "3D User Interfaces: Theory and Practice", ISBN-10: 9780201758672, Addison-Wesley Professional, 1 edition, 2004.
- [2] Lucas H., "Inside the Future: Surviving the Technology Revolution", ISBN-13: 978-0313348266, Addison-Wesley Professional, 1 edition, 2008.
- [3] Craig A., Sherman W., and Will J., "Developing Virtual Reality Applications: Foundations of Effective Design", ISBN: 0123749433, Addison-Wesley/Pearson Education, 2009.
- [4] Stanney K.M., Mourant R.R., and Kennedy R.S., "Human factor issues in virtual environments: A review of the literature", Presence, 7(4), 327-351, 2007.

- [5] Salzman M.C., Dede C., McGlynn D., and Loftin R.B., "ScienceSpace: Lessons for designing immersive virtual realities", In Proceedings of the CHI'08: ACM Conference on Human Factors in Computing Systems, (pp. 89-90), New York, 2008.
- [6] Winn W. "The impact of three-dimensional immersive virtual environments on modern pedagogy", HITL Technical Report R-97-15: Seattle, WA: University of Washington, Human Interface Technology Laboratory, 2008.
- [7] Noma T., Zhao L., and Badler N.I., "Design of a Virtual Human Presenter", IEEE Journal of Computer Graphics and Applications, Vol.20, No.4, July/August, 2000, pp. 79-85, 2000.
- [8] Magnenat-Thalmann N., and Kalra P., "The Simulation of a Virtual TV Presenter", Proc. Pacific Graphics 95, World Scientific, Singapore, 1995, pp. 9-21, 1995.
- [9] Baus J., Butz A., and Krüger A., "Incorporating a Virtual Presenter in a Resource Adaptive Navi-gational Help System", Workshop notes of the Workshop on Guiding Users through Interac-tive Experiences, Paderborn, Germany, April, 2000.
- [10] Nijholt A., van Welbergen A., and Zwiers J., "Introducing an embodied virtual presenter agent in a virtual meeting room", Proc. of the 23rd IASTED International Multi-Conference Artificial Intelligence and Applications, February, 2005 Innsbruck, Austria, pp. 579-584, 2005.
- [11] Rist T., André E., Baldes S., Gebhard P., et. al., "A Review of the Development of Embodied Presentation Agents and Their Application Fields in Life-Like Characters: Tools, Affective Functions, and Applications", Helmut Prendinger and Mitsuru Ishizuka (eds), Springer Series on Cognitive Technologies, pp. 377-404, 2003.
- [12] Rojc M., Rotovnik T., Brus M., Jan D., and Zdravko K., "Embodied Conversational Agents in Wizard-of-Oz and Multimodal Interaction Applications", Verbal and Nonverbal Communication Behaviours, Lecture Notes in Computer Science, vol. 4775: 294-309, Springer Berlin Heidelberg, ISBN 978-3-540-76441-0, 2007.
- [13] Mancini M., and Pelachaud C., Generating distinctive behavior for Embodied Conversational Agents, Journal on Multimodal User Interfaces, vol. 3(4): 249-261, Springer-Verlag, ISSN 1783-7677, 2009.
- [14] Cerekovic, A., and Pandzic. I., "Multimodal behavior realization for embodied conversational agents", Journal on Multimedia Tools and Applications 54(1):143-164, Springer US, ISSN 1380-7501, 2011.
- [15] Jacko J., "Human-Computer Interaction. Ambient, Ubiquitous and Intelligent Interaction", Proceedings 13th International Conference, HCI International 2009, Lecture Notes in Computer Science, Vol. 5612, ISBN 978-3-642-02579-2. 2009.
- [16] Farquhar A., Fikes R., and Rice J., "The Ontolingua Server: A tool for collaborative ontology construction". International Journal of Human-Computer Studies, 46,(6):707-727. 1997.
- [17] Noy N., "Semantic Integration: A Survey Of Ontology-Based Approaches", SIGMOD Record, vol. 33, 2004.
- [18] Serna T., Gómez H., and Rose C., "Herramientas Computacionales de Signo Escritura para la Comunicación y Aprendizaje en Lengua de Señas Mexicana para Personas con Deficiencia Auditiva", Revista Ciencia y Tecnología, Instituto Tecnológico de Hermosillo, Volumen 6 No. 1, 2011.
- [19] Domínguez O., Austria A., Carrillo E, et al., "Sistema para la Rehabilitación Oral y Auditiva basado en Tecnología de Realidad Virtual", 2do. Congreso Universitario de Información y Comunicaciones, ISBN: 970-769-101-8, México, 2007.
- [20] Martínez E., "La Realidad Virtual ya permite el aprendizaje en 3D", [http://www.tendencias21.net/La-Realidad-Virtual-ya-permite-el-aprendizaje-en-tres-dimensiones\\_a2047.html](http://www.tendencias21.net/La-Realidad-Virtual-ya-permite-el-aprendizaje-en-tres-dimensiones_a2047.html), ISSN 2174-6850, 2008.

# Tratamiento de Imágenes de Fibras Animales

Leticia Constable, Marcelo Arcidiacono, Juan Vázquez, Juan Picco

**Proyectos RNA / Departamento de Ingeniería en Sistemas de Información /  
Facultad Regional Córdoba / Universidad Tecnológica Nacional**  
Cruz Roja Argentina esq. Maestro López – Ciudad Universitaria – Córdoba – Argentina

{leticiaconstable, marceloarcidiacono, jcvazquez, primejepicco}@gmail.com

## Resumen

En el marco de la sustentabilidad productiva de fibra textil de origen animal, contar con un método ágil y seguro que permita obtener una medida de la calidad de la fibra, resulta sumamente valioso para los productores rurales. Una medida de calidad de la fibra textil puede obtenerse a partir del diámetro de la misma. El laboratorio del SUPPRAD lleva a cabo un procedimiento científico innovador en la obtención de este valor en la que se recoge un corte transversal del mechón, previamente peinado y preparado con acrilatos, en un portaobjetos, para ser fotografiado con un equipo adosado a un microscopio. La imagen obtenida es manualmente procesada para determinar el diámetro promedio de la fibra. Este procedimiento manual de medición, resulta lento, engorroso e introduce error por intervención humana. Se automatiza el procedimiento por medio de un software de procesamiento de imágenes. Se comentan los resultados obtenidos y se presentan las previsiones para la continuación de este trabajo.

Palabras clave: Fibra textil, Calidad, Imágenes, Procesamiento Automático.

## Contexto

En tres proyectos consecutivos desde 2004 a la fecha [1], [2], [3], se han estudiado modelos de redes neuronales artificiales y autómatas celulares buscando relaciones entre ellos; en el proceso se ha desarrollado software y se ha aplicado lo aprendido en la solución de problemas de ingeniería de software, ciencias sociales y de la salud, y producción animal.

El trabajo que se presenta se enmarca en el proyecto incentivado denominado *Redes*

*Neuronales Artificiales y Autómatas Celulares, Productos y Aplicaciones*, del Departamento de Ingeniería en Sistemas de Información, Facultad Regional Córdoba de la Universidad Tecnológica Nacional.

Este proyecto fue evaluado y aprobado por la Secretaría de Ciencia y Técnica de UTN e incluido en el programa nacional de incentivos. Además, ha recibido financiamiento de la Agencia Córdoba Ciencia (hoy Ministerio) para la transferencia de resultados (programa PROTRI 2008), ha trabajado en conjunto con el proyecto PICT 2007-02264 del Instituto de Investigaciones Geohistóricas (IIGHI-CONICET) financiado por FONCYT y ha obtenido derechos de autor sobre parte del software desarrollado el que fue transferido a instituciones de investigación en Cuba, Paraguay, Colombia, Brasil y Argentina, a la empresa Vates S.A. de Córdoba, mediante convenio de servicios y a la Municipalidad de Villa del Totoral, Provincia de Córdoba.

En particular, y referido al tema de este artículo, el sub-proyecto RNA-SU desarrolló para el programa SUPPRAD de la Universidad Católica de Córdoba (UCC), software para tratamiento automático de imágenes de preparados de fibra animal, con el objeto de medir parámetros de calidad de las mismas.

## Introducción

### 1. *Ámbito del problema*

En la República Argentina, el Programa Nacional “Fibras Animales” considera de gran valor la producción, comercialización e industrialización de lana, mohair, cashmere, llama, guanaco y vicuña [4].

El valor de la fibra textil está dado, fundamentalmente, por su finura promedio



además de otras propiedades que hacen a establecer su cotización tales como el índice de confort PF (*Prickle Factor*) que constituye el porcentaje de fibras con diámetros mayores a 32 micrones, la presencia o ausencia de medulación<sup>1</sup>, el crimpado<sup>2</sup> y la forma y altura de las escamas [5]. Para determinar una medida satisfactoria de calidad de la fibra textil de origen animal, la característica de mayor importancia es el diámetro medio. Fibras más finas tienen más aplicaciones industriales y en consecuencia tienen mayor valor económico [6].

## 2. Descripción del problema

En nuestro país existe poca información aún sobre los valores de Coeficiente de Variación de diámetros de fibra (CV) e índice de confort [7] que permita lograr mejoras genéticas por selección y elevar el porcentaje de especímenes con diámetros menores a los 23 micrones.

Investigaciones biomecánicas más recientes, demuestran que el análisis del corte transversal provee mediciones más directas y exactas de la finura y madurez de la fibra, usualmente utilizadas para validar y calibrar otras medidas indirectas de estas propiedades esenciales [8]. A pesar de su importancia e interés, los métodos transversales para análisis de imágenes, no se aplican más ampliamente aún a las mediciones de calidad, debido al complejo procesamiento de las imágenes o por requerir de la intervención de un operador calificado.

En cada medida se tiene que tener en cuenta que dada la gran variación de diámetros que tienen las fibras animales, un gran problema es la exactitud y la precisión. [9]

La evolución en los modelos y algoritmos de procesamiento de imágenes en fibras textiles, comienzan con algunos trabajos sobre fibras de algodón. Huang et al. [10] analiza el proceso de medición en el que la imagen de una fibra en corte longitudinal, capturada con diferentes condiciones de iluminación y enfoque, es convertida a escala binaria. En otro trabajo de

Huang et al. [11], se utiliza una técnica llamada *Umbral Adaptativo*, que consiste en la colocación de un umbral dinámicamente ajustado para separar objetos de fondo.

Debido a que resulta difícil mantener bien enfocadas las imágenes y además las fibras frecuentemente se tocan entre sí, se presentan contornos borrosos y dificultades de separación que pueden generar serias distorsiones en los datos medidos. [12][13].

En trabajos posteriores, Huang et al. [14] analiza imágenes de fibras de algodón en corte transversal. Para llevar a cabo la medición se recurre a una serie de procesos computacionales como segmentación, determinación de umbral adaptativo, inundación de fondo y esqueletización, que permiten separar los objetos a medir, preservar el detalle de los bordes y finalmente, obtener medidas geométricas.

## 3. Propuesta desarrollada

El aporte fundamental en cuanto a innovación tecnológica radica en el hecho de que los instrumentos actuales de análisis de fibras son costosos y permiten obtener la medida de diámetros en forma longitudinal. En cambio, en el presente trabajo se propone un llevar a cabo un proceso de medición de diámetros en forma transversal con hardware y software de bajos costos, en forma totalmente automatizada y que puede ser llevada a cabo por personal sin capacitación técnica alguna.

En base a las investigaciones previamente citadas, se desarrolla un sistema que permite obtener una medida del radio promedio de la fibra, a partir del procesamiento automático de la imagen de un corte transversal.

Se convierten las imágenes al estándar 24 bits por píxel, es decir, un byte para cada píxel y se las somete a un tratamiento en varias etapas para lograr identificar, separar y posteriormente medir la fibra.

a) Se lleva la imagen en colores a escala de grises.

b) Se la ecualiza para obtener un histograma de distribución más uniforme y convertir la figura a blanco y negro con referencia al umbral calculado.

c) Se procede a binarizar la imagen en sus valores extremos, con el fin de obtener una

<sup>1</sup> La medulación constituye un canal hueco en el centro de la fibra que supone un problema importante para la industrialización.

<sup>2</sup> El crimpado u ondulado, se refiere a un efecto mecánico producido para lograr cohesión entre fibras y se relaciona con la capacidad hidrófuga.

imagen en blanco y negro donde se puede distinguir más claramente forma y fondo.

La figura 1 ilustra la anterior secuencia de pasos aplicados al procesamiento de la imagen de un corte transversal de fibra de guanaco y los resultados obtenidos en cada etapa.

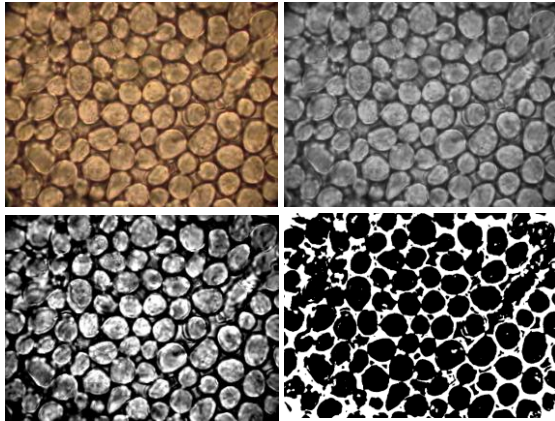


Fig. 1. Imagen superior izquierda: original. Imagen superior derecha: en escala de grises. Imagen inferior izquierda: escala de grises ecualizada. Imagen inferior derecha: binarizada.

Una vez binarizada la imagen, se procede a separar e identificar los objetos a medir.

Para ello se efectúan sucesivos barridos de la imagen binarizada, en cada uno de los cuales, se descartan primeramente las fibras que se encuentren en contacto con los bordes de la imagen ya que se desconocen sus dimensiones reales, se distingue entre fondo y forma, se rellenan sectores interiores, se seleccionan los objetos a medir tomando en consideración que todo aquello que presente interés en ser medido, no debe exceder ciertos rangos máximo y mínimo entre los cuales puede tratarse de una fibra. Por último, se identifica un centro geométrico a partir del cual se miden 32 radios como distancia a los bordes. Se calcula el radio promedio y se aproxima la figura a una circunferencia.

Las figuras 5, 6, 7 y 8 muestran una secuencia de imágenes que ilustran los pasos del proceso detallado en el párrafo anterior.

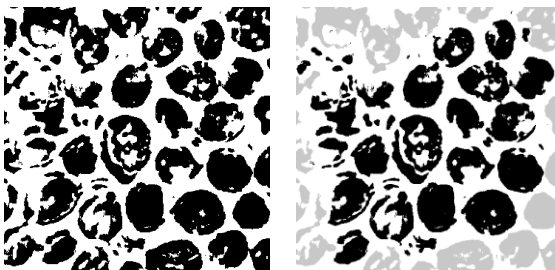


Fig. 5. Paso 1. Imagen binarizada

Paso 2. Eliminación de objetos en contacto con los bordes.

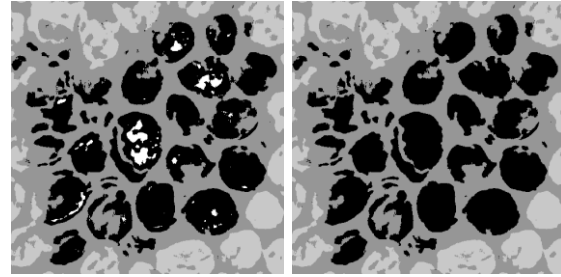


Fig. 6. Paso 3. Separación fondo-figura.

Paso 4. Relleno de blancos internos.

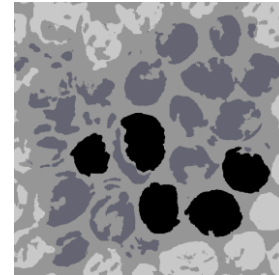


Fig. 7. Paso 5. Selección de objetos útiles a medir.

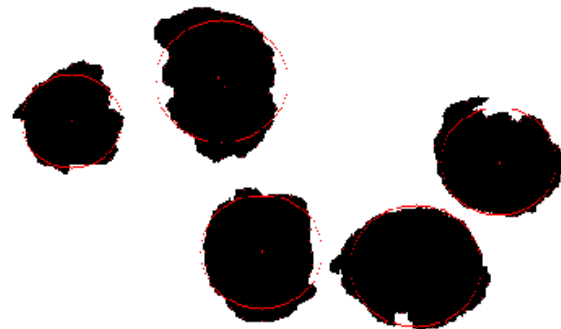


Fig. 8. Paso 6. Aproximación a circunferencia.

#### 4 *Discusión de Resultados*

El método resultó satisfactorio principalmente porque las mediciones que se obtuvieron en pixeles, con la equivalencia 2 pixeles = 1 micra de acuerdo al aumento del microscopio con el que se capturaron las imágenes, resultaron en valores adecuados para el radio promedio de las fibras, en comparación con los obtenidos a partir de otros métodos de medición en laboratorio y existe total independencia del operador en la medición, lo que asegura la precisión y exactitud requeridas. Es decir que el proceso es repetible y de exactitud conocida.

Por otra parte, como ya se dijo, se usan 32 medidas de distancia del centro geométrico de cada objeto a medir para calcular el radio de la fibra ya que experimentalmente se demuestra que no se presentan mejoras tangibles en las medidas por aumentar este número.

Atendiendo la problemática a campo que presenta la determinación de la calidad de la fibra a partir del conocimiento de su finura, el método es útil por cuanto presenta características de buena performance, bajo costo de equipamiento y no requiere operación por parte de personal calificado.

Por último, cabe destacar que el software fue desarrollado en Java lo que lo hace portable en cuanto a la plataforma y no involucra costos adicionales en licencias.

En adelante se deben centrar los esfuerzos en la evolución del sistema metrológico atendiendo tres factores principales:

- La revisión de la metodología de captura de imágenes.
- La adecuación de los algoritmos al tratamiento de fibras con otras características morfológicas.
- El reconocimiento de patrones morfológicos a través de redes neuronales.

### Línea de investigación y desarrollo

Los proyectos RNA desarrollados forman parte, con otros de la UTN-FRC, del Grupo de Investigación en Inteligencia Artificial (GIA).

En el primer proyecto se planteó la posibilidad de interpretar el entrenamiento de una red neuronal multicapa o el proceso de reconocimiento efectuado por una red Hopfield, como la evolución espacio-temporal de autómatas celulares unidimensionales, intentando descubrir un isomorfismo entre estos modelos.

En el segundo y tercer proyecto se continuó con estos estudios y además se inició el estudio de otros modelos de computación y herramientas (máquinas de vector-soporte, motor de búsqueda por similaridad, autómatas en general y algoritmos de tratamiento de imágenes), de nuevos algoritmos de entrenamiento de redes multicapa y de algoritmos para presentación gráfica de datos experimentales obtenidos.

Además, y mediante convenios con otras instituciones (UCC, IIGHI-CONICET, Vates S.A.) se inició el desarrollo de aplicaciones de lo aprendido, con software para Ciencias Sociales (cálculo de índice de riesgo para la salud de la vivienda urbana e historia clínica comunitaria y familiar), Ingeniería de Software

(estimador de esfuerzo y tiempo en proyectos de desarrollo de software) y Producción Animal (determinación de parámetros de calidad de fibras animales).

### Resultados y Objetivos

Los proyectos RNA han desarrollado software de entrenamiento y funcionamiento de redes neuronales artificiales (RVS 2.1) y de recopilación de datos para los mismos (Sistema de Historias Clínicas Familiar y Comunitaria), transferidos a instituciones de investigación, empresas y órganos de gobierno; simuladores de autómatas celulares y de redes Hopfield, un motor de búsqueda por similaridad con algoritmos innovadores para estudiar la posibilidad de emular el funcionamiento de una red multicapa, algoritmos para tratamiento de imágenes, rutinas de graficación de datos y bancos de prueba de estas cosas.

Además, ha obtenido derechos de autor sobre algunos de los productos desarrollados, ha organizado reuniones científicas sobre los temas bajo estudio, dirigido prácticas supervisadas y tesis de grado, promovido estudios de posgrado de sus integrantes y efectuado numerosas publicaciones en libros, revistas y congresos locales, nacionales e internacionales.

En particular, en la aplicación al estudio de la calidad de fibras textiles de origen animal, foco principal de este artículo, se diseñó y construyó software para proceso automático de imágenes de muestras de fibras textiles de camélidos y ovinos (transferido a UCC). En la actualidad se está desarrollando hardware y software para ser aplicado en campo por los investigadores de SUPPRAD.

### Formación de Recursos Humanos

El equipo de investigación está compuesto por investigadores formados, en formación y becarios alumnos y graduados de las carreras de Ingeniería en Sistemas de Información y de Ingeniería Electrónica de la UTN-FRC.

Adicionalmente, han colaborado en los distintos subproyectos Demógrafos del IIGHI-CONICET, Veterinarios de la UCC, Médicos y Psicólogos de la Universidad Nacional de Córdoba (UNC) y técnicos del Laboratorio de Investigación de Software de la UTN-FRC.

Durante el proyecto, seis becarios lograron su título de grado en Ingeniería, dos iniciaron la Maestría en Ingeniería en Sistemas de Información y dos su Doctorado (en Ciencias de la Computación y en Ingeniería). Adicionalmente dos de los investigadores formados lograron su Magíster en Ingeniería de Software.

Además, desde el proyecto se lanzaron dos nuevos proyectos relacionados y se esperan dos más en el corto plazo.

## Referencias

- [1] Martínez, Vázquez, Marciszack. *Redes Neuronales vs. Autómatas Celulares*. VII WICC, Río IV, Argentina. 2005.
- [2] Vázquez, Castillo, Rojas, Marciszack. *Redes neuronales aplicadas a las ciencias sociales*. X WICC, General Picco, La Pampa, Argentina. 2008.
- [3] Vázquez, Castillo, Cárdenas, Rojas. *Modelo computacional empleando redes neuronales artificiales para la estimación del riesgo para la salud de la vivienda urbana*. XIII WICC, Santa Fe, Argentina. 2011.
- [4] INTA. *Programa Nacional Fibras Animales*. Documento Base actualizado a noviembre de 2011.
- [5] Adot, O. *Introducción a la Industrialización de la Lana y las Fibras Especiales*. SUPPRAD N°2. 2010.
- [6] Mueller, J. *Objetivos de Mejoramiento Genético para Rumiantes Menores*. INTA EEA Bariloche, Comunicación Técnica, PA 238. 1993.
- [7] Mueller, J. *Novedades en la determinación del diámetro de fibras de lana y su relevancia en programas de selección*. Comunicación Técnica. INTA, Bariloche, 330pp. 2002.
- [8] Frank, E. *Camélidos Sudamericanos. Producción de fibra, bases físicas y genéticas*. Revista Argentina de Producción Animal. Vol. 28, pp. 112-119. 2008.
- [9] Frank, E., Hick M., Prieto, A., Castillo, M., 2009. *Metodología de Identificación Cualitativa y Cuantitativa de Fibras Textiles Naturales*. SUPPRAD N° 1. 2009.
- [10] Huang, Y., Xu, B. *Image Analysis for Cotton Fibers. Part I: Longitudinal Measurements*. Textile Research Journal, 72(8), 713-720. 2002.
- [11] Xu, B., Ting, Y. *Fiber Image Analysis. Part I: Fiber Image Enhancement*. Textile Research Journal, 87, 274-283. 1996.
- [12] Xu, B., Ting, Y. *Fiber Image Analysis. Part II: Measurement of General Geometric Properties of Fibers*. Textile Research Journal, 87, 284-295. 1996.
- [13] Rojas Vigo, D. A. *Caracterización del Espesor de las Fibras de Alpaca Basada en Análisis Digital de Imágenes*. Electrónica - UNMSM, N° 17. 2006.
- [14] Huang, Y. y Xu, B. *Image Analysis for Cotton Fibers. Part II: Cross-Sectional Measurements*. Textile Research Journal, 74(5), 409-416. 2004.



# Tracking de múltiples objetos aplicado a insectos

Diego Marcovecchio<sup>†‡</sup>, Natalia Stefanazzi<sup>‡</sup>, Claudio Delrieux<sup>†</sup>, Ana Maguitman<sup>‡</sup>, y  
Adriana Ferrero<sup>‡</sup>

<sup>†</sup>Laboratorio de Ciencias de las Imágenes (LCI)  
Departamento de Ingeniería Eléctrica y de Computadoras (DIEC)

<sup>‡</sup>Grupo de Investigación en Administración de Conocimiento y Recuperación de Información - LIDIA  
Departamento de Ciencias e Ingeniería de la Computación (DCIC)

<sup>‡</sup>Laboratorio de Zoología de Invertebrados II  
Departamento de Biología, Bioquímica y Farmacia (DBBF)

Universidad Nacional del Sur (UNS)  
Av. Alem 1253, (B8000CBP), Bahía Blanca, Argentina  
Tel: (0291) 459-5135 / Fax: (0291) 459-5136

## Resumen

La detección y el tracking de objetos múltiples es un aspecto muy estudiado dentro de las ciencias de las imágenes. El tracking de objetos múltiples es, en general, un problema exigente debido a una amplia gama de potenciales problemas que pueden surgir (como el movimiento brusco de los objetos, el cambio en apariencia de los objetos o la escena de fondo, frames de mala calidad, oclusión entre objetos o entre el objeto y la escena, o movimiento de la cámara). Usualmente, el tracking es realizado en aplicaciones que requieren la locación de los objetos en cada frame para realizar algún posterior procesamiento.

Conociendo la complejidad del problema de tracking, normalmente se asumen determinadas condiciones que hacen que la resolución sea más sencilla en el contexto particular. En este trabajo, describimos los métodos y técnicas utilizados para realizar una aplicación que detecta y trackea múltiples objetos (en particular, insectos en placas de Petri en videos obtenidos en un proyecto de integración con el Laboratorio de Zoología de Invertebrados II, del Departamento de Biología, Bioquímica y Farmacia), y analiza distintos aspectos del recorrido efectuado por cada uno de ellos.

El objetivo general y a largo plazo del proyecto es desarrollar una aplicación que permita realizar identificación y tracking de objetos genéricos.

**Palabras clave:** Detección de objetos, Tracking de objetos, Procesamiento de video.

## Contexto

El presente proyecto se da en el marco de la colaboración conjunta desarrollada por el Laboratorio de Ciencias de las Imágenes (IIIE-CONICET <http://www.imaglabs.org>), el Grupo de Investigación en Administración de Conocimiento y Recuperación de Información (<http://ir.cs.uns.edu.ar>) y el Laboratorio de Zoología de Invertebrados II, todos pertenecientes a la Universidad Nacional del Sur.

Esta línea de investigación se lleva a cabo dentro del ámbito del LCI, y está asociada a los siguientes proyectos de investigación:

- Procesamiento inteligente de imágenes. PICT 24/K047 (SECyT-UNS). Director: Claudio Delrieux.
- PICT Start-Up 2442/2010. Director: Dr. Claudio Delrieux.

- Soporte inteligente para el acceso a información Contextualizada en entornos centralizados y distribuidos. PIP: 11220090100863. Director: Dra. Ana G. Maguitman. Fuente de financiamiento: CONICET. 2010-2012.
- Aceites esenciales y nanoinsecticidas: nuevas alternativas de productos biorracionales para el control de insectos plaga de importancia en la sanidad vegetal y humana. PGI 24/B187 (SECyT-UNS). Director: Dra. Adriana A. Ferrero.

## Introducción

Dada la creciente facilidad de acceso a los dispositivos con capacidad de generación de contenido multimedia, en los últimos años ha habido un gran crecimiento en la cantidad de información disponible. Esto motivó el desarrollo de aplicaciones orientadas a procesar dicha información.

En función del gran volumen de contenido, comenzaron a desarrollarse sistemas automáticos y semi-asistidos que permiten el etiquetado de acciones en video con diferentes aplicaciones, como por ejemplo la detección de acciones sospechosas en cámaras de vigilancia [2], detección de incidentes de tránsito [4] y búsqueda de acciones en videos deportivos [1].

Este trabajo aborda el reconocimiento de insectos y el tracking de éstos en secuencias de video, con el objetivo a largo plazo de generalizar tanto la segmentación como el tracking a objetos genéricos. Este tema ya ha sido desarrollado por algunos otros autores: Balch, Khan y Veloso desarrollaron un sistema de tracking de hormigas [5] en la Universidad de Carnegie Mellon. Sin embargo, el sistema es muy limitado y presenta problemas como la oclusión de las hormigas por las paredes de la Placa de Petri, la pérdida del tracking cuando las hormigas pierden la mínima separación necesaria, la separación de los rectángulos minimax delimitantes (*bounding box*) de las hormigas (haciendo que, ocasionalmente, el programa confunda uno de los insectos con muchos de ellos debido a reflejos especulares), y la pérdida del tracking cuando las hormigas dejan de moverse por un determinado tiempo.

El sistema desarrollado en nuestro laboratorio pretende eliminar o reducir drásticamente dichos

problemas, utilizando técnicas más avanzadas de procesamiento de video, extracción de característicos, tracking y mejoras en las heurísticas, además de asumir ciertas condiciones que no son necesariamente más restrictivas, pero sí ayudan a acotar el problema.

Existen diversas motivaciones para buscar obtener un sistema de tracking generalizado; por ejemplo, la posibilidad de utilizarlo para mejorar los sistemas de control de plagas (orientados a evaluar la actividad repelente de diferentes productos), u observar el comportamiento de ciertas poblaciones de insectos sin perturbar su comportamiento natural utilizando métodos no invasivos. De la misma manera, es posible pensar en potenciales mejoras computacionales (como sugerir nuevos algoritmos biónspirados tales como la optimización basada en colonias de hormigas).

## Líneas de investigación y desarrollo

La principal línea de investigación está centrada en la detección, extracción de característicos y tracking de objetos. Para ello, se diseñó y se continúa implementando una aplicación encargada de utilizar algunas de las técnicas ya conocidas de procesamiento de video para trackear insectos, con el objetivo de intentar mejorar los algoritmos de tracking ya conocidos.

Adicionalmente se realizan algunos análisis estadísticos sobre los datos obtenidos durante el procesamiento, que son posteriormente reportados al Laboratorio de Zoología de Invertebrados II con el que se realiza el proyecto en conjunto.

## Propuesta y metodología

Se diseñó y realizó la implementación (que continúa en un proceso de mejora constante) de una aplicación encargada de procesar videos, segmentar, trackear y realizar análisis estadísticos del conjunto de insectos presentado. Indicamos a continuación el funcionamiento de cada subsistema.

## Videos

Los videos fueron realizados para evaluar la actividad repelente del aceite esencial extraído de una planta nativa del norte argentino en cucarachas. Para ello, discos de papel de filtro de 18 cm de diámetro fueron divididos en dos mitades; una de ellas se roció con 1 mL del aceite en tanto que la otra no recibió tratamiento, y sobre el papel se liberaron insectos. Se utilizaron anillos plásticos de 10 cm de alto para evitar el escape de las cucarachas. Las filmaciones se registraron en un cuarto cerrado en condiciones controladas de humedad y temperatura durante 30 minutos.

## Segmentación

La aplicación desarrollada procesa el video, reconociendo inicialmente a las cucarachas utilizando un centroide de color característico y aplicando algoritmos de clustering [3] adaptados, de manera que cada insecto quede contenido en una *bounding box*. En adelante, cada insecto será identificado frame a frame por la *bounding box* que lo contiene, así como su vector de movimiento, y un historial que permite identificar el rastro dejado.

## Tracking

En cada frame se detecta el movimiento de los pixels con color característico dentro de cada *bounding box*, haciendo uso de técnicas de erosión y dilatación para reducir problemas de ruido en el video, y reajustando la posición de cada caja según el nuevo centroide de pixels positivos.

Los problemas de oclusión entre objetos mencionados en el sistema de Balch, Khan y Veloso se solucionan parcialmente reaplicando algoritmos de *clustering* y un modelo probabilístico; de la misma manera, como se asume una cantidad constante de objetos en el video, la separación de las *bounding box* se soluciona, pues en todo momento se conocen las respectivas posiciones y vectores de movimiento de cada insecto, y las *bounding box* se repositionan dinámicamente frame a frame, en lugar de crear una nueva caja por cada potencial insecto detectado. Por último, al utilizarse las *bounding box* y ajustar sus posiciones según el contraste de los pixels característicos respecto al fondo, si los insectos

dejan de moverse las cajas no corren riesgo de desaparecer.

## Análisis estadístico

Los análisis estadísticos realizados posteriormente al procesamiento consisten, en primera instancia, en detectar qué porcentaje de tiempo pasa cada insecto en las zonas tratadas y no tratadas con insecticidas, para poder estudiar la efectividad de éstos. Como se mantiene un historial de las posiciones y el recorrido de cada insecto, también es factible realizar otros análisis como la tortuosidad del recorrido.

## Resultados y Objetivos

### Resultados preliminares

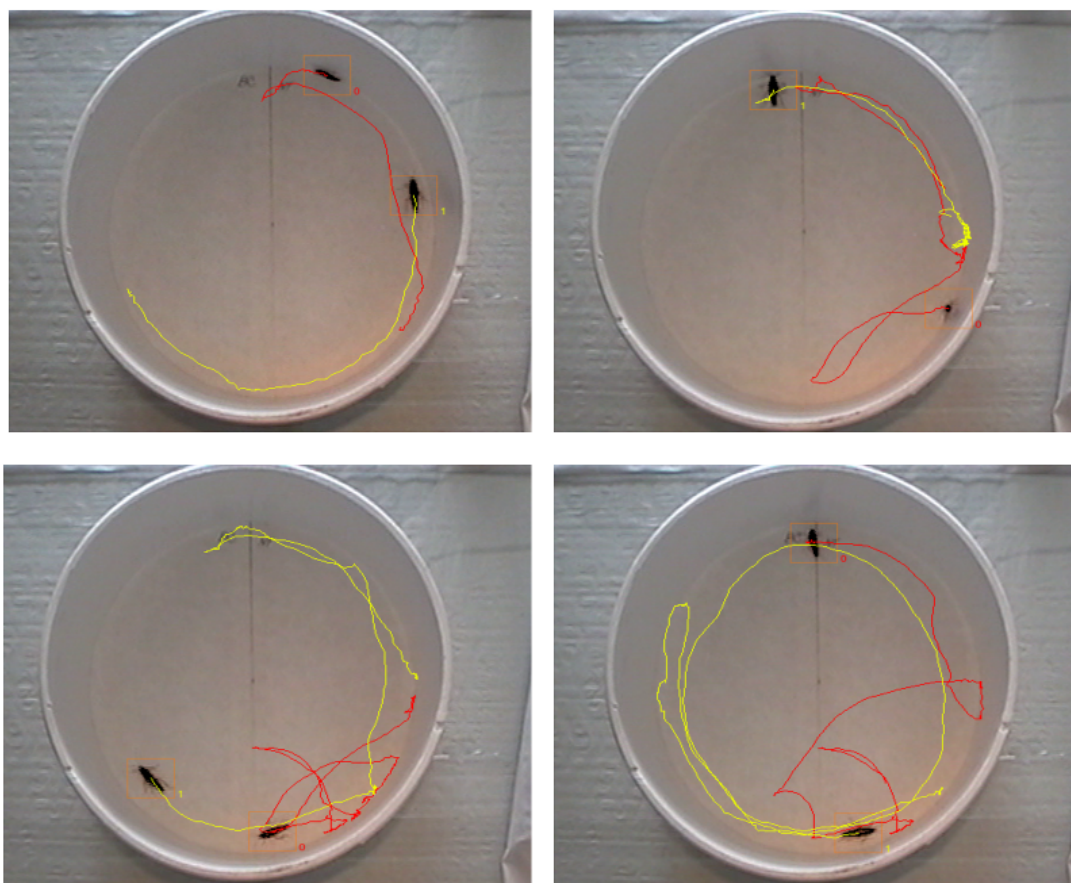
Actualmente, el sistema hace un tracking efectivo de los insectos en todas las condiciones normales presentadas en los videos, pudiendo generar estadísticas porcentuales respecto al tiempo de presencia de cada insecto en las regiones de interés con gran efectividad. La aplicación además demuestra robustez con cambios abruptos que se han presentado en la iluminación de la habitación en la que se filmaron los videos. Sin embargo, existen aún algunas limitaciones en la aplicación; entre ellas, la posibilidad de un tracking menos efectivo en caso de que dos o más insectos ocupen el mismo terreno durante mucho tiempo, pudiendo confundírseles, y fallos menores en las estadísticas debido a frames de mala calidad en el video.

En la figura 1 se puede observar a la aplicación en funcionamiento.

### Trabajo a futuro

Una de las principales características que se pretende agregar en el futuro es la capacidad de realizar la segmentación y el reconocimiento de los insectos sin utilizar un centroide de color característico; también es deseable el agregado de verificaciones de sensatez (*sanity checks*) entre cortos intervalos de tiempo, para comprobar que las *bounding box* efectivamente estén posicionadas en todo momento sobre un insecto, y en caso contrario, reanalizar el cuadro completo para repositionarlas; de esta manera, la aplicación resultaría aún más robusta, y

**Figura 1** La aplicación en ejecución. Se puede observar las *bounding box* de cada insecto, así como el rastro dejado por cada uno en diferentes momentos.





permitiría realizar los análisis estadísticos de manera aún menos asistida.

## Formacion de Recursos Humanos

Actualmente el equipo de trabajo de esta línea de investigación se encuentra formado por un becario de posgrado poseedor de una beca de ANPCyT, junto a su director y co-directora de tesis, más una estudiante de grado de biología y su directora de tesina.

Además, como parte de las actividades asociadas al proyecto se realizan cursos de grado y postgrado en Procesamiento Digital de Imágenes y Minería de datos.

## Referencias

- [1] Changsheng Xu y Qingming Huang Guangyu Zhu. Player action recognition in broadcast tennis video with applications to semantic analysis of sports game. In *in Proc. ACM Multimedia, 2006*, pages 431–440, 2006.
- [2] M. Fujii y S. Satoh M. Takahashi, M. Naemura. Human action recognition in crowded surveillance video sequences by using features taken from key-point trajectories. In *Computer Vision and Pattern Recognition Workshops (CV-PRW), 2011 IEEE Computer Society Conference on*, pages 9–16, June.
- [3] J. B. MacQueen. Some methods for classification and analysis of multivariate observations. In L. M. Le Cam and J. Neyman, editors, *Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 281–297. University of California Press, 1967.
- [4] Katsushi Ikeuchi y Masao Sakauchi Shunsuke Kamijo, Yasuyuki Matsushita. Incident detection at intersections utilizing hidden markov model, 1999.
- [5] Zia Khan y Manuela Veloso Tucker Balch. Automatically tracking and analyzing the behavior of live insect colonies, 2000.

# Líneas de Realidad Aumentada aplicadas a la Geología y los Libros Aumentados

Damián Flores<sup>(1,4)</sup>, Nicolás Gazcón<sup>(2,4)</sup>, Silvia Castro<sup>(4)</sup>

<sup>(1)</sup>Becario ANPCyT, <sup>(2)</sup> Becario Posgrado CONICET

<sup>(4)</sup>Laboratorio de Visualización y Computación Gráfica (VyGLab)

Dto. de Ciencias e Ingeniería de la Computación

Universidad Nacional del Sur

Av. L. N. Alem N° 1253, B8000CPB

Bahía Blanca, Buenos Aires, Argentina

Te.: (+54 291) 459-5135 Fax: (+54 291) 459-5136

{df, nfg, smc}@cs.uns.edu.ar

## Resumen

La manera en la que las personas interactúan con las computadoras no siempre es la misma. A medida que la tecnología avanza se desarrollan nuevas herramientas, nuevas propuestas de interacción emergen y se crean nuevas interfaces.

Las interfaces de Realidad Aumentada constituyen un ejemplo de tal desarrollo. Las mismas ofrecen al usuario un entorno combinando, por un lado, información del mundo real, y por otro, información sintética creada y manejada por la computadora. Una correcta fusión de estos dos mundos en un único entorno e interfaz de usuario es un elemento esencial en todo sistema de Realidad Aumentada. Las distintas áreas de aplicación de estos sistemas poseen cada una sus propios requerimientos.

En este artículo presentamos las diferentes líneas de investigación relacionadas con Realidad Aumentada que están actualmente en desarrollo en nuestro laboratorio.

**Palabras Clave:** Realidad Aumentada, Interacción Humano – Computadora, Computación Gráfica.

## Contexto

El trabajo se lleva a cabo en el Laboratorio de Investigación y Desarrollo en Visualización y Computación Gráfica (VyGLab) del Departamento de Ciencias e Ingeniería de la Computación de la Universidad Nacional del Sur.

La línea de Investigación presentada se encuadra en el proyecto “Representaciones Visuales e Interacciones para el Análisis Visual de Grandes Conjuntos de Datos” (24/N020), dirigido por la Dra. Silvia Castro. Este proyecto es financiado por la Secretaría General de Ciencia y Tecnología de la Universidad Nacional del Sur; y acreditados por la Universidad Nacional del Sur, Bahía Blanca.

## 1. Introducción

Los medios con los cuales las personas interactúan con las computadoras evolucionan rápidamente. Cada día surgen nuevas formas de comunicación con ésta, que se alejan de los esquemas de interacción tradicionales con teclado y mouse. Ejemplos de esto son las interfaces hápticas, las tangibles, las gestuales, las de reconocimiento de voz, las de realidad aumentada y virtual, e incluso las multimodales.

Cada nuevo tipo de interfaz desarrollada o propuesta trae consigo potencial para mejorar esta interacción en algunas áreas de aplicación. A su vez, también proponen desafíos y limitaciones que deben tenerse en cuenta a la hora de su utilización.

A continuación se introducen las interfaces de Realidad Aumentada y en la siguiente sección las líneas que están actualmente en desarrollo en el grupo de investigación.

### 1.1. Realidad Aumentada

En sus comienzos, la Realidad Aumentada (RA) surgió como una subárea de la Realidad Virtual (RV); pero con una diferencia muy importante: no sólo existen elementos virtuales generados por la computadora, sino que estos se combinan con elementos e información del mundo real.

Los sistemas de RA generan una visión que combina la escena real vista por el usuario y la virtual generada por la computadora integrándose ambas en un único entorno que se ha enriquecido con información adicional. Esta capacidad de combinar información real y virtual, la posibilidad de soportar interacciones en tiempo real y lograr una correcta registración tridimensional son los tres aspectos que caracterizan a estos sistemas [Azu97].

La información aumentada que recibe el usuario mejora el desempeño de la persona en su percepción del mundo y en la interacción con el mismo. Idealmente, el usuario debe interactuar naturalmente con los objetos virtuales y los reales, los cuales constituyen su mundo integrado. El objetivo ideal es contar con un sistema en el cual el usuario no sea capaz de discernir entre los aspectos del sistema que son reales y los que corresponden a la aumentación virtual.

### 1.2. Áreas de Aplicación

Las áreas de aplicación en las que estos sistemas podrían resultar efectivos se han ido ampliando a medida que distintos factores

tecnológicos limitantes van siendo superados. El mayor ejemplo de tal avance se puede observar en las aplicaciones que requieren movilidad. Los primeros sistemas móviles de RA requerían que el usuario incorporase varios elementos de *hardware*; esto los hacía incómodos e imprácticos de utilizar. Además, el costo de los mismos era generalmente prohibitivo. Actualmente se puede construir un sistema de RA móvil utilizando, por ejemplo, sólo un teléfono inteligente (*smartphone*), dispositivo al alcance del usuario promedio.

Tanto las aplicaciones tradicionales orientadas a las grandes industrias de manufactura, medicina, aviación, adiestramiento, diseño y construcción, entre otras, como las hogareñas orientadas a la educación, entretenimiento, marketing, etc., son ejemplos de áreas que han sabido sacar provecho de una manera cada vez más activa de la RA.

También debe mencionarse que las distintas áreas de aplicación imponen ciertos requisitos en cada uno de los subsistemas componentes de RA. Por ejemplo, los requerimientos en relación al *tracking* (subsistema utilizado para registrar la posición y orientación de los objetos) no son los mismos para una aplicación de entretenimiento hogareño [Min12] que para una aplicación del campo de la medicina [Mou09], en donde la vida del paciente puede ponerse en riesgo si no se cuenta con un *tracking* efectivo. La tecnología del *display* también se ve influenciada por el tipo de aplicación; en un sistema que no requiera movilidad se puede adaptar el entorno para utilizar *displays* espaciales, aspecto que no es viable en un sistema móvil que opere al aire libre, por ejemplo. La capacidad de cómputo, y en particular las capacidades gráficas, de los dispositivos móviles como *smartphones* o PDAs están acotadas debido al *hardware* utilizado. Esto condiciona las técnicas utilizadas en el proceso de *rendering*.

## 2. Líneas de Investigación y Desarrollo

A continuación describimos de manera sintética las dos líneas de investigación en el área de Realidad Aumentada.

### 2.1. Libros Aumentados

Acciones como sostener un libro, voltear sus páginas, transportarlo o sentir su peso, nos resulta placentero porque es a lo que estamos acostumbrados. Aumentarlos digitalmente no elimina estas ventajas; lejos de ello, nos abre un abanico de nuevas posibilidades, ya que podemos enriquecer los libros reales al combinar sus beneficios físicos con la interacción que nos ofrecen los medios digitales. Gracias a la versatilidad que ofrece la RA, la interacción con los contenidos digitales permitirá al usuario desenvolverse de manera más activa en su lectura [Gra08], adicionalmente permitiendo actividades en conjunto [Ha11].

La motivación de esta línea es involucrarnos en los distintos aspectos de diseño de los libros aumentados, es decir, aquellos libros que integren una combinación de contenido virtual y físico. El objetivo general consistirá en la exploración de los distintos aspectos que conducen al diseño de libros aumentados, que involucran desde su desarrollo hasta la experiencia del usuario con este tipo de medios. Podemos destacar que esta tecnología puede usarse tanto en el contexto educativo como también en contextos tales como entretenimiento, ingeniería, turismo y visualización de datos entre otros.

El fin es obtener libros que nos permitan dar soporte a la metáfora de la lectura tradicional y a su vez se vean enriquecidos con los elementos digitales que la RA nos permite incorporar de manera natural, no sólo aumentando sus contenidos sino también elementos resultantes de las interacciones. El desarrollo de esta subárea de la RA aportará seguramente al desarrollo de nue-

vas técnicas básicas de RA en los campos de tracking, interacciones con distintos dispositivos, etc. en el campo emergente de aplicaciones ubicuas de RA; además, los resultados obtenidos redundarán en beneficios para diversos dominios de aplicación.

### 2.2. Realidad Aumentada Móvil

Los sistemas de RA móviles poseen requerimientos que limitan los elementos disponibles para su construcción. Como se mencionó anteriormente, las capacidades gráficas y de cómputo de los elementos utilizados (*PDA*, *smartphone*) generalmente son limitadas. Sin embargo, en la actualidad existe un mercado cada vez mayor de equipos portátiles con capacidades cada vez más similares a las de una PC de escritorio, con una gran movilidad, portabilidad y con un costo moderado: las *netbooks* y/o *tablet-PCs*.

Estos nuevos dispositivos permiten trasladar hacia las aplicaciones móviles todos los procedimientos, métodos y técnicas utilizados en las aplicaciones de escritorio en lo referido al poder de cómputo y gráfico, elementos que no podían ser portados a los dispositivos anteriores debido a sus capacidades limitadas.

Debido a la naturaleza móvil, el subsistema de *tracking* debe ser tal que no requiera de configuraciones especiales en el entorno ni elementos de dimensiones considerables. En este sentido, los sistemas inerciales tienen la capacidad de cumplir con estos requisitos debido a su principio de funcionamiento y a su gran escala de integración. Además, los sistemas de navegación basados en ondas de radio o microondas, (GPS, Glonass, LORAN, TACAN) permiten rastrear al usuario en exteriores y prácticamente en cualquier parte del mundo. Estos dos tipos de tracking pueden combinarse y ofrecer una solución satisfactoria en este tipo de sistemas.



En esta línea se integran éstos y otros elementos para conformar un sistema de RA móvil con características similares a uno de escritorio.

### 3. Resultados y Objetivos

Exponemos a continuación los resultados obtenidos y los objetivos en curso y a futuro de las líneas de presentadas.

#### 3.1. Libros Aumentados

En este trabajo se plantea como objetivo encontrar un modelo para los libros aumentados, considerando además su diseño y desarrollo conducente a la implementación de un prototipo. Ahondaremos en los distintos aspectos de los elementos de diseño de los libros aumentados considerando especialmente diferentes técnicas de interacción que resulten en libros de gran riqueza sin perder la idea de lectura tradicional y de las tareas que la complementan.

Actualmente se está desarrollando la plataforma necesaria para abordar una propuesta con soporte colaborativo. Para esto es necesario contemplar las complicaciones de una alternativa distribuida, sumado al desarrollo que implica una aplicación de Realidad Aumentada.

Para esto también es necesario examinar los diferentes tipos de *registro/tracking* que se adaptan a los libros (con marcadores o los denominados *markerless*) como también los diferentes contenidos digitales que se utilicen.

#### 3.2. Realidad Aumentada Móvil

En esta línea se está trabajando en el diseño de un sistema de RA móvil orientado a la visualización *in situ* de datos del campo de la geología. El mismo integra la vista del mundo real, recogida a través de una cámara de vídeo, con información geológica geo-referenciada. Ejemplos de integración de sistemas GIS con sistemas de RA se presentan en los trabajos de [King05] y [Aba12].

Actualmente se encuentra desarrollado un primer prototipo, que puso de manifiesto los principales aspectos a tener en cuenta para que el sistema sea satisfactorio. En las pruebas realizadas, el problema de la registración (entre las imágenes del mundo real y los modelos de objetos y el terreno), fue el más evidente. Esto motiva a profundizar distintas técnicas de registro y modelos de representación de la información geográfica, temas que están siendo abordados en la actualidad.

La información a visualizar presenta una dependencia de las características del terreno y los principales accidentes geográficos, por lo que un correcto almacenamiento, representación e interacción con información del terreno resulta fundamental en este contexto.

La ubicación y condiciones geográficas del lugar de utilización de esta aplicación son tales que no es posible contar con acceso a redes de comunicación tradicionales como Internet o telefonía celular. Esto hace que toda la información del mundo virtual deba almacenarse en el propio dispositivo portátil. Sin embargo, se cuenta con acceso al servicio público de Posicionamiento Global (GPS), servicio que permite conocer la ubicación del usuario con una cierta precisión, la cual puede mejorarse utilizando distintas propuestas algorítmicas [Aco12]. Para completar esta información, se pretende contar con dispositivos de *tracking* inercial, que debido a su característica de no requerir ninguna configuración especial externa, se adaptan a las condiciones requeridas por la aplicación. Estos sistemas inerciales acumulan errores a lo largo del tiempo, hecho que obliga a estudiar y evaluar las distintas técnicas necesarias para mejorar la precisión de los mismos.

### 4. Formación de RRHH

La estructura del equipo de trabajo en las líneas presentadas está conformada por la directora del grupo, la Dra. Silvia Castro y

por los becarios/tesistas de posgrado Damián Flores y Nicolás Gazcón.

Se detallan a continuación las tesis en desarrollo y los cursos relacionados con la línea de investigación:

#### 4.1. Tesis de Doctorado y becarios en Cs. de la Computación

- Nicolás Gazcón. *La Exploración en los Libros Aumentados: Desafíos de las Interacciones*. Dir.: Dra. Silvia Castro.
- Damián Flores. *Realidad Aumentada en Visualización*. Dir.: Dra. Silvia Castro – Dr. Ernesto Bjerg.

#### 4.2. Trabajos de Final de Carrera

- Manuel Francisco Soto. *Realidad Aumentada Aplicada al Transporte Público Utilizando SmartPhones* Dir.: Dra. Silvia Castro y Dr. Martín Larrea. 2012.
- Daniel Eloy Sacomani. *Realidad Aumentada Aplicada al Desarrollo de Juegos*. Dir.: Dra. Silvia Castro y Dr. Martín Larrea. 2013.

#### 4.3. Cursos de pregrado

En la UNS, se dictaron cursos de Computación Gráfica, HCI, Procesamiento de Imágenes y Sistemas Embebidos.

## 5. Referencias

- [Aba12] F. Ababsa, I. Zedjebil, J. Didier, J. Pounderoux, and J. Vairon. Outdoor Augmented Reality System for Geological Applications. The 2012 IEEE/ASME International Conference on Advanced Intelligent Mechatronics. July 11-14, 2012, Kaohsiung, Taiwan.
- [Aco12] Nelson Acosta and Juan Toloza. Techniques to improve the GPS precision. International Journal of Advanced Computer Science and Applications. Vol. 3, No. 8, 2012. pp 125-130.
- [Azu97] R. Azuma. A Survey of Augmented Reality, Presence: Teleoperators and Virtual Environments, vol. 6, no. 4, Agosto 1997, pp. 355-385.
- [Gra08] R. Grasset, A. Dünser and M. Billinghurst. Edutainment with a mixed reality book: a visually augmented illustrative childrens' book. In Proceedings of the 2008 International Conference on Advances in Computer Entertainment Technology.
- [Ha11] Ha T., Lee Y., Woo W., Digilog book for temple bell tolling experience based on interactive augmented reality. Virtual Reality, 15:295-309, 2011.
- [Kin05] Gary King, Wayne Piekarski and Bruce Thomas. ARVino – Outdoor Augmented Reality Visualisation of Viticulture GIS data. Proceedings of the 4<sup>th</sup> IEEE/ACM International Symposium on Mixed and Augmented Reality. 2005. pp 52 – 55.
- [Min12] Mine M.R.; van Baar J.; Grundhofer, A.; Rose D.; Yang B, "Projection-Based Augmented Reality in Disney Theme Parks," Computer, vol. 45, no. 7, pp. 32-40, July 2012.
- [Mou09] Mountney P., Giannarou S., Elson D., Yang G., Optical biopsy mapping for minimally invasive cancer screening. In Proceedings of the 12th International Conference on Medical Image Computing and Computer-Assisted Intervention: Part I, 2009.

# Visión por Computador e Informática Gráfica. Realidad Virtual, Realidad Aumentada e Interfaces avanzadas

María José Abásolo<sup>1,2</sup>, Alejandro Mitaritonna<sup>3</sup>, Javier Giacomantone<sup>1</sup>, Armando De Giusti<sup>1</sup>,  
Marcelo Naiouf<sup>1</sup>, Cristina Manresa<sup>4</sup>, Francisco Perales<sup>4</sup>, Ramon Mas Sansó<sup>4</sup>, Silvia Mabel  
Castro<sup>5</sup>,

<sup>1</sup>Instituto de Investigación en Informática LIDI (III-LIDI)  
Facultad de Informática – Universidad Nacional de La Plata (UNLP)  
calle 50 y 120 (1900) La Plata, Buenos Aires, Argentina  
{mjabasolo, jog, degiusti, mnaiouf}@lidi.info.unlp.edu.ar

<sup>2</sup>Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CICPBA)

<sup>3</sup> Instituto de Investigaciones Científicas y Técnicas para la Defensa (CITEDEF)  
San Juan Bautista de La Salle 4397 (B1603ALO) Villa Martelli, Buenos Aires, Argentina  
amitaritonna@citedef.gob.ar

<sup>4</sup>Unidad de Gráficos y Visión por Ordenador(UGiV)  
Departamento de Ciencias Matemáticas e Informática – Universidad de las Islas Baleares (UIB)  
Edificio Anselm Turmeda. Ctra. Valldemossa km 7,5 (07122) Palma, Baleares, España  
{cristina.manresa, paco.perales, ramon.mas}@uib.es

<sup>5</sup>Laboratorio de Visualización y Computación Gráfica (VyGLab)  
Dpto. de Ciencias e Ingeniería de la Computación (DCIC) – Universidad Nacional del Sur (UNS)  
Avda Alem 1253 (8000) Bahía Blanca, Buenos Aires, Argentina  
smc@cs.uns.edu.ar

## Resumen

La línea de investigación y desarrollo presentada consiste en estudiar, desarrollar y evaluar métodos y algoritmos de sistemas de visión por computador así como también de informática gráfica. Los principales temas abordados son reconstrucción 3D, realidad virtual, realidad aumentada e interfaces basadas en visión. Uno de los principales objetivos abordados es el fortalecimiento de la investigación mediante el trabajo intergrupar entre diferentes instituciones, tanto nacionales (Universidad Nacional de La Plata, la Universidad Nacional del Sur y la Universidad Nacional del Centro de la Provincia de Buenos Aires), como extranjeras (Universidad de las Islas Baleares). Para posibilitar la formación de recursos humanos se realizó la definición y

puesta en marcha de una carrera de posgrado enfocada a Tratamiento de Señales e Imágenes, Visión por Computador, Informática Gráfica, Realidad Virtual y Aumentada, con la participación de docentes-investigadores de las instituciones mencionadas, acreditada por el Consejo Superior de la UNLP y actualmente en vías de acreditación por la CONEAU.

**Palabras Clave:** Visión por computador, Informática gráfica, Realidad Virtual, Realidad Aumentada, Interfaces basadas en visión

## Contexto

Esta línea de investigación y desarrollo (I/D) forma parte del Subproyecto “Tratamiento de imágenes digitales y

video. Visión 3D”, dentro del Proyecto “Procesamiento Paralelo y Distribuido y Aplicaciones en Sistemas Inteligentes y Tratamiento de Imágenes y Video” del Instituto de Investigación en Informática III-LIDI, acreditado por la Universidad Nacional de La Plata (UNLP) en el marco del Programa de Incentivos.

Esta línea de investigación es parte del proyecto “A1/037910/11 Formación de Recursos Humanos e Investigación en el Área de Visión por Computador e Informática Gráfica (FRIVIG)”, coordinado por la Universidad de las Islas Baleares (UIB) y la Facultad de Informática de la UNLP, financiado por la Agencia Española de Cooperación Internacional y Desarrollo (AECID) dentro del Programa de Cooperación Interuniversitaria e Investigación Científica. Dentro de dicho proyecto también participan el Laboratorio de Visualización y Computación Gráfica (VyGLab) de la Universidad Nacional del Sur (UNS) y el Instituto Pladema de la Universidad Nacional del Centro de la Provincia de Buenos Aires (UNICEN).

También dentro de esta línea se está desarrollando una tesis doctoral integrada al Programa de Investigación y Desarrollo para la Defensa (PIDDEF) 2012-2014 del Instituto de Investigaciones Científicas y Técnicas para la Defensa (CITEDEF).

## Introducción

La línea de investigación y desarrollo presentada involucra las áreas de visión por computador e informática gráfica. Los temas de investigación estudiados se agrupan en cuatro ejes temáticos: la reconstrucción 3D, realidad virtual, realidad aumentada e interfaces avanzadas.

## Reconstrucción 3D

La reconstrucción 3D consiste en la recuperación del modelo 3D a partir de una o más imágenes del modelo real. En la reconstrucción 3D se emplean diferentes

modelos y métodos. En particular se estudiaron los procesos de reconstrucción 3D basada en de visión estereocópica [1][2][3].

Más recientemente se comenzó a estudiar la reconstrucción 3D a partir de imágenes adquiridas con cámaras de tiempo de vuelo -en inglés Time of Flight (TOF). Las cámaras TOF obtienen sus píxeles a partir de determinar la distancia de la cámara al punto correspondiente en el objeto basándose en el tiempo de retardo. Dichas cámaras son capaces de crear mapas de distancias en tiempo real. Dichos mapas representan una reconstrucción 2-1/2 D ya que se trata de un mapa 2D de distancias o profundidades. Sin embargo el modelo 3D puede obtenerse a partir de combinar diferentes imágenes obtenidas de diferentes puntos de vista [4]. Las cámaras TOF también pueden usarse en otro de los temas de nuestro interés como es la detección y clasificación de gestos [5].

## Realidad Virtual

La Realidad Virtual (RV) es un término que se aplica a un conjunto de experiencias sensoriales sintéticas, es decir generadas por computador, comunicadas a un operador o participante. La mayoría de las aplicaciones de realidad virtual son experiencias visuales donde el participante se ve inmerso e interactúa en un ambiente o escena virtual 3D [6][7].

Una de las principales aplicaciones de RV son los simuladores. Al respecto se está trabajando en la implementación de un simulador de barco y un simulador de periscopio. Como antecedente de estos desarrollos se aprovecha la experiencia del Instituto Pladema el cual lleva años de desarrollo en este tipo de aplicaciones [7].

El simulador de barco cuenta con 3 pantallas de 42” LED 3D dispuestas una a continuación de la otra de forma de permitir una visión panorámica de 180° de un mundo virtual. Por otra parte, el periscopio contará con unas gafas de realidad virtual que permiten una visión



panorámica 360° del mundo virtual. Dichos desarrollos servirán de casos de estudio para alumnos de posgrado interesados en el campo. Asimismo se integrarán en una feria de ciencia abierta a alumnado de nivel primario y secundario, para explorar las posibilidades educativas de este tipo de desarrollos tecnológicos.

Existen dos temas en particular en los que se está trabajando como son la visualización de terrenos en tiempo real y la captura y visualización de video panorámico.

### **Realidad Aumentada**

Según la definición de R. Azuma [8] [9] la Realidad Aumentada (RA) se refiere a aplicaciones interactivas en tiempo real donde se visualiza la realidad con elementos sintéticos agregados (objetos 3D, sonidos, texto, etc.). En [7] puede encontrarse una introducción a esta área presentada por el grupo de investigación.

La combinación se realiza de forma coherente al punto de vista del usuario, de manera que los objetos 3D se encuentran registrados en el mundo real. Por esto es necesario realizar el seguimiento - en inglés *tracking* - del usuario para conocer su posición en el mundo en todo momento. A diferencia de las aplicaciones de RV las aplicaciones de RA generalmente necesitan la movilidad del usuario, incluso hacia ambientes externos- en inglés se denominan aplicaciones *outdoor*- para lo cual puede ser necesaria conocer la posición global del participante utilizando dispositivos como GPS y brújulas digitales, lo cual permite el acceso directo a información geo-referenciada. Las aplicaciones de RA son cada vez más populares debido a los notables progresos en los dispositivos de computación móviles, como celulares inteligentes, asistentes digitales personales y computadoras portátiles ligeras, los cuales tienen acceso a internet y dispositivos útiles para conocer la posición y orientación de usuario tales como

giroscopios, acelerómetros, al igual que GPS y brújulas. Algunas aplicaciones de RA realizan el análisis del video de la escena real capturada en tiempo real para el tracking del usuario. Las primeras aplicaciones de RA que realizan tracking basado en análisis de video utilizan marcadores para que el dispositivo tenga un punto de referencia sobre el cual superponer las imágenes. Recién en los últimos años el desarrollo de RA sin marcadores (en inglés *markerless*) está madurando ayudado por el aumento de capacidad computacional de los nuevos dispositivos, pudiendo el sistema de seguimiento captar un objeto real y superponer información sobre él [10].

La RA tiene aplicación en diversas áreas entre las que se enumeran medicina, soporte para realización de tareas complejas, asistencia en la navegación, publicidad, juegos, educación, etc.

En lo que respecta al ámbito militar, recientemente se está utilizando la RA como soporte para mejorar la conciencia situacional en la toma de decisiones en particular en operaciones militares. Muchas de las operaciones militares se desarrollan en entornos desconocidos. Estos complejos campos de batalla en 3D son muy exigentes e introducen muchos desafíos para el combatiente. Estos incluyen visibilidad limitada, falta de familiaridad con el medio ambiente, amenazas de francotirador, ocultamiento de fuerzas enemigas, mala comunicación y un problema de la localización e identificación de los enemigos y de las fuerzas aliadas. Para ello tener una conciencia situacional amplia del terreno es vital para que la operación sea un éxito minimizando los efectos colaterales. Una serie de programas de investigación han explorado los medios por los que la navegación y la coordinación de la información se pueden entregar a los soldados. El proyecto RAIOM (Realidad Aumentada para la Identificación de Objetivos Militares) que se está llevando a cabo como tesis doctoral consiste en el

desarrollo de un framework de RA para el reconocimiento, detección, identificación y suministro de información de objetos tridimensionales. El objetivo es que el prototipo sea desarrollado con software libre y ejecutado en un dispositivo móvil con sistema operativo de la misma característica (por ejemplo, *Android*). Consta de un sistema de sensores y pantallas que recogen (UAV, sensores externos, satélite, etc.) y proporcionan datos a cada soldado en el campo. Los dispositivos para la toma y representación de la información pueden ser gafas de RA translúcidas, tablet PC, notebooks, etc

### **Interfaces basadas en Visión**

Los estudios de investigación para obtener nuevos sistemas de interacción basados en fuentes de percepción de información como el sonido, el tacto o la visión se han convertido en un campo en auge que pretende desarrollar interfaces más naturales, intuitivas, no invasivas y eficientes. Hoy en día, las cámaras de bajo coste y con suficiente resolución se encuentran en la mayoría de las computadoras y en celulares. La información proveniente de la cámara se puede procesar mediante técnicas de visión por computador con fines de asistir la interacción persona-ordenador obteniendo las denominadas Interfaces basadas en visión (VBI) [7]. Según Turk [11] dichas interfaces pueden ofrecer diferentes funcionalidades como presencia y localización, identidad, expresión, gestos, foco de atención, postura del cuerpo y movimiento y actividad .

En el campo de la interacción persona-ordenador se utilizan pantallas táctiles (touchscreen), las cuales están superpuestas a un dispositivo de visualización, para la entrada táctil, el panel táctil. En algunos casos, las pantallas táctiles permiten la detección simultánea de varios puntos de contacto (multitáctiles). Esto ha hecho posible el desarrollo de metodologías de interacción

que permiten al usuario una comunicación más natural e intuitiva con el ordenador. Los habituales gestos utilizados para seleccionar, arrastrar, rotar o escalar objetos visuales han evolucionado a versiones en las que el usuario utiliza ambas manos o varios dedos para realizar las mismas acciones, proveyendo de mecanismos que se han convertido en gestos estándares.

Los displays de grandes dimensiones también se han hecho más y más necesarios por el auge del trabajo cooperativo en un mismo entorno. Han surgido nuevos términos para describir estos dispositivos: superficies (surfaces), mesas multitáctiles (multi-touch tables), *tabletops*. La utilización de cámaras y el posterior tratamiento de las imágenes mediante algoritmos de visión por computador se emplea en la construcción de mesas multitáctiles de bajo coste. En un contexto educativo, las mesas multitáctiles permiten a los estudiantes interactuar con objetos digitales en tareas colaborativas. En el caso de la educación especial, la tecnología puede proporcionar a los usuarios oportunidades de aprender, compartir información y ganar independencia. Se dedican esfuerzos para adaptar y desarrollar aplicaciones a las superficies multitáctiles para usuarios con limitaciones cognitivas y dificultades sociales.

En [12] se describe una aproximación realizada por nuestro grupo para la construcción de una mesa multitáctil de bajo coste pensando sobre todo en la seguridad de uso y en la accesibilidad y confort para usuarios discapacitados.

### **Líneas de investigación y desarrollo**

- Tratamiento de datos obtenidos con cámaras de tiempo de vuelo
- Reconstrucción 3D

- Simuladores de entrenamiento de Realidad Virtual
- Visualización de terrenos en tiempo real
- Captura y visualización de videos panorámicos en entornos inmersivos
- Aplicaciones de Realidad Aumentada en educación
- Aplicaciones de Realidad Aumentada militares
- Libros Aumentados
- Interfaces Basadas en Visión de bajo coste
  - 2 computadoras con tarjetas gráficas Geforce GTX 690
  - 3 TV LED 42" con soporte HDMI1.4 y 120 Hz
  - 1 kit Nvidia con gafas 3D
  - 3 tablet PC

## Resultados y objetivos

En el año 2012, las Universidades que participan del proyecto "FRIVIG: Formación de Recursos Humanos e Investigación en el área de Visión por Computadora e Informática" de la AECID pusieron en marcha una Especialización en Computación Gráfica, Imágenes y Visión por Computadora como posgrado de la UNLP, con la participación de docentes-investigadores de UNLP, UNS, UNICEN y UIB, con experiencia y formación en los temas propuestos. Dicha carrera ha sido aprobada por el Consejo Superior de la UNLP y está en vías de acreditación por la CONEAU.

- Durante los últimos 2 años se equipó al laboratorio III-LIDI de la UNLP con equipamiento específico para visión por computador e informática gráfica, tal como:
  - 1 cámara de tiempo de vuelo MESA SR4000
  - 1 gafas de realidad virtual con tracking Vusix Wrap 920 VR Bundle
  - 1 gafas de realidad aumentada "video see-through" Vuzix Wrap 920AR
- Se desarrolló un sistema, hardware y software, que permite la reconstrucción de modelos 3D basado en la combinación de cámaras estereoscópicas y la utilización de luz estructurada [1][2][3].
- Actualmente se están realizando pruebas con equipamiento de alto coste recientemente adquirido como la cámara de tiempo de vuelo MESA SR4000. Se espera en un corto a mediano plazo poder realizar reconstrucciones 3D.
- Se realizó la construcción de una mesa multitáctil basada en visión por computador para su uso en educación especial [12].
- Se están realizando pruebas de aplicaciones de realidad aumentada en exteriores en dispositivos móviles tales como tablets PC y smartphones [13].
- Se está realizando la construcción de un simulador de barco utilizando realidad virtual, en conjunto con el instituto Pladema de la UNICEN, con fines educativos y de difusión a la comunidad.
- Se está desarrollando un framework de software por parte de un doctorando del CITEDEF, para ser ejecutado en dispositivos móviles (tablets, notebooks y gafas) basado en realidad aumentada, visión por computador y sensores externos para el reconocimiento, detección, ubicación, identificación y suministro de información contextual.
- Se está desarrollando un framework de software por parte de un doctorando del VyGLab, para la generación de libros aumentados.

## Formación de recursos humanos

La formación de recursos humanos es prioritaria en esta línea, y por esto se implementó la Especialización en Computación Gráfica, Imágenes y Visión por Computadora.

Además de los cursos de posgrado presenciales, semipresenciales y a distancia que se realizan por parte del profesorado de las Universidades argentinas y españolas, también se han impartido cursos en las Escuelas Informáticas: “Realidad Virtual y Realidad Aumentada. Interfaces basadas en visión” [7] en la XV Escuela Internacional de Informática dentro del XVII Congreso Argentino de Ciencias de la Computación (CACIC 2011) y el curso “Interfaces Gestuales” en la XVI Escuela Internacional de Informática dentro del XVIII Congreso Argentino de Ciencias de la Computación (CACIC 2012).

En el marco de esta línea de investigación hay en curso:

- 4 tesis de doctorados en RA aplicadas respectivamente a educación, al ámbito militar, a la generación de libros aumentados, a geología en ambientes exteriores;
- 1 tesina de especialización en RA aplicada a educación;
- 3 tesinas de grado, una de visualización de terrenos, otra de libros aumentados con RA y otra de juegos con RA.
- 1 becario AECI trabajando en el desarrollo del simulador de barco

## Referencias

[1] F. Cristina, S. Dapoto, J. Vegas, V. Artola, C. Russo, C., M.J. Abásolo (2007) *Desarrollo de un escáner 3D mediante cámaras estereoscópicas e iluminación láser*. CACIC 2007.

[2] F. Cristina, S. Dapoto, J. Vegas, V. Artola, C. Russo, M.J. Abásolo, A. De Giusti. (2008) *3D Scanner Development with Stereoscopic Cameras and Laser Illumination*. IADIS Computer Graphics and Visualization CGV 2008. ISBN: 978-972-8924-63-8, p. 249-253.

[3] F. Cristina, S. Dapoto, J. Vegas, V. Artola, C. Russo, M.J. Abásolo (2008) *Reconstrucción 3D: de la adquisición a la visualización*. CACIC 2008.

[4] C. Munkelt, M. Trummer, P. Kühmstedt, G. Notni, J. Denzler (2009) *View Planning for 3D Reconstruction Using Time-of-Flight Camera Data*. Lecture Notes in Computer Science Volume 5748, pp 352-361

[5] A. Kollors, J. Penne, J. Hornegger, A. Barker A. (2008) *Gesture Recognition with a Time-of-Flight camera*. International Journal of Intelligent Systems Technologies and Applications, v. pp.334-343.

[6] Craig, A.; Sherman, W. and Will, J. (2009) *Developing Virtual Reality Applications: Foundations of Effective Design*, Morgan Kaufmann, ISBN 978-0-12-374943-7

[7] C. Manresa-Yee, M.J. Abásolo, R. Mas Sansó, M. Vénere. (2011) *Realidad Virtual, Realidad Aumentada e Interfaces Basadas en Visión*. XV Escuela Internacional de Informática, XVII Congreso Argentino de Ciencia de la Computación CACIC 2011. Editorial EDULP, ISBN 978-950-34-0765-3

[8] Azuma, R. (1997) *A Survey of Augmented Reality*. Presence: Teleoperators and Virtual Environments, 6(4), pp 355-385.

[9] Azuma, R.; Baillot, Y.; Behringer, R.; Feiner, S.; Julier, S. and MacIntyre, B. (2001) *Recent Advances in Augmented Reality*. IEEE Computer Graphics and Applications, 21(6).

[10] Haller, M; Billinghamurst, M. and Thomas, B. (2007) *Emerging Technologies of Augmented Reality. Interfaces and Design*, 1-22, Idea Group Publishing, ISBN1-59904-067-0.

[11] M. Turk, M. Kölsch (2005). *Perceptual interfaces*. En: G. Medioni and S.B Kang, editors, *Emerging Topics in Computer Vision*. Pearson Education

[12] C. Manresa Yee, R. Mas Sansó, G. Moyà, M. J. Abásolo, J. Giacomantone (2011) *Interactive multi-sensory environment to control stereotypy behaviours*. Computer Science & Technology Series XVII Argentine Congress of Computer Science Selected Papers. EDULP, Argentina 2012. Pág. 121-128. ISBN 978-950-34-0885-8.

[13] A. Rizzo, D. Flores, S.M. Castro, C. Manresa-Yee, S. Baldassarri. (2011) *AugNav: A location-based augmented reality system for outdoors navigation with smartphones*. CACIC 2011. Editorial EDULP, ISBN 978-950-34-0765-3



# IMÁGENES SATELITÁRIAS PARA DETECCIÓN Y MONITOREO OPERACIONAL DE LOS CAMBIOS EN EL BOSQUE NATIVO DEL NORTE DE LA PROVINCIA DE ENTRE RÍOS.

Francisco Darío Maldonado<sup>1</sup>, Walter Fabian Sione<sup>1</sup>, Fernando Raul Tentor<sup>1</sup>

<sup>1</sup>CEREGEO-Centro Regional de Geomática, Facultad de Ciencia y Tecnología. UADER-Universidad Autónoma de Entre Ríos.

Matteri y España s/n. (3105). Diamante. Entre Ríos. Argentina.

E-mail: {francisco.dario.maldonado; wsione; fernandotentor}@gmail.com

## Resumen

Este trabajo presenta el Proyecto trianual actualmente en ejecución en la Facultad de Ciencia y Tecnología de UADER-Universidad Autónoma de Entre Ríos. En este proyecto se desarrolla un sistema operacional para el monitoreo de cambios interanuales en la cobertura vegetal del bosque nativo del centro-norte de Entre Ríos utilizando imágenes digitales obtenidas en el espectro óptico por satélites orbitales. Para ello se adaptan técnicas digitales de procesamiento de imágenes para la detección de cambios en la vegetación en los ambientes sub-húmedos del norte provincial. La metodología posibilitará el monitoreo operacional a partir de imágenes satelitales de diferentes sensores. Las técnicas articuladas por esta metodología permiten la detección de cambios entre imágenes de sensores espectralmente compatibles, con o sin calibración atmosférica. Estas además son resistentes a la confusión producida por la dinámica compleja de la cobertura vegetal del área y capaces de incorporar las nuevas tecnologías satelitales actualmente en desarrollo. Los resultados y metodologías desarrolladas en el marco del proyecto permitirán establecer para el futuro un sistema operacional de monitoreo de la cobertura de bosques en la región permitiendo a la Provincia contar con herramientas para monitoreo del

cumplimiento de la Ley de Presupuestos Mínimos de Protección Ambiental de los Bosques Nativos de la Ley Nacional de Bosques 26331.

**Palabras clave:** Imágenes, técnicas digitales, teledetección, monitoreo, Landsat.

## Introducción

En este trabajo se presentan los lineamientos principales del proyecto trianual: “Desarrollo de un Sistema operacional para detección de cambios y monitoreo del bosque nativo usando imágenes satelitarias. Norte de la Provincia de Entre Ríos”. Este está en ejecución desde Marzo de 2013 en la Facultad de Ciencia y Tecnología de la Universidad Autónoma de Entre Ríos -FCyT/UADER. Uno de los motivos de la puesta en marcha del proyecto es el de la necesidad que se observa en los organismos de control y gestión de contar con herramientas de monitoreo para el control del cumplimiento de la Ley de Presupuestos Mínimos para el Bosque Nativo (Ley N° 26331) y para el ordenamiento territorial.

El proyecto utiliza imágenes digitales obtenidas periódicamente en el espectro óptico por satélites orbitales de estudio de los recursos naturales de la serie Landsat y LDCM-Landsat Data Continuity Mission. Estas imágenes son adecuadas para el monitoreo del ambiente a escala regional por su bajo costo y características espectro-radiométricas que

permiten analizar confiablemente uso y cobertura de las tierras. En otras regiones estas imágenes son usadas con buenos resultados para el mapeo y detección de cambios, como señalado por Asner et al. (1998) y Coppin et al. (2004). En la Provincia de Entre Ríos, el bosque nativo fue mapeado con estas imágenes, obteniendo buenos resultados en el año 2008, por Sabbatini et al. (2009). Este mapeo obtuvo la “Zonificación del bosque natural” solicitada por el Gobierno Provincial y según estos resultados la provincia contaba, en 2008, con más de un millón de hectáreas de monte nativo con fragmentación y degradación alta. En años posteriores a este mapeo se han producido cambios que deberán ser detectados para la actualización de mapas, monitoreo y análisis de los procesos que conducen a la fragmentación del bosque natural.

En la Provincia de Entre Ríos, las diferentes condiciones topográficas, de cobertura y de conservación, plantean la necesidad de crear o adaptar diferentes metodologías para la articulación de técnicas digitales de detección y de caracterización temática. Anteriormente en el CEREGEO - Centro Regional de Geomática de la UADER, se realizó una primera aproximación en un área piloto del norte provincial, con una detección histórica de los desmontes cada 5 años desde 1980 a 2010, en una faja estrecha, de 70km, entre los ríos Paraná y Uruguay, centrada en el Departamento Federal, publicado en Maldonado et al. (2012). Esta experiencia piloto mostró la necesidad de adaptar las técnicas y su articulación metodológica para un análisis operativo de la dinámica de bosques y áreas naturales en la región.

Un sistema operacional para el monitoreo continuo de los procesos ambientales a lo largo del tiempo enfrenta los cambios tecnológicos producidos en los sistemas satelitales y debe utilizar técnicas de procesamiento capaces de

incorporar imágenes producidas por los nuevos sensores satelitales desarrollados durante su tiempo de operación. La mayoría de las técnicas de detección de cambios no obtienen buenos resultados con imágenes de diferentes sensores, como mencionado por Coppin et al. (2004). Actualmente, la forma más usada para incorporar las nuevas imágenes como fuentes de datos a los sistemas de monitoreo es la “Detección post-clasificación”. Esta técnica consiste en la generación sucesiva de mapas temáticos independientes para cada fecha y el posterior cruzamiento entre estos. Esta técnica es intuitivamente correcta pero introduce al sistema de monitoreo el fenómeno llamado “propagación de errores”. Este amplifica los errores en los mapas finales por la multiplicación de los errores de los mapas antecedentes, y es en general inaceptable que para fines cuantitativos, según muestran Lunetta et al. (1981), Stow et al. (1990), Congalton y Green (1999), Pontius Jr y Millones (2011).

Un sistema operacional para el monitoreo necesita de una técnica capaz de la detección digital entre imágenes de diversas fuentes para evitar la necesidad de mapas intermedios, una de estas técnicas fue desarrollada y presentada en Maldonado et al. (2002), Maldonado et al. (2005b), Santos et al. (2005a) y Maldonado et al. (2007). Esta técnica de detección digital de cambios, “RCEN-Rotación Controlada por Eje de No-Cambio” permite el uso simple de imágenes de diferentes sensores y ya ha mostrado buenos resultados en regiones semiáridas de la catinga de Brasil y chaco árido en Maldonado (2007), regiones sub-húmedas del sur de Chile con explotaciones forestales en Frau et al. (2009), Bosque mixto del sur de Brasil en Scharlau et al. (2012), en bosque húmedo de la selva amazónica en Graça et al. (2008), Maldonado et al. (2009) y en el bosque atlántico por Arasato et al. (2012). La característica principal de esta técnica es la

rotación controlada del espacio bi-temporal de la radiometría de las imágenes digitales, produciendo buenos resultados usando imágenes sin calibrar o de diferentes sensores, como presentado en Maldonado et al. (2007).

El objetivo general del proyecto es el desarrollo de un sistema operacional para la detección y monitoreo de los cambios interanuales ocurridos en la cobertura vegetal del bosque nativo del centro-norte de Entre Ríos desde el año 2008 a la actualidad. Para alcanzar esos objetivos se desarrollará una metodología que articule las técnicas de detección de cambios con las técnicas de caracterización temática de los cambios. Posteriormente se aplicarán las técnicas RCEN para detección exploratoria y planificación del trabajo de campo. Se adecuarán y aplicarán técnicas de levantamiento de campo para la caracterización temática de los cambios en el bosque natural. Y se aplicarán las técnicas RCEN y RCEN multi-espectral para la detección de los cambios producidos y se actualizarán los mapas de vegetación.

### **Líneas de Investigación y Desarrollo**

Para suplir estas necesidades metodológicas el Proyecto es ejecutado por un equipo multidisciplinario con capacitación para desarrollar y adaptar Técnicas digitales de procesamiento de imágenes, adaptar y aplicar Técnicas de levantamiento Botánico y comprender la dinámica ambiental en una región donde predomina la actividad agropecuaria. La metodología ha desarrollar en este proyecto, articulará técnicas de pre-procesamiento de imágenes digitales, de detección de cambios radiométricos, mapeo de los cambios y trabajos de levantamiento de la verdad de campo. A seguir se detallan algunas etapas de la metodología:

Consolidación de la base de datos cartográficos en el SIG- Sistema de Información Geográfica.

Obtención de las imágenes detección con la técnica RCEN para orientar el trabajo de campo y la recolección a campo de muestras de cambios intra-anales e inter-anales. Posteriormente la adecuación de las técnicas digitales de rotación radiométrica controlada para la detección de cambios en región de condición ambiental húmeda y estacionalmente complejas del centro-norte de Entre Ríos. Levantamientos fito-fisonómicos para caracterizar los cambios, según Maldonado et al. (2005a). Aplicación de la técnica "RCEN multi-espectral" y obtención del mapa de cambios correspondiente al año, con una leyenda temática de intensidad y dirección de los cambios. Aplicación de la Técnica de actualización de mapas: Esta técnica se basa en un simple cruzamiento entre el mapa de uso anterior y el mapa de intensidad de los cambios, interpretando los cambios temáticos a través de una leyenda compleja en el mapa actualizado la que luego será simplificada para obtener el mapa final. Análisis de la confusión del mapeo según "Cantidad y localización del desajuste" según Pontius y Millones (2011). Ajuste y articulación de las técnicas para obtener un sistema operacional ajustado a la complejidad del ambiente del norte provincial. Formulación de la Metodología y test de la metodología.

### **Formación de recursos Humanos**

Durante la ejecución del proyecto serán ofrecidas dos becas de iniciación por año para alumnos de las licenciaturas en biología y en informática, para apoyo al muestreo y descripción de la vegetación y programación de rutinas en el SIG-Sistema de Información Geográfica. También se incorporarán dos becarios por año en la carrera de Licenciatura en Biología para desarrollar sus trabajos finales de formación de grado e actividades iniciales para su especialización futura en el uso de geotecnologías (subsedes Diamante y/o Paraná). Algunas actividades de actualización de mapas

de ambientes serán articuladas para la producción de trabajos finales de formación de los alumnos de grado. Los resultados intermedios de la etapa de detección, serán generados con participación de dos becarios de Doctorado del CONICET, para seguimiento y especialización de incendios

### Resultados esperados

El principal resultado es una metodología operacional para el monitoreo anual de los cambios en la cobertura vegetal del bosque nativo. Esta metodología operacional minimiza el efecto de las fuentes de error, errores que son propagados a los largo del tiempo. Estas fuentes de errores están asociadas con la subjetividad de la intervención humana, sobre todo cuando el personal en operación del sistema cambia durante el tiempo de funcionamiento de sistema de monitoreo. El sistema producirá periódicamente Mapas de intensidad de los cambios del bosque natural del centro-norte de Entre Ríos. Mapas de cambios de uso producidos en áreas de bosque natural del centro-norte de Entre Ríos y Tablas de cuantificación de los cambios.

### Referencias bibliográficas

- Asner G.P.; Wessman C.A.; Schimel D.S. 1998. Heterogeneity of savanna canopy structure and function from imaging spectrometry and inverse modeling. *Ecological applications*, 8(4): 1022-1036, 1998.
- Arasato, L.S.; Santos, J.R.; Maldonado, F.D.; Amaral, S.; Rennó, C.D. 2012. Detecção de mudanças da paisagem a partir de análise multisensor e multitemporal em associação com variáveis geomorfométricas no domínio da floresta atlântica. *Revista Brasileira de Cartografia*, 64(4): 475-486, 2012.
- Congalton, R.G.; Green, K. Assessing the Accuracy of Remotely Sensed Data: Principles and Practice. Mapping sciences series. New York: CRC Press Inc., 1999. 98p.
- Coppin, P.; Jonckheere, I. ; Nackaerts, K.; Muys, B.; Lambin, E. 2004. Digital change detection methods in ecosystem monitoring: a review. *International Journal of Remote Sensing*, 25(9): 1565-1596.
- Frau, C.M.; Santos, J.R.; Maldonado, F.D.; Valenzuela, J.G.; Valeriano, M.M.; Rojas, Y.O.; Hernandez, Y.M. 2010. Caracterización y monitoreo de paisaje semiárida en la Región del Maule mediante datos satelitales. *Revista Brasileira de Engenharia Agrícola e Ambiental*, 14: 600-668, 2010.
- Graça, P.M.L.; Maldonado, F. M.; Santos, J. R.; Soares, J. V. 2008. Detecção de corte seletivo de madeira por técnica de rotación radiométrica en la floresta amazónica. *Ambiência*, 4: 97-106.
- Lunetta, R.; Congalton, R.; Frenstermaker, L.; Jensen, J.; McGwire, K.; Tinney, L. 1981. Remote Sensing and geographic information system data integration: error sources and research issues. *Photogrammetric Engineering and Remote Sensing*, 57(6): 677-687.
- Maldonado, F.D.; Santos, J.R.; Carvalho, V.C. 2002. Land use dynamics in the semiarid region of Brazil (Quixabá-PE): characterization by principal components analysis. *International Journal of Remote Sensing*, 23(23): 5005-5013.
- Maldonado, F.D.; Carvalho, V.C.; Souza, C.L.; Martinelli, M.; Pinheiro, O.J. ; Santos, F.F. 2005a. Determinación de la longitud de transecta para el relevamiento fisonómico-estructural de la vegetación del semiárido para suministrar datos a las técnicas de percepción remota orbital. *Multequina*, 13: 1-14, 2005.
- Maldonado, F.D.; Santos, J.R.; Pastor, C.Q.; Manso, A.F. 2005b. Radiometric rotation technique to monitor degradation and regeneration features in Brazilian semi-arid region.. In: IEEE International Geoscience and Remote Sensing Symposium - 25th IGARSS'05, 2005, Seoul. IEEE International Geoscience and Remote Sensing Symposium IGARSS'05: Harmony between Man and Nature. Seoul, Korea: IGARSS, 2005. p. 2315-2318.



- Maldonado, F. D. ; Santos, J.R.; Graça, P.M. 2007. Change detection technique based on the radiometric rotation controlled by no-change axis, applied on a semi-arid landscape. *International Journal of Remote Sensing* , 28(8): 1789-1804.
- Maldonado, F.D.; Graça, P.M.L.; Santos, J.R. 2009. Aplicação da fração-solo por rotação radiométrica para detecção do corte seletivo em domínio da floresta Amazônica. In: Simpósio Brasileiro de Sensoriamento Remoto, 14. (SBSR), 2009, Natal. Anais... São José dos Campos: INPE, 2009. p. 2823-2830.
- Maldonado, F.D.; Sione, W.F.; Aceñolaza, P.G. 2012. Mapeo de desmontes en áreas de bosque nativo de la Provincia de Entre Ríos. *Ambiência*, 8(Es): 532-532.
- Pontius, R. G.; Millones, M. 2011. Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing*, 32(15): 4407-4429.
- Sabattini, R.A.; Ledesma,S.; Brizuela, A.; Sabattini, J. 2009. Zonificación de los bosques nativos en el Departamento La Paz (Entre Ríos) según las categorías de conservación. Informes FCA UNER y la Dirección General de Recursos Naturales, Sec. de la Producción Gob.de Entre Ríos. FCA UNER: Oro Verde, Octubre 2009. 30p.
- Santos, J.R.; Maldonado, F.D.; Graça, P.M.L. 2005a. New change detection technique using ASTER and CBERS-2 images to monitor Amazon tropical forest. In: International Geoscience and Remote Sensing Symposium-25th IGARS2005, Seoul, Korea. IEEE International Geoscience and Remote Sensing Symposium IGARSS-05 Harmony between Man and Nature. Seoul : IGARSS, 2005. v. 7. p. 5026-5028.
- Scharlau, L.A.; Disperatti, A.A.; Maldonado, F.D.; Araujo; A.J. Atualização de mapas de uso com mapas de intensidade das mudanças na cobertura vegetal. X Seminario de Atualização em Sensoriamento Remoto e Sistemas de Informações geograficas Aplicados à Engenharia Florestal. 15-18 octubre 2012, Curitiba, Parana-Brasil.
- Stow, D.A.; Collins; D.; McKinsey, D. 1990. Land use change detection based on multi-date imagery from different satellite sensor systems. *Geocarto International*, 5(3): 3-12.

# ESTRATEGIAS DIDÁCTICAS EN EL USO DE HERRAMIENTAS SOFTWARE PARA FAVORECER LA COMPRESIÓN DE LOS ALUMNOS EN LA ENSEÑANZA SOBRE COMPILADORES

Gladis Sequeira, Silvia Zajaczkowski, Cristian Kornuta, Gabriela Gómez

Departamento de Informática/Facultad de Ciencias Exactas Químicas y Naturales/  
Universidad Nacional de Misiones

Felix de Azara 1552, Posadas, Misiones

Te: 376-422186

gladis.sequeira@gmail.com, silviazaj@yahoo.com.ar, cristian.kornuta@yahoo.com.ar, gabrielaasc@gmail.com

## Resumen

En este artículo se expone la investigación que se realiza dentro de la asignatura Teoría de la Computación de la carrera Licenciatura en Sistemas de Información, perteneciente a la Facultad de Ciencias Exactas, Químicas y Naturales de la Universidad Nacional de Misiones. El objetivo de la experiencia es constituir una propuesta didáctica curricular en torno a las prácticas realizadas con herramientas de software dentro de la cátedra, con el propósito de lograr una mejor comprensión y afianzamiento de los conocimientos en temas teóricos y obtener un prototipo funcional por parte de los alumnos al finalizar el cursado.

**Palabras clave:** Educación universitaria, Estrategias didácticas, Teoría de la Computación, Compiladores.

## Contexto

Este trabajo se enmarca en el “Programa de Investigación en Computación” del Centro de Investigación y Desarrollo Tecnológico de la Facultad de Ciencias Exactas Químicas y Naturales de la Universidad Nacional de Misiones.

Dentro del proyecto se desempeñan docentes, tesistas, y becarios de las carreras de Analista en Sistemas de Computación, Licenciatura en sistemas de Información.

## Introducción

La planificación y ejecución de procesos de enseñanza-aprendizaje plantea un gran desafío a los docentes responsables de la cátedra Teoría de la Computación que se dicta en el primer cuatrimestre del tercer año de la carrera Licenciatura en Sistemas de Información de la Facultad de Ciencias Exactas, Químicas y Naturales perteneciente a la Universidad Nacional de Misiones. La tarea docente se complejiza a causa del grado de abstracción que requiere el cursado y la necesidad de relacionar todos los contenidos aprendidos en las cátedras anteriores de la carrera hasta ese momento.

El equipo docente, que se fue conformando a partir del año 2010, se abocó en su primer dictado a la enseñanza de los aspectos teóricos vinculados a los principios de diseño de compilación. Para la resolución de las guías prácticas se optó por aplicar el modelo “grupo cooperativo” organizándose, en esa oportunidad, diez grupos de trabajo al comenzar el cursado. Los prácticos consistieron en cuestionarios teóricos y construcción de mapas conceptuales integratorios. Al finalizar cada práctica, se realizaron la puesta en común del tema estudiado.

En dictados posteriores se incursionó en el desarrollo básico de prototipos de algoritmos, logrando que los alumnos trabajen en un proyecto donde implementaron un lenguaje de programación acotado tomando como referencia algún lenguaje imperativo aprendido

en cátedras anteriores. Así se logró un mejor afianzamiento de los contenidos teóricos.

A partir de los resultados obtenidos con el cursado del año 2011 la asignatura se planteó la necesidad de incluir herramientas informáticas, tales como software de simulación, que colaboren en el proceso de afianzar lo aprendido.

Como se expresa en [1], la posibilidad de probar un modelo, evaluarlo, ajustarlo y volverlo a probar hasta que el resultado sea satisfactorio contribuye al desarrollo de patrones o esquemas mentales adecuados para razonar y comprender. Así es posible que la aplicación de refinamiento sucesivo facilite al sistema cognitivo del sujeto la tarea de anticipar el comportamiento del modelo y permita redirigir ese esfuerzo hacia la comprensión de la teoría que respalda la práctica en proceso.

El software sería, en este caso, el elemento que simplificaría ese proceso de refinamiento sucesivo y ayudaría en la generación de relaciones pertinentes [1] que permitiese dotar de significado las expresiones regulares, autómatas finitos, gramática libre de contexto y árboles sintácticos que se emplean para describir el comportamiento de las distintas fases de un compilador [2] [3] [4].

## Objetivos

El objetivo de este proyecto es mejorar la comprensión de los alumnos dentro de la cátedra a partir de la incorporación de una estrategia didáctica que incluya herramientas software en las prácticas realizadas en la enseñanza de compiladores.

## Implementación de Simuladores y Metacompiladores para un Aprendizaje Constructivista

Nuestra experiencia nos llevó a proponer la utilización de simuladores, como ser el JFLAP [5] para la teoría de autómatas y el ANTRLWorks [6] [7] para el árbol sintáctico, que proveen un interesante enlace entre la teoría

y la práctica, entendiéndose que los simuladores son solo herramientas de software que ayudarían al alumno a comprender los temas centrales de la cátedra.

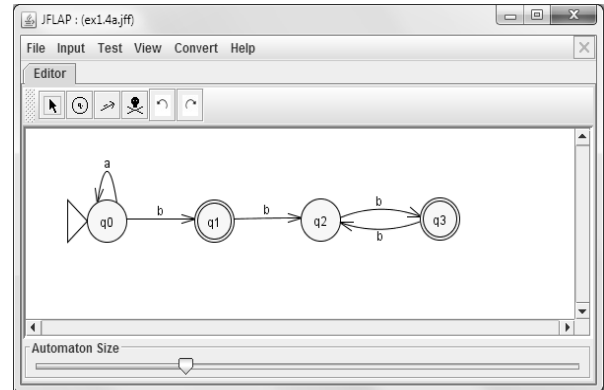


Fig. 1. JFLAP/ Autómatas finitos

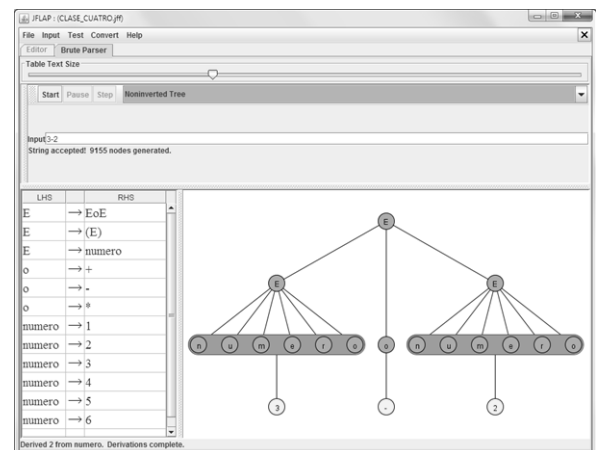


Fig. 2. JFLAP/ Arbol sintactico

Buscando estrategias didácticas que fortalezcan la comprensión de los alumnos y acompañen el Proceso Enseñanza-Aprendizaje dentro de la cátedra Teoría de la Computación, se optó por aplicar el enfoque constructivista. Este modelo define que el conocimiento se construye a partir de los esquemas que el individuo ya posee, sus conocimientos previos. Se trata de un proceso que se desarrolla día a día, centrado en la persona que aprende y que le permite adquirir, no solamente un nuevo conocimiento, sino también una nueva competencia que podrá aplicar a futuras situaciones [8].

Con esta intensión se planteó a los alumnos, en la segunda etapa del cursado, el desarrollo de

un trabajo integrador que consistió en la implementación de un prototipo de lenguaje. El estudiante tenía dos posibles caminos a seguir, optar por el uso de un lenguaje de programación que conociera o utilizar algún metacompilador; en ambos casos obtendría soporte y respuestas a consultas referentes a la construcción de su propio lenguaje.

Si bien se presentaron las características generales de diferentes metacompiladores, como ser PLY (Python, Lex y Yacc) [9], ANTRL, JFlex y Bison, un relevamiento expuso que en cinco de los siete grupos que permanecían activos en el último mes de cursado, se encontraba al menos un alumno que tenía conocimiento de Python por haber realizado un curso de Extensión el año anterior. Esta situación decidió la presentación más detallada del metacompilador PLY que se caracteriza por implementarse en su totalidad en Python, además es sencillo de usar y proporciona una extensa comprobación de errores. PLY proporciona la base lex / yacc.

Cada grupo decidió, según sus conocimientos de programación, las herramientas de software a utilizar para implementar las primeras fases del prototipo avanzado de su lenguaje.

Durante esta etapa se continuó trabajando en forma grupal, con el modelo de “aprendizaje cooperativo”, considerando que la construcción del conocimiento, según el enfoque utilizado, se produce cuando el alumno interactúa con otros [8]. La consigna incluyó la presentación del producto obtenido.

## Líneas de investigación y desarrollo

Algunos de nuestros propósitos próximos son:

- Elaborar un Cuaderno de Cátedra con la finalidad de presentar al alumno una base teórica, guía de trabajo a seguir, ejercitaciones prácticas y cuestionarios.
- Conformar un soporte multimedial que permita contener al alumno en el transcurso del dictado de la cátedra y

que le ayude a profundizar y ampliar todos aquellos temas que se tratan en la cátedra.

- Diseñar e implementar alguna herramienta de programación visual, orientado a un lenguaje específico con propósitos educativos.
- Diseño e implementación de un Framework de trabajo orientado a un lenguaje específico o a la conjunción de las herramientas de desarrollo.

## Resultados y Objetivos

De los diez grupos cooperativos que se conformaron al iniciar el dictado 2012, tres abandonaron el cursado antes de iniciar el trabajo integrador. Si bien no se investigaron las causas de tal decisión, se presume que las razones podrían deberse a sobrecarga de actividades académicas y laborales.

Se realizó una encuesta anónima a fin de analizar la opinión de los estudiantes respecto a la incorporación de las herramientas JFLAP y ANTLRWorks en las prácticas correspondientes a las fases analizador léxico y sintáctico como apoyo de los conceptos teóricos. La misma arrojó como resultado que el 72% de los alumnos, sobre un total de 18, concluyeron que las prácticas realizadas con software le sirvieron para entender los conceptos teóricos de la asignatura.

Los siete grupos que completaron el cursado 2012 lograron desarrollar, con diferentes grados de complejidad, el prototipo solicitado. Cinco grupos trabajaron con la herramienta PLY y dos lo implementaron usando .NET.

## Conclusiones

La experiencia que implicó la incorporación de las herramientas de software dentro de la cátedra produjo buenos resultados para ambas partes involucradas en el proceso: los docentes se encontraron con actividades productivas realizadas por los estudiantes en sus prácticas, y los alumnos demostraron mejor destreza y



confianza en ellos mismos al momento de trabajar con las herramientas y confeccionar su propio prototipo de lenguaje de programación.

Por ser éste el primer dictado en que la asignatura implementa un trabajo integratorio con estas características, no existen datos previos que puedan servir de comparación, aunque el equipo de cátedra reconoce el avance logrado en cuanto al grado de comprensión de los contenidos e interés en las actividades planteadas, cuestiones que se vislumbran al observar la calidad de los prototipos diseñados y la originalidad de los mismos.

Una encuesta posterior al trabajo integrador arrojó que para el 82% de los estudiantes fue un desafío motivador para seguir investigando y les permitió integrar los conceptos teóricos dados en la cátedra.

### Formación de Recursos Humanos

En esta línea de investigación se está desarrollando una tesis de grado en la carrera Licenciatura en Sistemas de Información. Conjuntamente se prevé conformar un grupo de investigación en el marco del Departamento de Informática de la F.C.E.Q. y N. orientando las líneas de investigaciones futuras a la construcción de compiladores.

### Referencias

- [1]. Santos, G., Otero, M.R., Fanaro, M.A.: ¿Cómo usar software de simulación en clases de física?. Cuaderno Catarinense de Ensino de Física vol. 17 n.1 (2000)
- [2]. Kenneth, L.: Construcción de compiladores, principios y práctica. International Thomson Ed. (2004)
- [3]. Ruiz Catalan, J.: Compiladores TEORÍA E IMPLEMENTACIÓN. International Thomson Ed. (2010)
- [4]. Aho, A., Lam, M., Sethi, R., Ullman, J., Compilers, Principles, Techniques & Tools. Pearson Addison – Wesley (2008)
- [5]. JFlap, <http://www.jflap.org/>
- [6]. Parr, T.: The Definitive ANTLR reference. Raleigh (North Carolina) (2007)
- [7]. ANTLRWorks: The ANTLR GUI Development Environment, [www.antlr.org/works/](http://www.antlr.org/works/)
- [8]. Carretero, M.: Constructivismo y Educación. Aique (1993)
- [9]. PLY (Python Lex-Yacc), [http://www.dabeaz.com/ply/ply.htm#ply\\_nn1](http://www.dabeaz.com/ply/ply.htm#ply_nn1)

# El Pensamiento Crítico y los Entornos Virtuales de Aprendizaje - Perspectiva Sistémica

Paola D. Budán<sup>3</sup>

Isabel Velázquez<sup>3</sup>

Av. Belgrano 1912, (4200CPB) Santiago del Estero, Argentina  
e-mail: pbudan@unse.edu.ar - kereyes@unse.edu.ar

## Resumen

Un *entorno virtual* de aprendizaje es una herramienta tecnológica que constituye un entorno dinámico, con determinadas condiciones físicas y temporales, que posibilitan y favorecen el aprendizaje. El *pensamiento crítico* desde un punto de vista práctico, es un proceso mediante el cual se usa el conocimiento y la inteligencia para llegar, de forma efectiva, a la posición más razonable y justificada sobre un tema. Se basa en valores intelectuales que tratan de ir más allá de las impresiones y opiniones particulares, por lo que requiere claridad, exactitud, precisión, evidencia y equidad. Tiene por tanto una vertiente analítica y otra evaluativa. Sin embargo, se trata de un enfoque que por ser analítico, es simplificador. De manera de desarrollar un modelo abarcador de la complejidad, que tenga en cuenta las características, el ambiente y los componentes del entorno virtual de aprendizaje, así como muchos otros aspectos relevantes para conformar el modelo, tales como pueden ser las emociones del aprendiz, es que se propone un desarrollo basado en una perspectiva *sistémica*.

**Palabras clave:** Entornos Virtuales de Aprendizaje - Pensamiento Crítico - Sistémica

## Contexto

Esta línea de investigación está inserta en el marco del desarrollo de un trabajo final para optar por el título de Especialista en Enseñanza de las Tecnologías de la Universidad Nacional de Santiago del Estero, trabajo éste denominado "*Propuesta de un modelo para incorporar el Pensamiento Crítico a los Entornos Virtuales de Aprendizaje (EVA), desde una perspectiva Sistémica*". Se vincula con el proyecto de investigación denominado "Demarcación disciplinar de la Informática Educativa para orientar el desarrollo, uso y evaluación de recursos que soportan el aprendizaje de personas y máquinas", C/113, dirigido por Isabel Velázquez.

## Introducción

Con frecuencia se afirma que los entornos virtuales promueven por sí mismos nuevas formas de aprendizajes, que responden a nuevas teorías. Se da por supuesto que los nuevos medios permiten superar los enfoques tradicionales de la enseñanza, centrados en las clases expositivas y en la reproducción acrítica de los contenidos por parte del alumno, para pasar a una apropiación autónoma, con espacios de discusión y construcción de conocimientos [6]. En los últimos años ha habido cambios importantes en la concepción del aprendizaje y

de la enseñanza, el concepto de aprendizaje se amplía o ramifica a través de diversos calificativos: natural, significativo, colaborativo, distribuido, entre otros. Los cambios en la cultura del aprendizaje históricamente están ligados al desarrollo de nuevas tecnologías. Numerosos aportes se destacan al respecto, aquellos que desarrollaron investigaciones del aprendizaje desde la perspectiva de los alumnos de nivel universitario, identificando tres maneras de analizar las tareas de aprendizaje: profundo, superficial y estratégico. El dilema principal parece radicarse en el hecho que se ha genera alrededor de cada disciplina; lo que implica que, para cumplir con su misión, la universidad contemporánea debe brindar a sus estudiantes una educación que los prepare tanto para las exigencias de hoy, como para las de los próximos 70 años de su vida. La capacidad de análisis, inferencia, interpretación, explicación y evaluación, sustentadas por la autorregulación y actitud investigativa, vigilante, honesta y flexible se convierte en lo que distingue al profesional que está capacitado para enfrentar los desafíos de la sociedad moderna.[1] Por otro lado, el pensamiento crítico [4] es un modo de pensar, en el cual el pensante mejora la calidad de su pensamiento al apoderarse de las estructuras inherentes del acto de pensar y al someterla a los estándares intelectuales. El aporte de este trabajo será la propuesta de un modelo que permita integrar el Pensamiento Crítico en los EVA, pero al ser el pensamiento crítico simplificador de la complejidad y analítico por naturaleza, se vuelve necesario incluir algún tipo de pensamiento que permita recuperar con inteligencia la complejidad, vista como los sistemas filosóficos, los sistemas metodológicos, los sistemas teóricos y los sistemas puramente de aplicación en la educación.

## Líneas de investigación y desarrollo

La presente línea de investigación estudiará la integración del Pensamiento Crítico en los EVA, mediante un modelo que permita aprender por *conectividad*, desde una perspectiva sistémica. Para ello:

- Se identificarán los lineamientos del Pensamiento Crítico y se los comentará.
- Se desarrollará un modelo que permita integrar dichos lineamientos en un EVA.
- Se tenderá a incorporar el sentido común a la praxis en un EVA.
- Se diseñará un modelo que permita integrar diferentes capacidades, incorporando otros tipos de pensamientos referidos a la observación, reflexión, entre otros.

## Soporte Informático Educativo Basado En Pensamiento Crítico y Creatividad

Existe una relación inversa entre la capacidad de un sistema computacional para suplantar actividades intelectuales y el grado en el cual estas actividades se caracterizan por la individualidad y la creatividad [2]. Entre las dimensiones que pueden emplearse para pensar sobre el aprendizaje, o los niveles de pensamiento crítico aplicables a este contexto, se tiene:

- La evaluación personal,
- Los criterios de consenso,
- Los criterios autónomos,
- La evaluación de pares.

Los desarrollos basados en computadora que soporten el pensamiento crítico, deben tolerar formas de actividad intelectual divergente (pensamiento creativo) y convergente (estructuras simples). Esto significa que deben dar soporte a la representación del conocimiento, al modelado de la estructura del conocimiento, realizar vínculos con inferencias basadas en ontologías, agrupamientos por palabras claves y establecer correspondencias entre palabras claves.

### Estándares Intelectuales Universales

Paul [4] sostiene que los estándares intelectuales universales son estándares que deben usarse cuando se quiera verificar la calidad del razonamiento sobre un problema, asunto o situación. Pensar críticamente implica dominar estos estándares. Para ayudar a los estudiantes a aprenderlos, los profesores deben formular preguntas que exploren su capacidad de pensar críticamente; preguntas que provoquen que los estudiantes se responsabilicen por su pensamiento; preguntas que, al formularse con regularidad en el aula, se vuelvan parte de las preguntas que los estudiantes necesitan formular. Algunos estándares del Pensamiento Crítico son:

- *Claridad*: Es un estándar esencial. Si un planteamiento es confuso, no se puede saber si es exacto o relevante.
- *Exactitud*: Un enunciado puede ser claro pero inexacto.
- *Precisión*: Un planteamiento puede ser claro y exacto pero impreciso.
- *Relevancia*: Un planteamiento puede ser claro, exacto y preciso pero irrelevante al asunto o a la pregunta.
- *Profundidad*: Un enunciado puede ser

claro, exacto, preciso y relevante pero superficial.

- *Amplitud*: Una línea de razonamiento puede ser clara, exacta, precisa, relevante y profunda pero carecer de amplitud.
- *Lógica*: Cuando las ideas combinadas se apoyan entre sí y tienen sentido, el pensamiento es lógico.

### El Conectivismo

Según Siemens [5], el conductismo, el cognitivismo y el constructivismo son las tres grandes teorías de aprendizaje utilizadas más a menudo en la creación de ambientes instruccionales. Estas teorías, sin embargo, fueron desarrolladas en una época en la que el aprendizaje no había sido impactado por la tecnología informática. En los últimos veinte años, la tecnología informática ha reorganizado la forma en la que vivimos, nos comunicamos y aprendemos. Las necesidades de aprendizaje y las teorías que describen los principios y procesos de aprendizaje, deben reflejar los ambientes sociales subyacentes. El conectivismo es orientado por la comprensión que las decisiones están basadas en principios que cambian rápidamente. Continuamente se está adquiriendo nueva información. Tal como menciona Siemens, los principios del conectivismo son:

- El aprendizaje y el conocimiento dependen de la diversidad de opiniones.
- El aprendizaje es un proceso de conectar nodos o fuentes de información especializados.
- El aprendizaje puede residir en dispositivos no humanos.
- La capacidad de saber más, es más crítica que aquello que se sabe en un momento dado.



- La alimentación y mantenimiento de las conexiones es necesaria para facilitar el aprendizaje continuo.
- La habilidad de ver conexiones entre áreas, ideas y conceptos es una habilidad clave.
- La actualización (conocimiento preciso y actual) es la intención de todas las actividades conectivistas de aprendizaje.
- La toma de decisiones es, en sí misma, un proceso de aprendizaje. El acto de escoger qué aprender y el significado de la información que se recibe, es visto a través del lente de una realidad cambiante. Una decisión correcta hoy, puede estar equivocada mañana debido a alteraciones en el entorno informativo que afecta la decisión.

### Los Entornos Virtuales de Aprendizaje

Caracterizar el aprendizaje en entornos virtuales como un proceso de construcción supone, esencialmente, afirmar que lo que el alumno aprende en un entorno virtual no es simplemente una copia o una reproducción de lo que en ese entorno se le presenta como contenido a aprender, sino una reelaboración de ese contenido mediada por la estructura cognitiva del aprendiz. El aprendizaje virtual, por tanto, no se entiende como una mera traslación o transposición del contenido externo a la mente del alumno, sino como un proceso de (re)construcción personal de ese contenido que se realiza en función, y a partir, de un amplio conjunto de elementos que conforman la estructura cognitiva del aprendiz: capacidades cognitivas básicas, conocimiento específico de dominio, estrategias de aprendizaje, capacidades metacognitivas y de autorregulación, factores afectivos, motivaciones

y metas, representaciones mutuas y expectativas. La actividad mental constructiva que el alumno, al poner en juego este conjunto de elementos, desarrolla en torno al contenido se configura, desde esta perspectiva, como clave fundamental para el aprendizaje, y la calidad de tal actividad mental constructiva, por lo mismo, se configura como clave fundamental para la calidad del aprendizaje: ni toda actividad que el alumno realiza cuando aprende conlleva actividad mental constructiva, ni toda actividad mental constructiva es igualmente deseable ni óptima para un aprendizaje de calidad.[3]

### El Pensamiento Sistémico

El Pensamiento Sistémico es el proceso de comprender cómo los sistemas interactúan o se influyen entre sí constituyendo un todo. Se define como un enfoque para la resolución de problemas, en donde los problemas forman parte de un sistema global. Es un conjunto de hábitos o prácticas dentro de un marco de trabajo que se basa en la creencia de que las partes componentes de un sistema pueden entenderse mejor en el contexto de las relaciones entre sí y con otros sistemas, en lugar de manera aislada. El pensamiento sistémico se centra en cíclico y no lineal de causa y efecto. Cuando las percepciones simples no son suficientes, surge la necesidad de contar con otras herramientas que respondan qué, quién, cuánto, dónde. Se requiere entonces ampliar el sentido común, ya sea en la forma de desarrollar mejores herramientas para realizar percepciones simples, métodos mejores para simplificar percepciones complejas, o mejores enfoques para hacer juicio sobre la base de estas percepciones. Tal como se sostuvo en las conversaciones de la IFSR (International Federation for Systems Research) Abril 2012) se están haciendo muchos intentos para refinar la noción de sentido común, pues bien, las actuales concepciones de la cien-

cia enfrentan el mismo obstáculo de ampliar el repertorio más allá de lo simple.

## Resultados y Objetivos

El objetivo general de esta línea de investigación es *contribuir a la integración del Pensamiento Crítico en los Entornos Virtuales de Aprendizaje, mediante un modelo que permita aprender por conectividad, desde una perspectiva sistémica*. El objetivo particular del plan es desarrollar un modelo, aunque más no sea esquemáticamente como orientación, que permita integrar dichos lineamientos en un Entorno Virtual de Aprendizaje, y permitir la integración de diferentes capacidades, incorporando otros tipos de pensamientos referidos a la observación, reflexión, entre otros.

## Formación de Recursos Humanos

Actualmente el equipo de trabajo de esta línea de investigación se encuentra compuesto por estudiantes avanzados de la carrera de Licenciatura en Sistemas de Información, de la Especialización en Enseñanza de las tecnologías, sociólogos y docentes del Laboratorio de Informática de la Facultad de Ciencias Exactas y Tecnologías de la UNSE.

## Referencias

- [1] F. Andrei, N. Fëdorov. *Siglo XXI, la universidad, el pensamiento crítico y el foro virtual*.
- [2] Nigel Ford. *Web-Based Learning through Educational Informatics. Information Science Meets and Educational Computing*.
- [3] Javier Onrubia. *Learning and teaching in virtual environments: joint activity, teacher assistance and knowledge construction*.
- [4] Elder L Paul R. *Mini guía para el pensamiento crítico. Fundación para el Pensamiento Crítico*.
- [5] George Siemens. *Conectivismo. Una teoría de aprendizaje para la era digital*.
- [6] Espiro Susana. *El aprendizaje en Entornos Virtuales*.

# Motores de Juegos e Inteligencia Artificial para la Enseñanza

Frittelli, V. – Strub, A. M. – Destéfánis, E. – Steffolani, F – Teicher, R. – Tartabini, M. – Bett, G. – Fernández, J. – Serrano, D..

Proyecto: *Diseño de Motores de Juegos y Componentes para Enseñanza y Aplicaciones de la Inteligencia Artificial (Código: UTN1707)*

Carrera de Ingeniería en Sistemas de Información – UTN: Facultad Regional Córdoba

Dirección: M. López eqz. Cruz Roja Argentina – Ciudad Universitaria (5000) – Córdoba (Capital)  
Teléfono: (0351) 5986000

{ [vfrittelli@gmail.com](mailto:vfrittelli@gmail.com) - [anastrub@gmail.com](mailto:anastrub@gmail.com) - [edestefanis1@gmail.com](mailto:edestefanis1@gmail.com) - [fsteffolani@gmail.com](mailto:fsteffolani@gmail.com) - [rteicher@gmail.com](mailto:rteicher@gmail.com) - [marcelatartabini@hotmail.com](mailto:marcelatartabini@hotmail.com) - [gfbett@gmail.com](mailto:gfbett@gmail.com) - [jujulifer@gmail.com](mailto:jujulifer@gmail.com) - [diegojserrano@gmail.com](mailto:diegojserrano@gmail.com) }

## Resumen

En la organización curricular de la carrera de Ingeniería en Sistemas de Información es poco común la integración de asignaturas de diferentes áreas aun cuando existan temáticas, prácticas o incluso docentes en común. Si bien es habitual identificar estrategias de integración entre asignaturas correlativas (integración vertical), no ocurre lo mismo con asignaturas de un mismo nivel (integración horizontal) ni con asignaturas de otras áreas. Distintas materias incluyen el estudio de temas que permitirían al alumno desarrollar software para dominios diversos. Por otro lado, otras asignaturas solicitan el desarrollo de agentes inteligentes que suelen plantearse como “jugadores” virtuales para simulaciones de estrategias frente al accionar de un adversario; y la evaluación de estos trabajos requiere un esfuerzo considerable por parte del docente. Se plantea aquí una propuesta de integración consistente diseñar actividades prácticas que involucren a alumnos de asignaturas distintas,

mediante la cual algunos desarrollen motores para diferentes juegos o escenarios, y alumnos de otras asignaturas desarrollen los jugadores o agentes. El desafío es que en el desarrollo del motor del juego, se plantee un marco de trabajo general, que permita el diseño de nuevos escenarios desde un modelo de reusabilidad, pero que permita incorporar agentes externos con relativa sencillez.

**Palabras clave:** Algoritmos - Aplicación Educativa - Componentes - Integración - Inteligencia Artificial - Modelos de Juegos -

## Contexto

La propuesta se planteó en el proyecto que se designa oficialmente como *Diseño de Motores de Juegos y Componentes para Enseñanza y Aplicaciones de la Inteligencia Artificial*, presentado para su aprobación definitiva en los primeros días de Marzo de 2013. El proyecto está integrado por docentes

de varias Cátedras de la UTN Córdoba con orientación a la programación y/o a los paradigmas de programación, y surgió como una forma de estudiar, poner en práctica y analizar diferentes ideas de integración entre trabajos de distintas cátedras. Estas ideas surgían de reuniones y conversaciones entre pares y notando que muchos docentes sugerían desde hace tiempo la posibilidad del desarrollo de actividades conjuntas, con algún tipo de marco de trabajo diseñado y prefijado, es que se decidió el planteo de un proyecto de investigación que las formalice.

## Introducción

Diversas asignaturas de la carrera de Ingeniería en Sistemas de Información (como Algoritmos y Estructuras de Datos (AED), Paradigmas de Programación (PPR), y las electivas Tecnología de Software de Base (TSB) y Diseño de Lenguajes de Consulta (DLC)) incluyen en sus programas el estudio de un amplio abanico de estructuras de datos y sus algoritmos relacionados, permitiendo al alumno desarrollar software para dominios diversos. Esta situación es muy aprovechada en evaluaciones parciales, trabajos prácticos y exámenes finales en las cuales se solicita el desarrollo y entrega de programas aplicables a escenarios muy diferentes entre sí. Por otro lado, en otras asignaturas (como Inteligencia Artificial (IAR)) se pide a los alumnos el desarrollo de agentes inteligentes tales como redes de Hopfield, redes neuronales o agentes de búsqueda mediante heurísticas varias. Entre tales actividades suele solicitarse a los alumnos la creación de "jugadores" para ciertos juegos ampliamente conocidos, o bien el desarrollo de componentes a modo de "agentes"

(Dignum, Bradshaw, Silverman, & van Doesburg, 2009) para simulaciones de estrategias frente al accionar de un adversario, de modo que los jugadores o agentes consisten en alguna pieza de software cuya interfaz o protocolo es definida con exactitud por el cuerpo docente y cuyo comportamiento es desarrollado por cada grupo de alumnos. Para la evaluación de esta actividad práctica el docente puede hacer competir a los "jugadores" de diferentes grupos entre sí. Sin embargo, esta tarea requiere un esfuerzo considerable por parte del cuerpo docente que debe programar un "motor" y/o un "tablero" para el juego, es decir, un software que seleccione la estrategia de dos o más grupos y las ejecute en forma simultánea comunicando a cada una de las partes el estado del juego y el movimiento del adversario. Además el motor debe identificar si un jugador intenta un movimiento o una acción no válida según las reglas o restricciones del juego y debe poder detectar una situación ganadora o que identifique la finalización del juego. El desarrollo del motor del juego es una tarea de programación compleja y debe ser realizada una vez por cada juego, ya que es muy difícil utilizar un mismo motor para juegos con reglas diferentes. Por lo tanto se requiere un esfuerzo importante por parte del plantel docente para programar ese motor, ya sea que lo realicen los mismos docentes o que requieran el auxilio de un ayudante de cátedra o becario. En los lenguajes de programación que se utilizan en la carrera resulta relativamente simple y natural construir un marco de trabajo (o *framework*) (Salen & Zimmerman, 2003) a modo de contenedor, dentro del cual se pueda correr el motor de las reglas del juego para que tome a los jugadores o agentes como



componentes. Además, al mismo tiempo que se programan componentes para dicho framework, se continúa el perfeccionamiento de la programación flexible y extensible aprovechando a fondo un paradigma como el de la Programación Orientada a Objetos. En ese sentido, entonces, la propuesta de integración consiste en la realización de actividades prácticas que involucre a los alumnos de asignaturas diversas. Los alumnos de una u otra asignatura desarrollan motores para diferentes juegos y los alumnos de otras asignaturas desarrollan los jugadores o agentes. El desafío es que en el desarrollo del motor del juego, se llegue al planteo de un framework que permita el diseño de esos motores y escenarios en forma genérica (Salen & Zimmerman, 2003), de forma que pueda reusarse y facilitar el trabajo posterior cuando se pidan nuevos escenarios, pero que permita integrar agentes externos con relativa sencillez.

## Líneas de Investigación y Desarrollo

Las principales líneas de acción, investigación y desarrollo previstas para el proyecto son las siguientes:

1. Selección de los juegos o situaciones base y determinación de las reglas: Deben seleccionarse juegos o escenarios de competencia que ofrezcan dificultades razonables y planificadas para el aprendizaje de los alumnos (Schell, 2008), de forma tal que los motores requieran el uso de estructuras de datos y algoritmos que puedan aprender en asignaturas previas y que los jugadores requieran estrategias de juego que puedan desarrollar con los conceptos adquiridos en materias posteriores. Son aplicables

también diversas estrategias, como los algoritmos evolutivos (Ashlock, 2006) o los algoritmos genéticos (Chisholm & Bradbeer, 1997). Esta actividad requiere del análisis del estado del arte actual en la materia, estudiando publicaciones que se hayan realizado al respecto, y experiencias llevadas a cabo por otros grupos de investigación.

2. Selección de las tecnologías de desarrollo: Un aspecto muy importante es que deben seleccionarse tecnologías de desarrollo que sean ampliamente dominadas por todos los alumnos para que dicha elección no resulte en una complejidad adicional. Para ello deben considerarse todas las tecnologías que los alumnos conozcan en profundidad por haberlas aprendido en el transcurso de su carrera (no necesariamente las que hayan adquirido en estas asignaturas específicas).

3. Definición del protocolo de comunicación entre el motor y los jugadores: El protocolo de comunicación seleccionado también influye en esta decisión. Si se decide que todos los grupos utilicen el mismo lenguaje de programación, podría seleccionarse Java por ser el más estudiado en la actualidad en nuestras materias. Si se decide que los jugadores se comuniquen con el motor por medio de la entrada y salida estándar, cada equipo puede elegir prácticamente cualquier lenguaje de programación moderno.

4. Desarrollo del motor: Esta tarea requiere trabajo en equipo y dominio acabado de una plataforma de desarrollo que incluya capacidades gráficas. Aquí se aplicarán algoritmos y estructuras de datos básicas y avanzadas, así como elementos de diseño y análisis de algoritmos. El motor puede plantearse como un componente evaluador y controlador de reglas del

juego. De esta forma, el desarrollo termina enfocándose en un servidor que sea capaz de correr el motor, incorporando agentes a modo de jugadores.

5. Desarrollo de los jugadores: Se requiere el conocimiento y el dominio de elementos de Inteligencia Artificial aplicados al contexto específico de situaciones de juegos y presencia de adversarios. En el contexto de lo expuesto en el punto anterior, los jugadores pueden desarrollarse como componentes que ejecutan las estrategias impuestas por el motor.

6. Ejecución de las competencias: Aquí se integran los desarrollos de motores, escenarios y agentes o jugadores, para llevar a la práctica el juego o competencia diseñado.

7. Emisión de resultados y reportes: Se miden los resultados obtenidos en cuanto a logro de objetivos del juego desarrollado, el cumplimiento de las metas por parte de cada agente, el nivel de eficiencia alcanzado por cada uno (y por lo tanto de las estrategias programadas para ellos) y la posibilidad de realimentación y refinamiento de esas estrategias.

## Resultados y Objetivos

El objetivo general del proyecto es el *"desarrollo de un Motor o Servidor de Estrategias de Juego, que permita el planteo de Plataformas para el diseño de Juegos y Escenarios con Adversarios, con el fin de facilitar la integración de conocimientos y prácticas de diversas asignaturas de la Carrera de Ingeniería en Sistemas de Información, y posibilitar el desarrollo futuro de aplicaciones similares en contextos profesionales"*. A nivel de resultados esperados, se apunta a la posible integración de desarrollos y aplicaciones

prácticas entre asignaturas diversas de la carrera de Ingeniería en Sistemas de Información, pero es claramente extrapolable a otras situaciones y aplicaciones no académicas, tales como: simulaciones de estrategias de agentes adversarios en cualquier ámbito de aplicación de la Teoría de Juegos, desarrollo profesional de software para juegos, e incluso aplicaciones en la programación de software de control de componentes de hardware para robótica (software, motores de juego y escenarios para placas programables). En cuanto a posibles dificultades, un proyecto de esta naturaleza requiere considerar que existirán muchos alumnos y grupos trabajando desde cátedras distintas (aunque esto puede manejarse en parte mediante el planteo de estas actividades en forma opcional), además del tiempo de clase ocupado con los desarrollos. Si varias cátedras acuerdan un trabajo conjunto y colaborativo en esta línea de acción, debe preverse adecuadamente el tema en las planificaciones de todas ellas.

## Formación de Recursos Humanos

*Director del Proyecto:* Valerio Frittelli (Magister en Docencia Universitaria, Ingeniero en Sistemas de Información, Director de la Carrera de Licenciatura en Tecnología Educativa de la UTN Córdoba, Docente de grado y posgrado e investigador en la UTN Córdoba)

*Codirector:* Ana María Strub (Ingeniera en Sistemas de Información, titulada como Profesora Universitaria en Ingeniería por la Universidad Católica de Córdoba), docente e investigadora en la UTN Córdoba.

*Investigador Formado y Asesor:* Eduardo Destéfánis (Doctor en

Ingeniería por la UTN Córdoba, Ingeniero Electricista Electrónico, Investigador Categoría II en Programa de Incentivos, Profesor Titular Ordinario en la UTN Córdoba)

*Investigadores de Apoyo Categorizados:* Felipe Steffolani, Romina Teicher (ambos Ingenieros en Sistemas de Información, Categorizados E y F respectivamente en el programa interno UTN)

*Investigadores de Apoyo en Formación:* Marcela Tartabini, Gustavo Bett, Julieta Fernández y Diego Serrano (todos Ingenieros en Sistemas de Información y docentes en la UTN Córdoba)

Del objetivo del proyecto se desprende un fuerte trabajo en formación de recursos humanos. Los alumnos de las cátedras involucradas serán beneficiarios directos de la experiencia, así como los propios docentes e investigadores. Algunos de estos docentes se incorporan por primera vez a un proyecto de investigación, aportando sus conocimientos y prácticas, pero también buscando experiencia y antecedentes que les posibiliten el acceso futuro a la categorización como investigadores. En esta primera etapa, en que el proyecto se está planteando, todo el equipo se conforma con docentes de diversas cátedras. A medida que el trabajo avance, se prevé incorporar estudiantes de diversos años de cursado de la carrera. Esto permitirá contribuir a su formación como investigadores y a aportar elementos de aplicación profesional. Al diseñar e implementar el Servidor de Estrategias en una arquitectura de dos niveles de componentes (servidor de reglas y estrategias vs. agentes que ejecutan

reglas y estrategias) se contribuye además en la formación profesional de investigadores y alumnos, ya que el grupo estaría en condiciones de extender el alcance a otros proyectos orientados a la construcción de un servidor de aplicaciones amplio (que evalúe y controle reglas de negocio en general, y no solo aplicables al contexto de juegos, simulaciones e integración entre contenidos de asignaturas).

## Referencias

- [1.] Ashlock, D. (2006). Evolutionary Computation for Modeling and Optimization. New York: Springer.
- [2.] Chisholm, K., & Bradbeer, P. (1997). Machine Learning Using a Genetic Algorithm to Optimise a Draughts Program Board Evaluation Function. Proceedings of IEEE International Conference on Evolutionary Computation, (ICEC'97. Indianapolis: IEEE Conference Publications.
- [3.] Dignum, F., Bradshaw, J., Silverman, B., & van Doesburg, W. (2009). Agents for Games and Simulations: Trends in Techniques, Concepts and Design. Berlín: Springer.
- [4.] Salen, K., & Zimmerman, E. (2003). Rules of Play: Game Design Fundamentals. Cambridge: MIT Press.
- [5.] Schell, J. (2008). The Art of Game Design: A book of lenses. Burlington: Morgan Kaufmann.

## RemoteBot: una Aplicación que Combina Robots y Dispositivos Móviles

Claudia Queiruga, Claudia Banchoff Tzancoff, Fernando López  
LINTI/Facultad de Informática/Universidad Nacional de La Plata  
La Plata, B1900ASD, ARGENTINA  
{claudiaq,cbanchoff, flopez}@linti.unlp.edu.ar

### Resumen

**RemoteBot** es una aplicación cliente-servidor que permite controlar los robots del proyecto "Programando con Robots y Software Libre" mediante dispositivos móviles Android.

**RemoteBot** es una aplicación innovadora que combina robots con dispositivos móviles.

**RemoteBot** es el resultado de la articulación del proyecto de I+D "Programando con Robots y Software Libre" con la cátedra "Laboratorio de Software" de 4to. año de las carreras Lic. en Informática y Lic. en Sistemas de la Facultad de Informática de la UNLP.

El objetivo del proyecto "Programando con Robots y Software Libre" es acercar a docentes y estudiantes de escuelas secundarias a la programación, convencidos que las habilidades que se obtienen programando promueven el pensamiento analítico, sistemático, fomentan la creatividad y el trabajo colaborativo, todas ellas habilidades muy requeridas en la sociedad del siglo 21. En el mencionado proyecto, se utilizan robots fabricados por la empresa [RobotGroup](#) [8], basados en hardware libre para enseñar las primeras nociones de programación.

**RemoteBot** integra diferentes tecnologías y su aplicación concreta es en el campo de aplicaciones no tradicionales. A su vez, permitió articular un proyecto de I+D de la Facultad de Informática con una cátedra de años avanzados de la misma casa de altos estudios, resultando esto último sumamente motivante para los estudiantes que participaron de la actividad.

**Palabras clave:** Enseñar a Programar - Programar con robots - Python - JAVA - Dispositivos móviles.

### El proyecto "Programando con Robots y Software Libre"

Es común que cuando nos referimos a nuestros niños y jóvenes hablemos de "nativos digitales" haciendo alusión a su fluidez digital o habilidad para manejarse con las tecnologías digitales. En los niños y jóvenes de hoy está naturalizado el envío de mensajes de texto, los juegos en línea, navegar en Internet, etc. [1]. Sin embargo, a pesar de interactuar todo el tiempo con medios digitales, muy pocos de estos jóvenes pueden construir sus propios juegos, animaciones y simulaciones. La fluidez digital requiere también de la habilidad para diseñar, crear e inventar con los nuevos medios [2] y para ello, es necesario aprender a programar. Saber programar tiene múltiples beneficios, amplía las posibilidades de las cosas que podemos crear con la computadora y en general de las cosas que se pueden aprender.

"Programando con Robots y Software Libre" [4] es un proyecto de I+D de la Facultad de Informática de la Universidad Nacional de La Plata, iniciado en el año 2009, cuyo objetivo es promover en los jóvenes de nuestra región el interés en la programación, entendiendo que aprender a programar es el fundamento del "pensamiento computacional" [3]. Este proyecto está basado en la iniciativa [RobotEducation.org](#) presentada en el evento "Latinamerican Academic Summit 2008", organizado por Microsoft Research en la ciudad de Panamá. Este proyecto se lleva adelante realizando un trabajo articulado con docentes y alumnos de escuelas secundarias de nuestra región. Mediante el uso de robots sencillos, se presenta a los jóvenes estudiantes una nueva forma de resolver problemas,



mediante el desarrollo de algoritmos. El uso y manipulación de robots es una manera atractiva y simple de acercar a nuestros jóvenes estudiantes a la programación. Introducir a docentes y jóvenes al mundo de la programación no debe ser considerado sólo un aprendizaje técnico, es el fundamento para aprender estrategias de diseño y resolución de problemas como la modularización, el pensamiento analítico y sistémico y, el trabajo colaborativo, todos ellos valorados como habilidades muy requeridas en la sociedad de hoy.

En el proyecto “Programando con Robots y Software Libre” se trabaja con algoritmos escritos en el lenguaje Python, mediante los cuales se programa al robot y de esta manera moverlo, hacer que evite obstáculos, etc. Desde la perspectiva educativa, la característica más relevante es que los estudiantes aprenden los conceptos básicos de programación en forma intuitiva y lúdica, explorando instrucciones y sentencias del lenguaje para manipularlos, moverlos y darles órdenes para emitir sonidos, experimentando sus resultados en forma interactiva y mediante la observación directa del robot. Utilizando los robots, se pueden programar actividades artísticas (pintar / bailar), sociales (realizar obras de teatro) y lúdicas (carreras de obstáculos / batallas), y de esta manera fomentar la creatividad y el trabajo colaborativo entre los estudiantes.

El lenguaje Python es un lenguaje interpretado, característica que simplifica el proceso de programación en una comunidad con escasa experiencia en programación.

Por otro lado, nuestro grupo de investigación promueve el uso de software libre, es por ello que todas las herramientas que se utilizan en el proyecto cuentan con licencias libres y la documentación y materiales de difusión del proyecto [5] se publican y difunden bajo licencias Creative Commons[6].

## Nuestras Experiencias en Escuelas

Las primeras experiencias del proyecto se realizaron con robots importados, denominados scribblers[7] y, en el año 2011, después de trabajar en forma colaborativa con los técnicos de la empresa [RobotGroup](#) [8] se pudo construir un robot con características similares a los scribblers, pero fabricados con hardware libre [9] y disponibles para su adquisición en nuestro país. RobotGroup desarrolló el módulo para Python duinobot [10], publicado como software libre, que permitió obtener un comportamiento similar al de los scribblers originales.

Durante el año 2012, a través de un subsidio de la Fundación YPF y con el auspicio de la Dirección de Escuelas Técnicas de la Provincia de Buenos Aires, se trabajó con 10 (diez) escuelas técnicas de la provincia. En dicha experiencia se capacitaron a más de 140 docentes y 40 alumnos en cursos dictados en las diferentes escuelas intervinientes [11].

Otras experiencias llevadas a cabo en esta línea de trabajo, fueron la implementación de las denominadas “pasantías académicas” realizadas con los colegios preuniversitarios de la UNLP, Liceo “Víctor Mercante” y Nacional “Rafael Hernández” y, durante el año 2012 se incorporó la Escuela Media N° 3 de Los Hornos. A través de estas pasantías, jóvenes del último año de la escuela secundaria concurren a la Facultad a realizar actividades coordinadas en el marco de proyectos de extensión y, de esta manera comienzan su interacción con el ámbito universitario. En el caso de “Programando con Robots y Software Libre” participaron de la experiencia más de 30 alumnos.

## Robots y Dispositivos Móviles

Actualmente vivimos en un mundo en el que las comunicaciones digitales han modificado la manera en que las personas se comunican. La telefonía móvil tiene un rol central en este cambio ya que no solamente es

utilizada para “hablar” sino que también para capturar y reproducir videos, tomar fotos, jugar, consultar la agenda de trabajo, páginas de noticias, usar mapas dinámicos, etc. Los teléfonos celulares cada vez tienen más prestaciones, sus pantallas son de mayor precisión y tamaño, tienen cámaras fotográficas y de video incorporadas, reproducen música y cuentan con múltiples sensores, tienen la capacidad de estar siempre conectados (always on), representando esto un desafío para el desarrollo de nuevas aplicaciones que dejan de ser entidades aisladas que intercambian información a través de la interfaz de usuario. El desarrollo de aplicaciones innovadoras que permita integrar diferentes tecnologías y que involucre dispositivos móviles, es un tema sumamente actual y motivante para nuestros jóvenes estudiantes universitarios. Combinar robots con dispositivos móviles, resultó una propuesta interesante, que permitió articular el proyecto “Programando con Robots y Software Libre” con las actividades desarrolladas por los estudiantes en la cátedra “Laboratorio de Software” de cuarto año de las carreras Lic. en Informática y Lic. en Sistemas de la Facultad de Informática de la UNLP. En esta asignatura los estudiantes adquieren conocimientos específicos sobre la construcción de, aplicaciones orientadas a servicios, con acceso a bases de datos y aplicaciones nativas para dispositivos móviles inteligentes utilizando tecnologías JAVA [14]. Los estudiantes obtienen las habilidades necesarias para desarrollar un trabajo integrador que signifique la aplicación concreta de los conocimientos adquiridos hasta el momento en la carrera. En este marco, resultó interesante proponerle a los estudiantes que construyan una aplicación que permita controlar a los robots mediante un teléfono celular inteligente, pensando en el celular como un control remoto no convencional, que mediante movimientos emitiera órdenes simples al robot, por ejemplo

que avance, retroceda, gire en una dirección determinada, etc.

## RemoteBot

**RemoteBot** es una aplicación cliente servidor que permite controlar los robots del proyecto "Programando con Robots y Software Libre " mediante dispositivos móviles Android que funcionan como controles remotos.

La Figura 1, ilustra una instalación típica de **RemoteBot** en la que se pueden apreciar todas componentes intervinientes.



Figura 1- Instalación Típica de **RemoteBot**

La aplicación consta de dos componentes, un servidor escrito en Python (Remotebot) que se comunica con el robot utilizando el módulo para Python duinobot [10] y un cliente para dispositivos Android (Remotebot4Android) escrito en Java que envía al servidor las acciones a realizar por el robot y que recibe retroalimentación desde el servidor respecto de las acciones realizadas.

La comunicación entre el cliente y el servidor es a través de mensajes POST de HTTP por el puerto 8000 y se utiliza JSON [12] para codificar los mensajes. Se diseñó un protocolo de capa de aplicación que permite instanciar los robots, enviarles mensajes y recibir los resultados de ejecutar los métodos correspondientes. El protocolo se adapta sin modificaciones a cualquier extensión que se le pueda hacer a las clases del módulo duinobot.

En la Figura 2 se muestra la arquitectura de **RemoteBot**, las cajas verdes, (Remotebot4Android y Remotebot) constituyen los desarrollos realizados para esta solución.

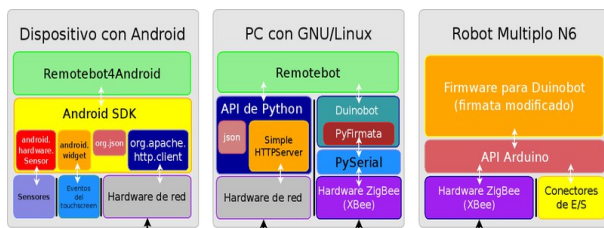


Figura 2- Arquitectura de **RemoteBot**

### El Servidor Remotebot

Considerando que el módulo original que controla los robots está escrito en Python, se decidió escribir **Remotebot** también en ese lenguaje. El servidor mantiene una colección de objetos que representan al dispositivo de comunicaciones (Board) y una colección de robots instanciados (Robot). Estas instancias nunca se liberan durante la ejecución del servidor, sin embargo esto no es necesariamente malo ya que cada instancia puede ser reutilizada por sucesivos clientes y estas colecciones sirven como una suerte de caché.

En el manejo de las peticiones se utiliza la técnica de reflexión para acceder a las funciones y a los métodos de los robots y placas. El uso de reflexión permite que el servidor siga funcionando sin modificaciones a pesar que se alteren, amplíen o reduzcan los métodos de las clases Robot y Board en el módulo duinobot.

Como se especificó anteriormente el servidor acepta peticiones utilizando el método POST de HTTP, la respuesta a ese POST contendrá los valores de retorno de los métodos invocados o bien un mensaje de excepción si algo falló.

### El Cliente Remotebot4Android

El cliente en Android contiene un *wrapper* completo de las clases Board y Robot que

puede ser reutilizada sin modificaciones en otros proyectos Android o bien con algunas modificaciones en la clase Board (agregando el paquete org.json y Apache HTTPComponents al proyecto) puede ser utilizada en aplicaciones Java regulares.

**Remotebot4Android** cuenta con una GUI compuesta por dos Activities (pantallas en el vocabulario de Android): una que permite configurar la conexión, seleccionando la dirección IP del servidor, el dispositivo que representa la placa y el robot al cual conectarse. En la Figura 3 se muestra una captura de pantalla de configuración. En determinadas ocasiones el módulo duinobot retornará una lista vacía de robots encendidos, en esos casos el cliente muestra una lista por defecto con los (supuestos) robots 1 a 6, luego de determinar esos parámetros se pasa al siguiente Activity (pantalla).

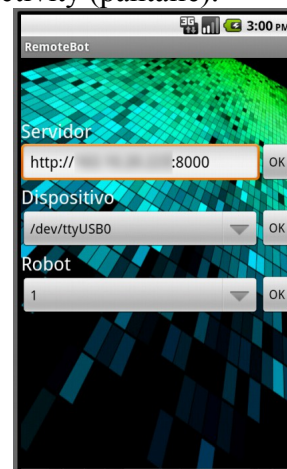


Figura 3- Cliente Android **Remotebot4Android**: configuración de la conexión

En el segundo Activity se encuentran los controles para manejar el robot que permiten: controlar la velocidad del robot (de 0 a 100), configurar el modo de avance (por ejemplo "avanzar sin chocar"), mostrar los valores del sensor de obstáculos del robot, girar a la mitad de la velocidad indicada, mover al robot hacia adelante, atrás, izquierda y derecha, detener al robot. La Figura 4 muestra una captura de la pantalla del cliente Android que permite manejar al robot.

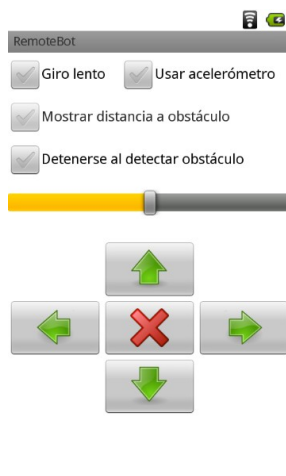


Figura 4- Cliente Android **Remotobot4Android**: controles para interactuar con el robot

Asimismo la interacción entre el cliente móvil y el robot puede realizarse usando los acelerómetros del celular. Para ello desde la pantalla de controles de la aplicación cliente es posible elegir este modo interacción y de esta manera se puede mover el robot simplemente inclinando el celular en la dirección deseada; el nivel de inclinación determina la velocidad. Los movimientos en la GUI se hacen de forma asincrónica y no se espera la respuesta del servidor (incluso se ignoran algunos errores) todo esto es para que la interfaz responda de forma rápida y se ignoren problemas de conexión intermitentes, naturales en las conexiones inalámbricas, que de otra forma resultan muy molestos.

### Líneas de investigación y desarrollo

El LINTI, Laboratorio de Investigación en Nuevas Tecnologías Informáticas, tiene definida una línea de investigación sobre Software Libre, tanto en términos de desarrollo como de uso y difusión especialmente en el ámbito de escuelas. El proyecto Lihuen GNU/Linux [13] con más de 6 años de trabajo, nos da un marco de interacción con las escuelas. Asimismo otra línea de investigación en crecimiento es la de desarrollo de aplicaciones móviles.

El desarrollo de RemoteBot no sólo permite trabajar en aspectos de integración de

diferentes tecnologías de desarrollo, sino que sirve como un elemento disparador tanto para los jóvenes que participan de los distintos espacios del proyecto “Programando con Robots y Software Libre” como así también y, fundamentalmente para los estudiantes universitarios que realizan una práctica sobre un desarrollo no convencional, innovador, que integra múltiples tecnologías y dispositivos.

Como líneas de trabajo en esta temática podemos mencionar también la evaluación y análisis de frameworks abiertos para el desarrollo de juegos (que se complementa con las actividades mencionadas anteriormente) y el desarrollo de otras actividades con escuelas que también promueven el acercamiento a la programación entre docentes y jóvenes estudiantes secundarios.

### Resultados y Objetivos

Este trabajo ofrece una herramienta más para los alumnos secundarios que participan del proyecto. En una primera instancia, todos aquellos que poseen celulares con Android, podrán descargarse la aplicación RemoteBot e interactuar con los robots. Luego, se introducirán las herramientas y nociones de programación necesarias para que puedan hacerlo escribiendo sus propios programas en Python.

Un aspecto importante a destacar es el hecho de lograr la implementación de aplicaciones reales en el ámbito de las cátedras. El hecho de trabajar sobre requerimientos y usos reales le otorga a los estudiantes una motivación adicional.

### Formación de Recursos Humanos

Este tipo de trabajo promueve la articulación entre proyectos de I+D y cátedras, involucrando a los estudiantes en la búsqueda de soluciones a problemas reales, no convencionales, que integran múltiples dispositivos y tecnologías. Asimismo se



fomenta la participación de los estudiantes de las distintas carreras en proyectos de I+D de interés para la Facultad, estimulando la elaboración de tesinas de grado y trabajos finales de carrera.

## Referencias

- [1] Prensky, Marc. Digital Natives, Digital Immigrants. On the Horizon, MCB University Press, Vol. 9 No. 5, October 2001, 1-6.
- [2] Resnick, Mitchel. Sowing the Seeds for a More Creative Society. Learning and Leading with Technology, Dec.2007, 18–22.
- [3] Wing, Jeannette. Computational thinking. Communications of ACM. Vol 49 N° 3, Mar. 2006, 33–35.
- [4] <http://robots.linti.unlp.edu.ar>
- [5] Manual de Programación con Python y Robots: [http://robots.linti.unlp.edu.ar/material\\_disponible](http://robots.linti.unlp.edu.ar/material_disponible)
- [6] Licencias Creative Commons: <http://www.creativecommons.org.ar/licencias>
- [7] Especificación del scribbler: [http://wiki.roboteducation.org/Myro\\_Hardware](http://wiki.roboteducation.org/Myro_Hardware)
- [8] <http://robotgroup.com.ar/>
- [9] [http://www.ecured.cu/index.php/Hardware\\_libre](http://www.ecured.cu/index.php/Hardware_libre)
- [10] Repositorio de código: <http://repo.lihuen.linti.unlp.edu.ar/lihuen/pool/lihuen4/main/r/robot/>
- [11] Aprendiendo a programar con juegos y robots. Javier Díaz, Claudia Banchoff, Sofia Martín, Fernando López. Aceptado y presentado en TEyET 2012. Junio 2012.
- [12] <http://www.json.org/>
- [13] Proyecto Lihuen GNU/Linux: <http://lihuen.info.unlp.edu.ar>
- [13] Aplicaciones para Dispositivos Móviles-Laboratorio de Software: <http://wiki.labmovil.linti.unlp.edu.ar>

## Propuesta de un portal de replicación de experimentos - Análisis de la personalidad en los equipos en el desarrollo de software

Alicia Mon<sup>1</sup>, Diego Fontdevilla<sup>1</sup>, Marcelo Estayno<sup>2</sup>, Diego Serra<sup>2</sup>,  
Marisa Panizzi<sup>3</sup>, Mauro De Mitri<sup>3</sup>, Nicolas Rivero<sup>3</sup>

<sup>1</sup>Departamento de Ingeniería Mecánica e Industrial, Facultad de Ingeniería, Universidad Nacional de Lomas de Zamora  
Ruta 4 – Km. 2 – Lomas de Zamora (CP 1832)  
Tel: 4282-7880/3454

<sup>2</sup>Departamento de Ingeniería e Investigaciones Tecnológicas. Facultad de Ingeniería, Universidad Nacional de La Matanza.  
Florencio Varela 1903 - San Justo (CP 1754)  
Tel: 4480-8900

<sup>3</sup>Facultad de Informática, Ciencias de la Comunicación y Técnicas Especiales, Universidad de Morón  
Cabildo 134 – Morón (CP 1708)  
Tel: 5627-2000

[alicialmon@gmail.com](mailto:alicialmon@gmail.com), [dfontde@gmail.com](mailto:dfontde@gmail.com), [mestayno@gmail.com](mailto:mestayno@gmail.com), [diego.g.serra@gmail.com](mailto:diego.g.serra@gmail.com)  
[marisapanizzi@speedy.com.ar](mailto:marisapanizzi@speedy.com.ar), [dmmauro@hotmail.com](mailto:dmmauro@hotmail.com), [nicorivero@hotmail.com](mailto:nicorivero@hotmail.com)

**Resumen:** El propósito de este trabajo pretende mejorar las replications de experimentos para agilizar la solución de problemas sociales a los investigadores en Ingeniería de Software. Se realizará mediante un portal web, el cual brindará soporte a las replications, permitirá el almacenamiento de materiales experimentales como así también el almacenamiento masivo de datos experimentales. En esta primera propuesta se trabajará con las replications del cuasi-experimento de factores de personalidad y equipos de desarrollo de software. Este modelo de portal puede tomarse como base para dar soporte a las replications de cualquier tipo de experimento dentro de la Ingeniería de Software Experimental (ESE). En esta primera etapa de su construcción se ha trabajado fuertemente en la definición de los requerimientos, el modelo de análisis y una definición a priori de la arquitectura del portal.

**Palabras clave:** *Experimentación en Ingeniería de Software (ESE)/ Portal de replications / Análisis de influencia de la personalidad/ Equipos de desarrollo*

### Contexto.

El presente proyecto será desarrollado por el Grupo GIS, integrado por investigadores de la Universidad Nacional de La Matanza y de la Universidad Nacional de Lomas de Zamora. Asimismo se trabajará en forma conjunta con grupos de investigación consolidados en el área en la Facultad de Informática de la Universidad Politécnica de Madrid, así como de la Universidad ORT de Uruguay.

En colaboración con uno de los objetivos propuestos en el proyecto, se incorporan alumnos de la Facultad de Informática, Ciencias de la Comunicación y Técnicas Especiales de la Universidad de Morón, para el desarrollo del portal de replicación de experimentos.

### Introducción.

Es sabido que no es fácil mantener una armonía cuando se habla de grupos de trabajo. Mucho menos cuando hablamos de grupos de trabajos de desarrollo de software, donde el trabajo en equipo es una herramienta clave para que el desarrollo sea

eficiente. La importancia de esta falta de armonía radica en las dificultades existentes para lograr un buen trabajo en equipo cuando sus integrantes tienen distintas características o personalidades.

*¿Por qué analizar los equipos de desarrollo de software?*

La respuesta a esta pregunta surge de la necesidad de relevar y analizar el comportamiento de los distintos equipos de trabajo en el desarrollo de software, más particularmente la influencia de la personalidad y las características de los distintos equipos. Las dificultades existentes para lograr un buen trabajo en equipo atentan directamente a la calidad del software desarrollado.

De los antecedentes bibliográficos revisados, podemos mencionar la definición de experimento propuesta por Campbell y Stanley, un experimento es: “una porción de la investigación en la cual se manipulan variables y se observan sus efectos sobre otras variables” (Campbell, 1963).

Estas variables pueden ser de dos tipos: *independientes* y *dependientes*. Las *variables independientes* son aquellas que inicialmente se consideran implicadas en llevar a cabo la consecución de un determinado efecto sobre las variables dependientes. Las *variables dependientes* son aquellas sobre las que se espera encontrar cambios significativos en su valoración debido a la influencia de las *variables independientes*.

La principal ventaja de la experimentación es que permite variar iterativamente aspectos de la realidad para estudiar el impacto que tienen las manipulaciones.

Estos experimentos se pueden clasificar en 3 tipos: *experimentos verdaderos*, *pre-experimentos* y *diseños cuasi-experimentales*.

Los *experimentos verdaderos* son los que manipulan las *variables independientes* para ver sus efectos sobre las *variables dependientes*. Para obtener evidencia de esta supuesta relación causal, el investigador manipula o hace variar la *variable independiente* y observa si la dependiente varía o no. La *variable dependiente* no se manipula, sino que se mide para ver el efecto que la manipulación de la *variable independiente* tiene en ella.

Los *pre-experimentos* se llaman así porque su grado de control es mínimo. No son adecuados para el establecimiento de relaciones entre la *variable independiente* y la *variable dependiente*. Se deberían de usar sólo como ensayos de otros experimentos con mayor control (Hernández Sampieri, 2006).

Por último, los *diseños cuasi-experimentales* son los que serían utilizados en este proyecto en particular. Estos manipulan deliberadamente al menos una *variable independiente* para ver su efecto y relación con una o más *variables dependientes*, solamente que difieren de los *experimentos verdaderos* en el grado de seguridad o confiabilidad que pueda tenerse sobre la equivalencia inicial de los grupos.

Los *cuasi-experimentos* difieren de los experimentos verdaderos en la equivalencia inicial de los grupos (los primeros trabajan con grupos intactos y los segundos utilizan un método para hacer equivalentes a los grupos). Sin embargo, esto no quiere decir que sea imposible tener un caso de *cuasi-experimento* donde los grupos sean equiparables en las variables relevantes para el estudio (Hernández Sampieri, 2006).

La ejecución de estos experimentos por parte de los investigadores se denominará *replicación*. La *replicación* consiste en la ejecución del experimento en un sitio concreto, utilizando un diseño experimental específico. En las *replicaciones* participan

personas (investigadores) que interactúan con materiales para observar el comportamiento de las técnicas estudiadas (Hernández Sampieri, 2006).

El *paquete de laboratorio* contiene las instrucciones para que los sujetos apliquen las técnicas y el investigador pueda analizar los resultados.

*¿Por qué desarrollar un portal de replicaciones de experimentos?*

Como resultado de la inexistencia de herramientas tecnológicas relacionadas con las replicaciones en la experimentación en ingeniería de software, se llegó a la conclusión que sería de gran ayuda el desarrollo de un portal que brinde soporte a la ejecución de los experimentos requeridos y permita administrar el material necesario para los mismos.

**Líneas de investigación y desarrollo.**

El portal propone brindar herramientas a los investigadores para que puedan realizar las replicaciones. Deberá ser capaz de registrar las replicaciones deseadas, permitiendo descargar todos los materiales necesarios (cuestionarios, etc.). Dicho sistema deberá disponer de una interfaz amigable y simple, de manera que los investigadores puedan acceder a toda la información de manera eficaz. El mismo será desarrollado siguiendo las buenas prácticas de la Ingeniería Web propuestas por Pressman (Pressman, 2005).

La especificación de los requerimientos de software se realizó siguiendo los lineamientos de la ERS-IEEE ANSI 830-1998 (IEEE ANSI 830-1998, 1998).

El portal contará con tres módulos bien definidos:

- *Administración de replicaciones*

- *Administración de materiales*
- *Módulo de seguridad,*

El *módulo de seguridad* permitirá al usuario autenticarse y acceder solamente a su propia información.

Los otros dos módulos permitirán la creación, modificación y eliminación de replicaciones y materiales.

Considerando los conceptos brindados anteriormente y ante la necesidad de desarrollar un sistema de fácil acceso que permita simplificar la administración de los experimentos, se determinó que lo más apropiado sería bajo entorno Web.

La particularidad de estos sitios webs, además de ser de muy fácil acceso y muy intuitivos, es que se pueden acceder a los mismos desde cualquier lugar distante, lo que en este caso brindaría una disponibilidad para cualquier grupo de investigadores de cualquier universidad que desee trabajar en esta línea de investigación.

La funcionalidad que deberá tener el mismo es la siguiente:

- Ingreso seguro: el mismo deberá ser capaz de validar los usuarios que podrán acceder al mismo.
- Administración de replicaciones: deberá ser capaz de registrar las replicaciones con toda la información relacionada a la misma.
- Administración de materiales: deberá ser capaz de almacenar todos los materiales disponibles para poder ejecutar las replicaciones.

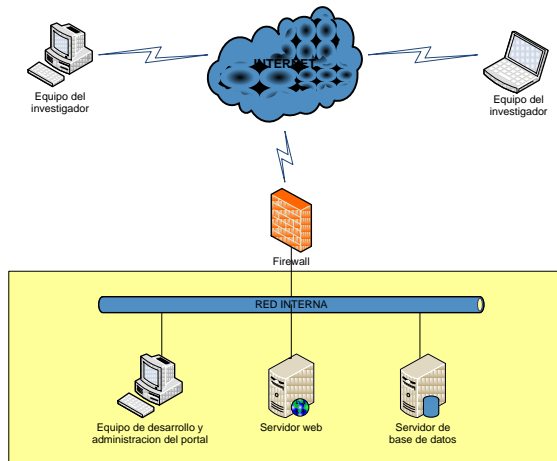
Dentro de los *requerimientos no funcionales*, podemos mencionar los



requerimientos de hardware necesarios para el funcionamiento del portal, estos son:

- *Equipo cliente con navegador web (Ej.: Internet Explorer, Firefox, Mozilla, etc.)*
- *Servidor web*
- *Servidor de base de datos*

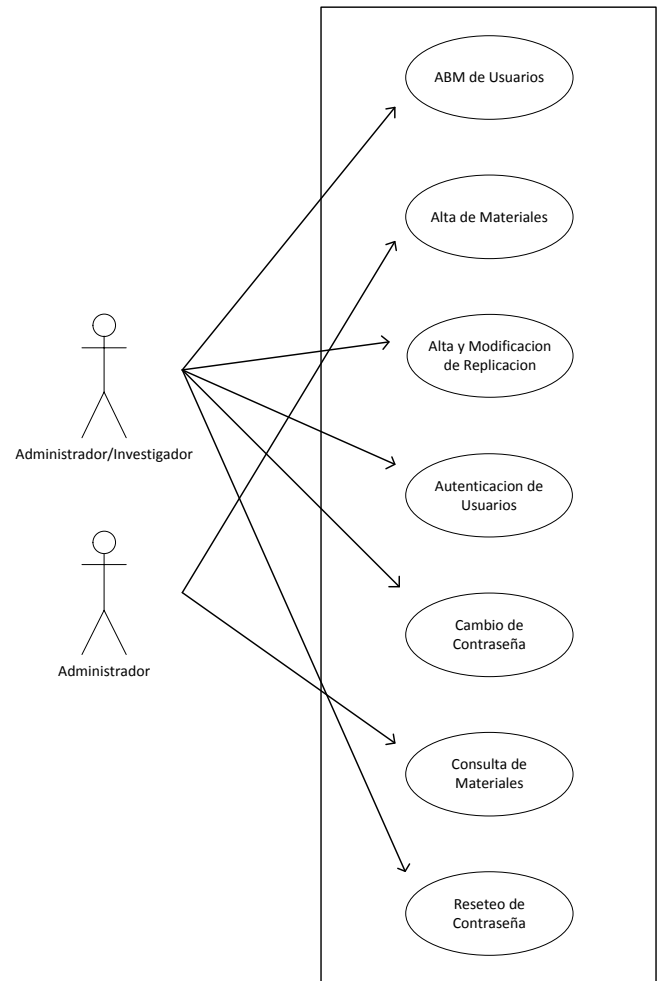
La **Figura Nro. 1**, muestra cómo se accedería al portal y cuáles serán los componentes de hardware que intervendrían en el funcionamiento.



**Figura Nro. 1.** Modalidad de acceso al portal y componentes de hardware

Para la construcción de los modelos de análisis y diseño se empleará RUP y UML (Jacobson Ivar , 2000).

En la **Figura Nro.2**, se detalla un caso de uso preliminar para la visualización del comportamiento del portal:



**Figura Nro. 2.** Caso de uso preliminar para la visualización del comportamiento del portal.

**Resultados Obtenidos/ Esperados.**

Se pretende finalizar con el desarrollo del portal y realizar las pruebas de dicho producto con investigadores de diferentes universidades.

Se intenta dejar como antecedente este portal para que sirva de guía para otro tipo de replications que se pretendan hacer en el marco de la Ingeniería de Software Experimental.

Se pretende sumar más Universidades para el trabajo en Red en esta línea de investigación con el propósito de generar mayor cantidad de evidencias empíricas de este tipo de experimentos.

Como resultados se plantea la incorporación de esta línea de investigación en trabajos de fin de carrera.

### **Formación de Recursos Humanos.**

La formación de recursos humanos está dada por la incorporación de investigadores y alumnos en el campo de la Ingeniería de Software Experimental.

La transferencia a la docencia será desarrollada por los investigadores pertenecientes a cada Universidad, que realizan repeticiones en el contexto de sus cátedras o en contextos de la industria del software.

### **Bibliografía.**

Booch Grady, M. R. (1998). *Object Oriented Analysis and Design With Applications* (2a edición ed.). Addison-Wesley.

Campbell, D.T., Stanley, (1963). *J.C.: Experimental and Quasi-Experimental Designs for Research*. Houghton Mifflin Company.

Craig Larman. (2005). *Applying UML and Patterns: An Introduction to Object-Oriented Analysis and Design and Iterative Development*. Pearson Education.

Davis, A. M. (1995). *201 Principles of Software Development*. New York, NY, USA: McGraw-Hill Inc.

Hernández Sampieri Roberto, Fernández Collado Carlos y Baptista Lucio Pilar. (2006). *Metodología de la investigación* (Cuarta ed.). Mexico: Mc Graw Hill.

IEEE ANSI 830-1998. (1998). *IEEE Recommended Practice for Software Requirements Specifications*. USA: Institute of Electrical and Electronics Engineers.

Itzik Ben-Gan, Lubor Kollar, Dejan Sarka, Steve Kass (2009). *Inside Microsoft SQL*

*Server 2008: T-SQL Querying*. Microsoft Press.

Jacobson Ivar , Booch Grady, Rumbaugh James . (2000). *El proceso Unificado de Desarrollo de Software*. Pearson - Addison Wesley.

John Sharp (2007). *Microsoft Visual C# 2008 Step by Step*. Microsoft Press.

Pfleeger, S. L. (2002). *Ingeniería de Software. Teoría y Práctica*. Bs As: Prentice Hall.

Pressman, Roger; (2005). *Ingeniería de Software, Un Enfoque Práctico*. Mc Graw-Hill.

Sommerville, I. (2007). *Ingeniería de Software*. España: Pearson Addison Wesley.

Tanenbaum Andrews S. (2003). *Redes de computadoras*. Editorial PEARSON ADDISON-WESLEY.

Will Panek, James Chellis: *Windows Server 2008 Active Directory Configuration* (2008).

## Accesibilidad web como medida de calidad en el marco del proyecto “Sistemas y TIC: técnicas y herramientas”

Sonia I. Mariño<sup>1,2</sup>, María V. Godoy<sup>1,2</sup>, Pedro Alfonzo<sup>1,2</sup>

<sup>1</sup>Área de Ingeniería Web. Facultad de Ciencias Exactas y Naturales y Agrimensura  
<sup>2</sup>Departamento de Informática. Facultad de Ciencias Exactas y Naturales y Agrimensura.  
9 de Julio 1449. CP: 3400. Corrientes. Argentina.

Universidad Nacional del Nordeste.

simarinio@yahoo.com, mvgodoy@exa.unne.edu.ar, plalfonzo@hotmail.com

### CONTEXTO

En el trabajo se exponen los resultados obtenidos en el proyecto “Sistemas y TIC: técnicas y herramientas” (2012-2016), como continuación de los alcanzados en el proyecto “Tecnologías de la información y comunicación como herramientas del desarrollo local” (2009-2012), ambos acreditados por la Secretaria General de Ciencia y Técnica (UNNE).

### RESUMEN

Se resume la línea de investigación vinculada con el estudio y análisis de accesibilidad, sus métodos y herramientas y la aplicación de los mismos para la evaluación de sitios web enfatizando los representativos de organizaciones del NEA orientados a diversos dominios del conocimiento. Se enfatiza la formación de recursos humanos en la temática, incorporando los estudios realizados en el desarrollo de becas otorgadas por la UNNE y tesinas de grado.

**Palabras clave:** accesibilidad web, técnicas y herramientas, formación de recursos humanos.

### 1. INTRODUCCIÓN

En el marco del proyecto de I+D “Sistemas y TIC: técnicas y herramientas” se continúa tratando el tema de accesibilidad iniciado en el anterior “Tecnologías de la información y comunicación como herramientas del desarrollo local”. Su eje principal es el estudio de las metodologías y herramientas orientadas a la generación de software de

calidad en coincidencia con lo expuesto por la Ingeniería del Software (IS).

En la IS existen tres elementos clave: i) los métodos, ii) las herramientas y iii) los procedimientos. Estos facilitan el control del proceso de construcción de software y brinda a los desarrolladores las bases de la calidad de una forma productiva. Siendo una de sus principales áreas de estudio e investigación la calidad del software. La misma se define como “Grado con el que un sistema, componente o proceso cumple los requerimientos especificados y las necesidades o expectativas del cliente o usuario” (IEEE Std. 610-1990). Atendiendo esta definición, donde el objetivo es la satisfacción del usuario y la importancia otorgada a los organismos internacionales para desarrollar y evaluar la calidad de un producto software, se considera que el mismo, no puede estar exenta de la aplicación de estándares, siendo uno de los referentes la accesibilidad.

En este sentido, existen organismos no oficiales como la W3C (Consortio World Wide Web), a través de la Iniciativa para la Accesibilidad a la Web o WAI (Web Accessibility Initiative) (Oficina Española, 2008), que definen pautas que facilitan la inclusión social y el acceso universal a la información.

Como normativa oficial sobre accesibilidad Web, se mencionan: i) ISO 9241-171:2008 (Accesibilidad al Software (ISO, 2008 a), ii) ISO 9241-20:2008 (Accesibilidad en productos y servicios TIC (ISO, 2008 b) y iii)

ISO 9241-151:2008 (Ergonomía de interfaces Web (ISO, 2008 c).

La iniciativa desarrollada por este equipo de trabajo de la UNNE, coincide con las desarrolladas por otros equipos universitarios como los expuestos en Díaz, (2008), Díaz (2011), Díaz et al. (2012a), Díaz et al. (2012b), Martín et al. (2012), Mendez y Caraldi (2012), Toledo et al. (2012) y Trigueros et al. (2012).

## 2. LINEAS DE INVESTIGACION y DESARROLLO

En la línea de I+D, con mira a la transferencia de sistemas accesibles se trabaja en:

- El relevamiento, selección y estudio de metodologías (González Flórez, 2006; Montero y Fernández, 2004; Segovia, 2008;) para el tratamiento de la accesibilidad web.
- La elección, análisis y estudio de herramientas informáticas utilizadas para la medición de accesibilidad, entre las que se mencionan: Brailersurf, Lynx Browser, TAW, HERA, EXAMINATOR, TextAloud, NVDA, WEBBIE, entre otras.

## 3. RESULTADOS OBTENIDOS/ESPERADOS

En esta sección se mencionan los logros alcanzados en la línea temática especificada y vinculados al mencionado proyecto que facilita la concreción de actividades de investigación aplicada, desarrollo y transferencia hacia el contexto de influencia de la UNNE

Los resultados se plasmaron en presentaciones de congresos y reuniones científicas de carácter nacional (Mariño et al., 2012a; Fernández Vázquez et al., 2012), una publicación internacional (Mariño et al., 2012b) y generación de productos software transferibles al contexto de influencia de ésta Universidad que cumplen las pautas estudiadas.

Uno de los propósitos del proyecto de I+D es la divulgación del conocimiento y experiencia adquirida en ámbitos de Educación Superior a fin de instalar estos temas y su aplicación en la comunidad académica. En el segundo semestre del año 2012 se desarrolló un taller avalado por la FACENA (UNNE) destinado a alumnos, adscriptos y docentes.

## 4. FORMACIÓN DE RECURSOS HUMANOS

En referencia a la formación de RRHH se mencionan como logros del año 2012

- Continuaron el estudio de metodologías y herramientas de accesibilidad y su introducción en diseños de sistemas de información becarios de grado y postgrado de la SGCyT – UNNE
- Recientes graduados de la carrera incorporados al proyecto iniciaron la formación de alumnos desempeñándose como profesores orientadores del Trabajo Final de Aplicación en esta temática. Se resalta este logro considerando que es una forma de incrementar la masa crítica comprometida con la docencia e investigación
- Iniciaron el Trabajo Final de Aplicación, en la mencionada temática, cuatro alumnos de carrera Licenciatura en Sistemas de Información: Vergara (2012), desarrolla un sistema de gestión para la empresa Distribuciones Litoral y Maidana (2012), elabora un Sistema Informático para la gestión de un club deportivo. Solís (2012), aborda el estudio de la accesibilidad y su inclusión en módulos del sistema SIU. En los proyectos de TFA se abordó como tema de calidad la accesibilidad incluida desde etapas tempranas del ciclo de vida, es decir se incorporó este concepto desde la etapa de análisis y diseño del sistema de información. Los temas abordaron la selección y estudio de herramientas, la aplicación de pautas de accesibilidad en sitios web de diferentes dominios y sus resultados serán difundidos en el presente año.



- Se desarrolló un taller gratuito denominado “Accesibilidad en sistemas de información web”, avalado por la Unidad Académica, cuyos asistentes superaron en un 33% el cupo máximo previamente establecido.

## Referencias

- Brailersurf, Disponible en:  
<http://www.snv.jussieu.fr/inova>
- Consortio World Wide Web (W3C),  
Disponible en: <http://www.w3c.es/>
- Díaz, J. 2008. Responsabilidad social universitaria como mecanismo de inclusión digital: experiencia de la Universidad Nacional de La Plata, XIII Congreso Internacional del CLAD sobre la Reforma del Estado y de la Administración Pública, Buenos Aires, Argentina, 4 - 7 nov. 2008
- Díaz, F. J., Banchoff, C. M., Harari, I., Osorio, M. A., Amadeo, A. P. (2011). Accesibilidad Web en la Práctica. Iniciativas Académicas en Informática. Anales Simposio sobre la Sociedad de la Información 2011. 40° Jornadas Argentinas de Informática. Argentina.
- Díaz, F., Schiavoni, A., Amadeo, A. y Charnelli, M. (2012a). ”Construyendo Objetos de Aprendizaje utilizando estándares abiertos y sistemas open source. Una experiencia sobre un curso de Accesibilidad Web”. Congreso de Tecnología en Educación y Educación en Tecnología. TEYET 2012.
- Díaz, F. Harari, I. y Amadeo, P. (2012b). “Propuesta sobre Aprender Enseñando: desarrollo de un curso a distancia sobre Accesibilidad Web en manos de alumnos”. Congreso de Tecnología en Educación y Educación en Tecnología. TEYET 2012.
- Godoy, M. V. y Mariño, S. I. (2010). “El Área de Ingeniería WEB de la FACENA – UNNE.” Área “Innovación en Sistemas de Software”. Anales Workshop de Investigadores en Ciencias de la Computación. WICC 2010 (Argentina). 513-516pp.
- González Flórez, J. A. (2006). Pautas de accesibilidad en la web para bibliotecas. Alfagrama Ediciones.
- Fernández Vázquez, A., Acevedo, J. J., Mariño, S. I., Godoy, M. V., Alfonzo, P. L. (2012). "Comunicación y accesibilidad en sitios web municipales de la región del Nordeste Argentino, su evaluación mediante validadores automáticos", Revista Question, N° 35. ISSN 1669-6581.
- IEEE STD 610-1990. IEEE Standard Glossary of Software Engineering Terminology.
- ISO (2008 a). International Standards for Business, Government and Society (ISO). 9241-171:2008. Ergonomics of human-system interaction -- Part 171: Guidance on software accessibility. 2008.
- ISO (2008 b). International Standards for Business, Government and Society (ISO). ISO 9241-20:2008, Ergonomics of human-system interaction -- Part 20: Accessibility guidelines or information/communication technology (ICT) equipment and services ISO 9241-20, 2008.
- ISO (2008 c). International Standards for Business, Government and Society (ISO). ISO 9241-151:2008, Ergonomics of human-system interaction -- Part 151: Guidance on World Wide Web user interfaces, 2008.
- Lynx Browser. Disponible en: <http://invisible-island.net/lynx/>
- Maidana, A. B. (2012). “Sistema de Gestión de club deportivo”. Trabajo Final de Aplicación. Cerrera Licenciatura en Sistemas de Información. Prof. Orientador: Pedro Alfonzo.
- Mariño, S., Godoy, M., Acevedo, J., Alfonzo, P., Gómez Solis, L., Fernández Vazquez, A. (2012a). “Estudio de accesibilidad web en el marco del proyecto “Tecnologías de la información y comunicación como herramientas del desarrollo local””.

- Anales Workshop de Investigadores en Ciencias de la Computación. WICC 2012 (Argentina). 389-391pp.
- Mariño, S. I., Godoy, M. V., Alfonzo, P. L., Acevedo, J. J., Gomez Solis, L., Fernández Vázquez, A. (2012b). "Accesibilidad en la definición de requerimientos no funcionales. Revisión de herramientas", *Multiciencias*, vol 12, Nº 3. ISSN 1317-2255.
- Martín, A.; Gaetán, G.; Saldaño, V.; Miranda, G.; Molina, S.; Pastrana, S. (2012). "Diseño y Evaluación tempranos para priorizar la Accesibilidad en la WWW". *Anales Workshop de Investigadores en Ciencias de la Computación. WICC 2012 (Argentina)*.
- Méndez, P. y Cataldi, Z. (2012). Inclusividad en los campus virtuales: condiciones de accesibilidad y usabilidad. *Revista Quaderns Digitals*. 73. 4
- Montero, Y. H.; Fernández, F. J. (2004). *Revista española de documentación científica*. Vol. 27, Nº 3, 2004, págs. 330-344. Disponible en: [http://www.nosolousabilidad.com/hassan/DCU\\_accesible.pdf](http://www.nosolousabilidad.com/hassan/DCU_accesible.pdf)
- NVDA. Lector de pantalla. Disponible en: <http://nvda.softonic.com/descargar>
- Oficina Española (2008) *Word Wide Web- - Guía Breve de Accesibilidad Web*, Disponible en: <http://www.w3c.es/divulgacion/guiasbreves/accesibilidad>
- Segovia, C. (2008). *Accesibilidad e Internet...para que todas las personas, con distintas capacidades o recursos, puedan acceder a Internet*. Disponible en: [http://www.archena.es/files/accesibilidad\\_e\\_internet.pdf](http://www.archena.es/files/accesibilidad_e_internet.pdf)
- Solis, M. A. (2012). "Aplicación de conceptos de integración persona ordenador al diseño de interfaz para la gestión de expedientes de la Facultad de Derecho y Ciencias Sociales y Políticas". Trabajo Final de Aplicación. Cerrera Licenciatura en Sistemas de Información. Prof. Orientador: Juan J. Acevedo y Laura Gomez Solis.
- TAW. Test de Accesibilidad Web "TAW". Disponible en: <http://www.tawdis.net/>
- TextAloud. Disponible en: [http://www.nextup.com/files/manualtrams/Spanish\\_TextAloudManual.html](http://www.nextup.com/files/manualtrams/Spanish_TextAloudManual.html)
- Toledo, G., González, A. y Malbrán, M. (2012). "Accesibilidad digital para usuarios con limitaciones visuales". *Congreso de Tecnología en Educación y Educación en Tecnología. TEYET 2012*.
- Trigueros, A., Giulianelli, D., Rodríguez, R., Vera P., Fernández, V. (2012). "Sitio Web Móvil Universitario – Priorizando la Accesibilidad". *Anales Workshop de Investigadores en Ciencias de la Computación. WICC 2012 (Argentina)*.
- Vergara, M. (2012). "Sistema de gestión para la empresa Distribuciones Litoral". Trabajo Final de Aplicación. Cerrera Licenciatura en Sistemas de Información. Prof. Orientador: Pedro Alfonzo.
- WEBBIE. Explorador web para usuarios ciegos y deficientes visuales. Disponible en: <http://www.webbie.org.uk/es/index.htm>
- Web Developer Extensión. Disponible en: <https://addons.mozilla.org/es-ES/firefox/addon/60/>

# Caracterización de los riesgos inherentes a la Ingeniería Reversa

## Herramientas para la Ingeniería Reversa

Inés Gimenez Zens, Juan Carlos Cuevas, Franco Mana

Universidad Tecnológica Nacional

Facultad Regional Córdoba

Departamento de Ingeniería de Sistemas de Información

Grupo de Investigación en Ingeniería Reversa

Maestro López Esq. Cruz Roja Argentina – Ciudad Universitaria  
C.P 5016 – Córdoba – Argentina

### RESUMEN

Se presenta en este documento el resultado de la línea de investigación, “Herramientas de Ingeniería Reversa”, en el marco del proyecto “Caracterización de los riesgos inherentes a la Ingeniería Reversa”. Este proyecto lo desarrolla el grupo de investigación de la Cátedra de Servicios del Software de la Universidad Tecnológica Nacional Córdoba. Se describe aquí el proceso de investigación y el desarrollo de un gráfico que tiene como objetivo informar las posibles herramientas de ingeniería de software a utilizar según las variables del proyecto que se haya encarado.

### PALABRAS CLAVES

Ingeniería reversa, Herramientas de Ingeniería reversa, Ingeniería de Software.

### CONTEXTO

La Cátedra de Servicios del Software de la Universidad Tecnológica Nacional Córdoba desarrolla desde el año 2008 investigaciones tendientes a arribar a un Framework para la Gestión de la Calidad en Ingeniería Reversa. Como primer proyecto de investigación se presentó “Caracterización de los riesgos inherentes a la Ingeniería Reversa” ante la Secretaría de Ciencia y Tecnología de la UTN, el cual fue acreditado por la mencionada Secretaría y se encuentra en su fase final de desarrollo. Este proyecto tiene, entre sus objetivos aportar a una

gestión más eficiente del software en aspectos tales como funcionalidad, confiabilidad, usabilidad, mantenibilidad, portabilidad, seguridad, costos y satisfacción de los clientes-usuarios.

### INTRODUCCION

En la actualidad el hombre tiene una alta dependencia de los sistemas de software para la realización de todo tipo de tareas. Estos sistemas permiten automatizar procesos, simplificar tareas y un sin fin de objetivos que mejoran la calidad de vida de las personas. Esta dependencia junto con las necesidades empresariales y sumadas a los avances tecnológicos, han ocasionado que constantemente se necesiten crear o adaptar sistemas de software. Si el sistema de software es creado sin tener un precedente, se debe realizar un análisis profundo de los requerimientos para que este cumpla con el objetivo y resuelva las necesidades. En el caso de que el sistema sea heredado<sup>1</sup>, o que se necesite un nuevo sistema pero con un precedente, la tarea a realizar es la investigación de la información existente que permita comprender y posteriormente actuar de

---

<sup>1</sup> Un sistema heredado (o sistema legacy) es un sistema informático (equipos informáticos o aplicaciones) que ha quedado anticuado pero continúa siendo utilizado por el usuario (típicamente una organización o empresa) y no se quiere o no se puede reemplazar o actualizar de forma sencilla.

acuerdo a los objetivos prefijados. Este último caso se denomina ingeniería reversa de software.

La ingeniería reversa de software se enmarca dentro del mantenimiento de software y es definida por Chikofsky y Cross como el proceso de analizar un sistema para crear una representación del mismo, pero a un nivel más elevado de abstracción<sup>2</sup>. Esto se logra a través de procesos de compresión y no a través de la modificación del sistema. La recopilación de la información es fundamental para elevar los niveles de conocimiento del programa y es una tarea que debe realizar el ingeniero de software a través la identificación de artefactos, el descubrimiento de relaciones y la generación de abstracciones. La recuperación del entendimiento del sistema no es tarea fácil, se deben tener habilidades específicas para realizar este trabajo y conocimiento sobre herramientas de ingeniería reversa específicas que permitan simplificar el trabajo.

Las herramientas clásicas de ingeniería reversa son:

**El Decompilador:** es un programa que realiza la operación inversa a un compilador. Esto es, traducir código o información de bajo nivel de abstracción (sólo diseñado para ser leído por un ordenador, por ejemplo el código máquina) a un lenguaje o medio de mayor nivel de abstracción.

**El depurador (en inglés, debugger):** es un programa usado para probar y depurar (eliminar los errores) de otros programas (el programa "objetivo").

**El desensamblador:** es un programa que traduce el lenguaje de máquina a lenguaje ensamblador, la operación inversa de la que hace el ensamblador.

Las herramientas CASE (Computer Aided Software Engineering, Ingeniería de Software Asistida por Computadora): son diversas aplicaciones informáticas destinadas a la recuperación de los modelos en los que fue originalmente desarrollado el software heredado, en el caso de la ingeniería reversa. Debemos destacar que es muy escasa la información y bibliografía que el ingeniero de software dispone

<sup>2</sup> La abstracción, en la ingeniería de software, es entendida como las diferentes cualidades aisladas del sistema de software heredado que permiten mejorar la información para comprenderlo más fácilmente.

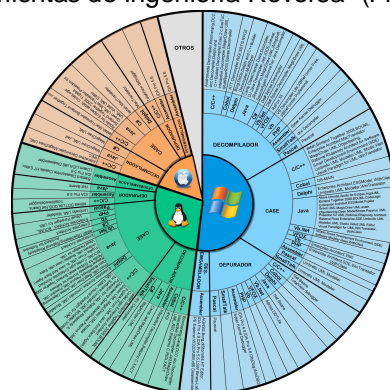
para hacer uso de una en herramienta en particular. Esto se debe, en gran medida, a que la elección de la herramienta dependerá de diversas variables que complejizan la búsqueda. Las principales variables limitantes en el uso de una herramienta de ingeniería reversa son el sistema operativo, el tipo de herramienta y su salida, el tipo de lenguaje de programación, y la funcionalidad de la herramienta, entre otros. Cuando el ingeniero de software necesita una herramienta en particular, la definición de estas variables en que opera limitará significativamente el universo de posibilidades. Por ejemplo, un caso común podría ser la búsqueda de un decompilador que funcione para el lenguaje Java y que se opere bajo la plataforma Windows. Pero muy distinta será la misma búsqueda de herramientas bajo la plataforma MacOS, en donde el universo es significativamente menor.

## LINEAS DE INVESTIGACION Y DESARROLLO

En el marco del proyecto "Caracterización de los riesgos inherentes a la Ingeniería Reversa" se abrió una línea de investigación en el tema "Herramientas para la ingeniería reversa" con la intención de aportar el conocimiento necesario sobre la temática para poder avanzar hacia la definición de los riesgos en la ingeniería reversa.

## RESULTADOS Y OBJETIVOS

Como resultado de este relevamiento se obtuvo una lista de herramientas que fueron categorizadas y puestas a disposición de un diseñador que transformó esta lista en un gráfico denominado "Herramientas de Ingeniería Reversa" (Fig.1).



- Fig.1 Herramientas de Ingeniería Reversa

Este gráfico (expuesto en sus reales dimensiones en el anexo 1) presenta la información de una



manera que el ingeniero de software puede acceder rápidamente, seleccionando las variables limitantes, a un listado de los nombres de las herramientas de ingeniería reversa disponibles. Esta información minimiza el esfuerzo de investigación de herramientas específicas necesarias al encarar un proyecto de ingeniería reversa de un software heredado. El gráfico aporta datos concretos según el sistema operativo, el tipo de herramienta, y lo que es fundamental, el tipo de lenguaje de programación.

## **FORMACION DE RECURSOS HUMANOS**

El proyecto de investigación está destinado a contribuir a la formación de recursos humanos como docentes de la Cátedra, Docentes de otras asignaturas, Estudiantes, Graduados, Especializandos, Mastrandos y Doctorandos y Profesionales de la Industria del Software.

### **Transferencia de conocimiento**

La especialización adquirida en el estado del arte de la temática objeto de estudio como así también el eventual nuevo conocimiento generado, está destinado a ser transferido, en primera instancia a los estudiante de la cátedra de Servicios del Software y como segundo paso a los distintos estamentos académicos, gobierno, industria y sociedad.

### **Metodología**

La base de ejecución del programa de investigación es el método científico. La metodología propuesta para el abordaje de dicho programa se fundamenta en desarrollar una serie de actividades que se basan en el método que permite relacionar el estudio y evaluación de las relaciones entre los cuerpos teóricos disponibles y la evidencia empírica de los fenómenos estudiados, como por ejemplo: los estudios de índole exploratorios, descriptivos y explicativos, los principios, métodos y herramientas estadísticas y teorías probabilísticas, diseños experimentales, etc.

Para el caso particular de la investigación en las herramientas para la ingeniería reversa, se analizaron los tipos de herramientas disponibles en la web y se categorizaron. Con esta información se

lanzó un relevamiento de un año de duración en busca de las herramientas para las plataformas Windows, Linux y MacOS. Los resultados fueron tabulados por tipo de lenguaje de programación y luego se diseñó el gráfico para simplificar las búsquedas.

### **Impacto del Proyecto**

El proyecto apunta a reducir la brecha entre el amplio conocimiento existente para el desarrollo del software y el escaso conocimiento para el mantenimiento y la ingeniería reversa. Para esto el gráfico presentado "Herramientas de ingeniería reversa" aporta la información necesaria para que el ingeniero de software pueda reducir significativamente el esfuerzo en la búsqueda de herramientas que funcionen en entornos particulares, facilitando así la tarea de ingeniería reversa.

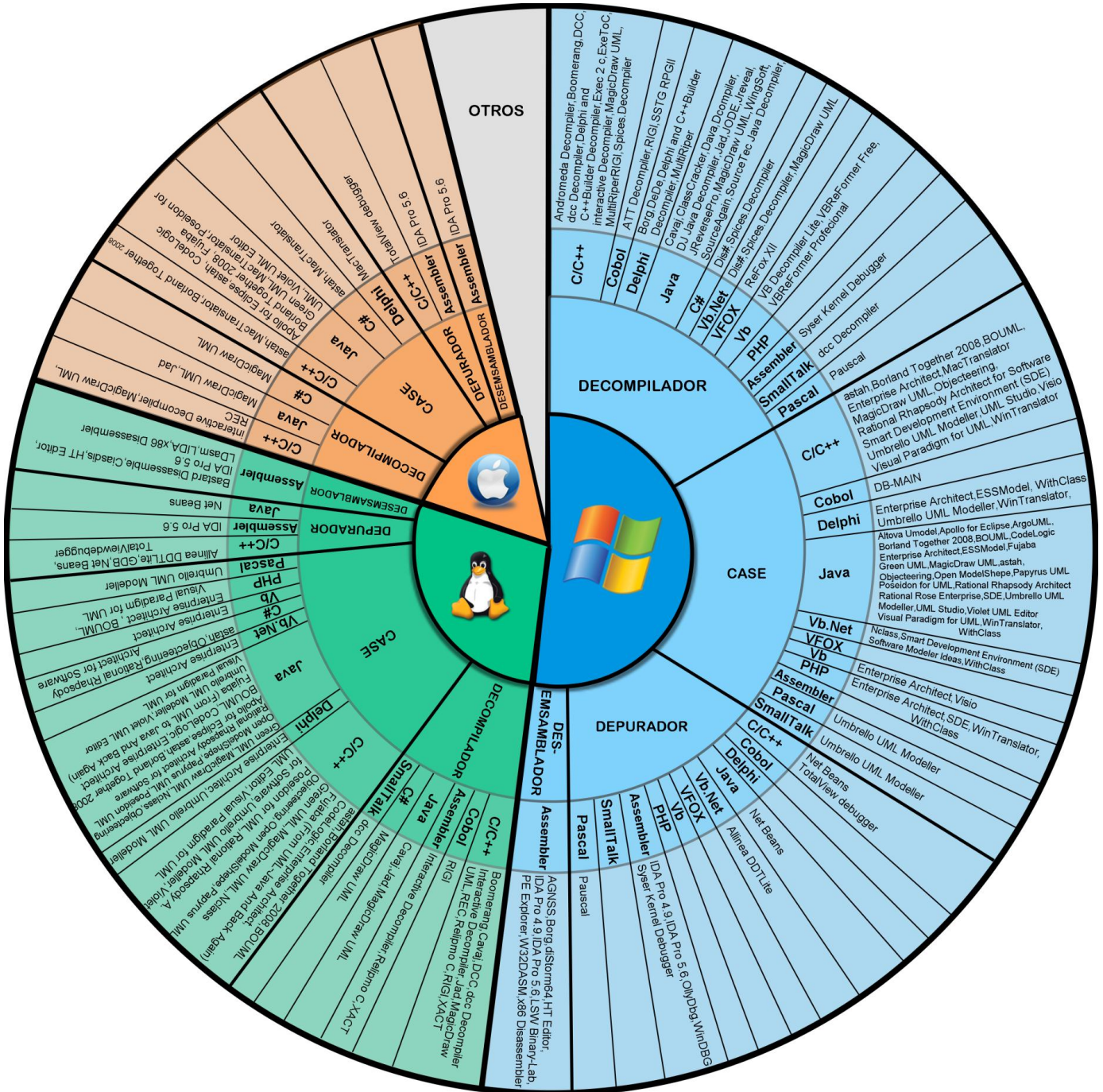
**BIBLIOGRAFIA**

Convergence Information Technology - 0-7695-3038-9/07 - IEEE - Computer Society - 2007.

1. [Chikofsky:1990] Chikofsky, E. and Cross, J. (1990). Reverse engineering and design recovery: A Taxonomy. IEEE Software.
2. [Jarzabek:2007] Stanislaw Jarzabek; "Effective Software Maintenance and Evolution. A Reuse-Based Approach". Auerbach Publications. NY - EEUU - 2007.
3. [Grubb-Takang:2005] Penny Grubb and Armstrong A. Takang; "Software Maintenance - Concepts and Practice" Second edition. World Scientific - Singapore - 2005.
4. [Tonella-Potrich:2005] Paolo Tonella and Alessandra Potrich; "Reverse Engineering of Object Oriented Code". Springer Science + Business Media, Inc. - Boston - EEUU - 2005
5. [Eilam:2005] Eldad Eilam "Reversing - Secrets of Reverse Engineering". Wiley Publishing, Inc - IN - EEUU - 2005.
6. [Davis:2005] Kathi Hogshead Davis " August-11: A Tool for Step- by-Step Data Model Reverse Engineering" IEEE - 0-8186-7111-4/95 - 1995
7. [Gannod-Cheng:1996] Gerald C. Gannod and Betty H. C. Cheng " Using Informal and Formal Techniques for the Reverse Engineering of C Programs" - 0-8186-7674-4/9 - IEEE - 1996.
8. [Bellay-Gall: 1997] Bemdt Bellay and Harald Gall; "A Comparison of four Reverse Engineering Tools" 0-8186-8162497 - IEEE - 1997.
9. [Davis:2001] Kathi Hogshead Davis; "Lessons Learned in Data Reverse Engineering". 1095-1350/01 - IEEE - 2001.
10. [Jahnke-Valenstein:2000] Jens H. Jahnke and Andrew Walenstein "Reverse Engineering Tools as Media for Imperfect Knowledge" 1095-1350/00 - IEEE - 2000.
11. [Tilley:1998] Scott Tilley; "A Reverse-Engineering Environment Framework". TECHNICAL REPORT
12. CMU/SEI-98-TR-005. Carnegie Mellon University - Software Engineering Institute - PA - EEUU. 1998.
13. [Zhou et al:1999] Shikun Zhou, Hongji Yang, Paul Luker and Xudong He; "A Useful Approach to Developing Reverse Engineering Metrics". 0-7695-0368-3/199 - IEEE - 1999.
14. [Chiang-Barron:1995] Roger H. L. Chiang and Terence M. Barron "Quality Issues in Database Reverse Engineering: An Overview". 95 ENGINEERING MANAGEMENT CONFERENCE - 0-7803-2799-3/95/ - IEEE - 1995.
15. [Garcia et al:2004] Vinicius C. Garcia, Daniel Lucr'edio, Antonio F. do Prado, Alexandre Alvaro and Eduardo S. de Almeida "Towards an effective approach for Reverse Engineering". Proceedings of the 11th Working Conference on Reverse Engineering (WCRE'04) - 1095-1350/04 - IEEE - Computer Society - 2004.
16. [D'Ambros-Lanza:2006] Marco D'Ambros and Michele Lanza "Reverse Engineering with Logical Coupling". Proceedings of the 13th Working Conference on Reverse Engineering (WCRE'06) - 0-7695-2719-1/06 - IEEE - Computer Society - 2006.
17. [Fahmi-Choi:2007] Syed Ahsan Fahmi and Ho-Jin Choi; "Software Reverse Engineering to Requirements". 2007 International Conference on

ANEXO I

Gráfico - Herramientas de Ingeniería Reversa





# Optimización de la gestión del avalúo fiscal en la provincia de San Luis

Erica Daniela Bonilla<sup>1</sup>, Germán Montejano<sup>2</sup>, Ana Funes<sup>2</sup>

<sup>1</sup>Dirección de Catastro y Tierras Fiscales – Gobierno de la Provincia de San Luis  
Pedernera e Ituzaingó  
5700 San Luis, Argentina

ericadanielabonilla@gmail.com

<sup>2</sup>Universidad Nacional de San Luis, Ejército de los Andes 950  
5700 San Luis, Argentina  
{gmonte, afunes}@unsl.edu.ar

## Resumen

El avalúo de las tierras rurales, con fines catastrales, se fundamenta en la observación de las características y cualidades inherentes a la tierra misma, los recursos naturales, la infraestructura y la disponibilidad de los servicios de equipamiento, tales como vías de comunicación, acueductos y electricidad, componentes que contribuyen a la conformación de los precios del mercado inmobiliario.

Para la valuación territorial, la Provincia de San Luis conformó 82 zonas económicas, cuyos límites y valor fiscal fue determinado por una comisión de expertos, basando dicho avalúo en la correspondencia de cada parcela con las 82 zonas predefinidas. La aplicación de esta zonificación presentó inconvenientes, generando reclamos por parte de propietarios en relación al límite de las zonas. La falta de actualización de los valores de referencia de cada zona, una distorsión en el valor fiscal y una desactualización del valor del impuesto inmobiliario rural con valores muy por debajo de lo real fueron otros de los inconvenientes que surgieron de dicha forma de avalúo.

La decisión de actualizar el avalúo fiscal de las tierras rurales de la provincia de San Luis, sugirió la necesidad de revisión del

método utilizado y de ofrecer la posibilidad de establecer un nuevo método que utilice las posibilidades tecnológicas actuales.

Bajo tales premisas, en esta línea de investigación, nos hemos avocado al desarrollo de un modelo para el cálculo del avalúo fiscal, basándonos en el uso de un sistema de información geográfica [1] para lograr la estandarización de los datos alfanuméricos y cartográficos que permitan vincular un inmueble con la disponibilidad de recursos naturales e infraestructura pública, ofreciendo mejoras en la objetividad, precisión y facilidad de actualización, y resultando, a su vez, en una mayor equidad de la carga impositiva.

**Palabras clave:** avalúo fiscal, sistemas de información geográfica, SIG, metadatos, zonificación.

## Contexto

El presente trabajo de investigación se encuentra enmarcado en una colaboración entre investigadores del Proyecto de Incentivos código 22/F822 “Ingeniería de Software: Conceptos, Métodos y Herramientas en un contexto de Ingeniería de Software en Evolución”, de la Universidad Nacional de San Luis y la Dirección de Catastro y Tierras Fiscales de la Provincia de San Luis.

Esta propuesta de investigación se desarrolla como una tesis de la Maestría en



“Ingeniería de Software” de la Universidad Nacional de San Luis.

## Introducción

Desde el año 1973, a partir de la Ley N° 3583, para la valuación territorial, la provincia de San Luis se conformó en 54 zonas geoeconómicas. Con el transcurso del tiempo surgieron nuevas legislaciones hasta llegar a la actual conformación consistente en 82 zonas económicas relativamente homogéneas, cuyos límites y valor fiscal fue determinado por una comisión de expertos públicos y privados. A su vez, la valuación fiscal fue basada en la correspondencia de cada parcela con las 82 zonas predefinidas en la Resolución Ministerial 1180 HyOP-SEF95 del año 1995.

Esto trajo aparejado que en una misma zona coexistieran parcelas rurales con distintas características de servicios y disponibilidad de recursos, lo que implica un desequilibrio entre las mismas.

Asimismo, la falta de actualización de los valores de referencia de cada zona produjo una distorsión en el valor fiscal y una desactualización del valor del impuesto inmobiliario rural con valores muy por debajo de lo real.

Por otro lado, el uso de la información espacial es cada vez más amplio en el manejo y la administración del medio ambiente y recursos naturales, en la prevención y mitigación de desastres, en el ordenamiento del territorio, planeamiento de infraestructuras (transporte, servicios básicos de saneamiento, edificaciones, etc.), defensa y seguridad ciudadana, y muchos otros temas de vital importancia tales como la toma de decisiones a nivel gobierno para promover el desarrollo sostenible y la mejora del bienestar social y económico de un territorio.

La utilización de los Sistemas de Información Geográfica (SIG) y la estandarización de los datos alfanuméricos

y cartográficos permiten vincular un inmueble con la disponibilidad de recursos naturales e infraestructura, ofreciendo mejoras en la objetividad, precisión y facilidad de actualización, y resultando, a su vez, en una mayor equidad de la carga impositiva.

La constante actualización de la información de las tierras rurales nos sugirió la necesidad de revisión del método hasta entonces utilizado para el cálculo del avalúo, pretendiéndose de esta manera, disponer de un nuevo modelo, el cual disolviera el análisis zonal y lograra el análisis por parcela, utilizando como herramienta para el cálculo un Sistema de Información Geográfica para el análisis espacial de las variables consideradas [2].

Al mismo tiempo, en base a experiencias realizadas, pudimos constatar que se presentan algunas dificultades a la hora de usar esta información geográfica, sobre todo referidas a la escasez, dificultad de localización, desconocimiento de sus características, calidad, elevado costo de generación y en especial a los obstáculos para su integración en otros sistemas de información diferentes del organismo que los genera. Asimismo, es sabido que, a medida que crecen el número, la complejidad y diversidad de los datos geográficos, también crece la necesidad de disponer de un mecanismo que facilite la comprensión de todas las características de dichos datos. En este sentido, una documentación apropiada proporciona a todos un conocimiento claro y un mejor manejo en la producción, almacenamiento, actualización y reutilización de sus datos.

Teniendo en cuenta lo anteriormente expuesto, y debido a que las variables que creemos relevantes para el cálculo han sido generadas por diversos organismos provinciales y nacionales, resulta necesario revisar y analizar el estándar ISO 19115 “Geographic Information - Metadata” [3] y definir un perfil de metadatos mínimos que permita describir totalmente los geodatos utilizados en nuestro nuevo modelo de

avalúo, pudiendo los usuarios, de esta manera, entender las presunciones y limitaciones así como evaluar la calidad y aplicabilidad de los datos para el uso específico de su interés, a la vez de proporcionar a los productores de datos geográficos la información apropiada para que ellos caractericen sus datos de manera que faciliten su descubrimiento, recuperación y reutilización e intercambio.

## Resultados y Objetivos

Como resultado de esta investigación nos planteamos obtener un modelo integrado por un método de cálculo para el avalúo fiscal [4] [5] [6] [7], implementando un Sistema de Información Geográfica que permita el análisis espacial de las variables involucradas. Como parte del modelo también nos propusimos incluir un Perfil de metadatos basado en la norma ISO 19115 “Geographic Information - Metadata” que permita describir de forma completa y precisa los geodatos utilizados [8] [9] [10] [11].

### Objetivos generales

- Definir un modelo para el cálculo del avalúo fiscal de inmuebles. basado en un conjunto de variables o atributos geográficos.
- Implementar el modelo en un Sistema de Información Geográfica que permita el análisis espacial de las variables involucradas en el modelo.
- Proponer nueva legislación que reglamente la aplicación del método de avalúo propuesto.
- Definir un perfil de metadatos de acuerdo a la norma ISO 19115 [3] para documentar los datos geográficos involucrados en el cálculo del avalúo, a través de estándares que permitan, de manera unificada, obtener información

acerca del estado de la calidad y el origen de los datos existentes [12] [13] [14] [15].

### Objetivos específicos

- Lograr mayor confiabilidad y precisión en el cálculo del avalúo de inmuebles rurales.
- Proporcionar un método que permita una valuación masiva y sistematizada.
- Establecer un modelo flexible que permita representar la heterogeneidad de los atributos que hacen al valor de las parcelas y que sea capaz de facilitar la actualización y parametrización de las variables, así como la modificación de los criterios de valuación considerados.
- Lograr mayor equidad en el avalúo.
- Proporcionar, a los productores de información geográfica (organismos de gobierno, municipios, universidades, empresas), las herramientas necesarias para describir correctamente sus datos y facilitarles, a su vez, el mantenimiento y la organización de los mismos.
- Facilitar el descubrimiento, la recuperación y la reutilización de los datos.
- Facilitar la transferencia e intercambio de datos y su interpretación.
- Evitar la duplicidad de esfuerzos en la producción de información.
- Distribuir información confiable.

## Líneas de Investigación y Desarrollo

La línea de investigación aquí presentada, la cual está siendo llevada adelante entre investigadores del proyecto de incentivos de la Universidad Nacional de San Luis, código 22/F822: "Ingeniería de Software: Conceptos, Métodos y Herramientas en un Contexto de Ingeniería de Software en Evolución" y la Dirección de Catastro y Tierras Fiscales de la Provincia de San Luis, se trata de una nueva línea surgida de la ejecución de una tesis de Maestría en Ingeniería de Software de la Universidad Nacional de San Luis, y de la cual hemos presentado aquí los lineamientos generales para la elaboración de un modelo integrado que optimice la gestión del avalúo fiscal.

## Formación de Recursos Humanos

El presente trabajo es la base para el desarrollo de una tesis de posgrado de la carrera "Maestría en Ingeniería de Software" de la Universidad Nacional de San Luis. La misma está siendo llevada adelante por la Analista de Sistemas Erica Bonilla, alumna de la Universidad Nacional de San Luis, y siendo dirigida por investigadores de la Universidad Nacional de San Luis, Dr. Germán Montejano y Magister Ana Funes. El trabajo de tesis se encuentra en la etapa final de escritura del informe, habiéndose cumplido ya casi por completo con los objetivos planteados para este trabajo de investigación.

## Referencias

[1] Bosques Sendra, Joaquín. (2000) Sistemas de Información Geográfica. Segunda edición corregida. Editorial RIALP S.A. Madrid.

[2] Buzai, Gustavo y Baxendale, Claudia (2006). Análisis socioespacial con Sistemas de Información Geográfica. Editorial Lugar, Buenos Aires.

[3] International Organization for Standardization (2003). "International Standard ISO19115. Geographic information – Metadata". <http://www.iso.org>. Fecha de último acceso: 10/12/2012

[4] Alcázar Molina, M y Ariza López, F. (2004). Situación actual de la valoración catastral rústica en España: propuesta de un modelo alternativo. CT: Catastro, ISSN 1138-3488, N° 52, 2004, págs. 85-112. <http://dialnet.unirioja.es/servlet/articulo?codigo=1064193>

[5] CANO GUERVÓS, R. A., CHICA OLMO, J. M. Y HERMOSO GUTIÉRREZ, J. A. (2002). "A Geo-Statistical method to define districts within a city". Journal of Real Estate Finance and Economics, 27, 1, 61-85.

[6] CABALLER y MOYA (1997): "Valoración de empresas españolas", Ed. Pirámide.

[7] GARCÍA, A. (2004): "Redes Neuronales Una aplicación a la predicción del IBEX-35" Ponencia publicada en "Novedades en la teoría general de valoraciones. Aplicaciones", Universidad de Granada.

[8] IPGH, IGAC (2011). "Perfil Latinoamericano de Metadatos Geográficos". <http://www.igac.gov.co/igac>

[9] Perfil IDEC para conjunto de Datos. <http://www.geoportal-idec.cat>. Fecha de último acceso: 10/12/2012

[10] Subgrupo de Trabajo del Núcleo Español de Metadatos. Núcleo español de Metadatos (NEM). <http://www.ideo.es/>. Fecha de último acceso: 10/12/2012

[11] ICDE. «Lineamientos de Política Nacional de Información Geográfica y consolidación de la infraestructura colombiana de Datos Espaciales: Productos y Servicios Geográficos para el Desarrollo», Documento Pre-CONPES, Bogotá, Marzo de 2006.

[12] Dublin Core Metadata Initiative, 2003. "Using Dublin Core – The Elements". <http://dublincore.org/documents/usageguide/elements.shtml>

[13] Noguera, J.; Rodríguez, A.; Gould, M.; Zarazaga, F.J. y Muro, P.R.; 2004. "Recomendaciones de Metadatos para la IDEE, V 1.0. <http://ideo.unizar.es/>. Fecha de último acceso: 10/12/2012

[14] Metadata Ad Hoc Working Group. Document FGDC-STD-001-1998 "Content Standard for Digital Geospatial Metadata", Federal Geographic Data Committee (USA), 1998. <http://www.fgdc.gov/metadata/csdgm/>. Fecha de último acceso: 10/12/2012

[15] F. J. Zarazaga-Soria, J. Nogueras-Iso and M.Ford. Guidance material for mapping between Dublin Core and ISO in the Geographic Information domain. CWA 14856. CEN/ISSS Workshop - Metadata for Multimedia Information - Dublin Core. Sep. 2003.  
<http://www.cenorm.be/cenorm/businessdomains/businessdomains/iss/cwa/cwa14856.asp>. Fecha de último acceso: 10/12/2012



# GENERACIÓN DE SISTEMAS DE RECUPERACIÓN DE INFORMACIÓN PARA LA GESTIÓN DOCUMENTAL EN EL ÁREA DE LAS CIENCIAS DE LA COMPUTACIÓN

H. Kuna<sup>1</sup>, M. Rey<sup>1</sup>, J. Cortes<sup>1</sup>, E. Martini<sup>1</sup>, L. Solonezen<sup>1</sup>, R. Sueldo<sup>1</sup>, G. Pautsch<sup>1</sup>

1. Depto. de Informática, Facultad de Ciencias Exactas Quím. y Naturales Universidad Nacional de Misiones.

hdkuna@gmail.com , m.rey00@gmail.com

## RESUMEN

La búsqueda de información en bibliotecas digitales es una actividad en la cual los resultados a presentar al usuario, a partir de una consulta, deben responder a sus expectativas. Los Sistemas de Recuperación de Información buscan optimizar el proceso de búsqueda de contenido en la web a través de diversas herramientas, entre ellas los meta-busadores. Los mismos amplían el espectro de cobertura en la búsqueda, a partir de la capacidad para utilizar las bases de datos de varios buscadores en simultáneo; además de poder incorporar diversos métodos para el ordenamiento de los documentos, que mejoren la relevancia de los resultados para el usuario. En este trabajo se presenta el desarrollo de un Sistema de Recuperación de Información para la búsqueda de documentos científicos en el área de las ciencias de la computación, haciendo especial hincapié en el algoritmo de ranking utilizado para ordenar el listado final de resultados.

**Palabras clave:** recuperación de información, algoritmo de ranking, búsqueda web, indicadores bibliométricos.

## CONTEXTO

Esta línea de investigación articula el "Programa de Investigación en Computación" de la Facultad de Ciencias Exactas Químicas y Naturales de la Universidad Nacional de Misiones; el Grupo de Investigación Soft Management of Internet and Learning (SMILE) de la Universidad de Castilla-La Mancha, España; y el Departamento de

Bibliotecología de la Universidad Nacional de Mar del Plata, Argentina.

## 1 INTRODUCCION

### 1.1 Sistemas de Recuperación de Información

Un Sistema de Recuperación de Información (SRI) se puede definir como un proceso capaz de almacenar, recuperar y mantener información (Salton & McGill, 1983) (Kowalski, 1997). Existen en la literatura diversas propuestas sobre la estructura básica que debiera tener un SRI, un ejemplo es la de (Baeza-Yates & Ribeiro-Neto, 1999) que lo considera a partir de la unión de cuatro elementos como son: los documentos que forman parte de la colección sobre la que se realizará la recuperación; las consultas que representan las necesidades de información por parte de los usuarios; la forma en la que la modelan las representaciones de los documentos, consultas y las relaciones presentes entre ellos; y una función de evaluación que determina para cada consulta y documento el orden que ocupará en los resultados a presentar.

En la actualidad los principales modelos de SRI que operan sobre internet son: los directorios, los buscadores y los meta-busadores (Olivas, 2011). En el contexto de la presente investigación cobran mayor relevancia los meta-busadores, debido a que posibilitan la utilización de bases de datos de otros buscadores replicando las consultas de los usuarios sobre los mismos y procesando los resultados obtenidos de la manera que se crea conveniente para generar un listado único. Para ello, los meta-busadores deben contar con

algoritmos de evaluación, que son aplicados para la fusión de las listas de resultados obtenidas de cada buscador (Serrano-Guerrero, Romero, Olivas, & De la Mata, 2009).

## 1.2 Métricas para la Evaluación de Documentos Científicos

Dada la naturaleza del SRI a generar, los métodos para la evaluación de los resultados deben ser desarrollados en forma particular. Para la evaluación de documentos científicos se debe considerar una serie de características evaluables, como ser (Bollen, Van de Sompel, Hagberg, & Chute, 2009) (Pendlebury, 2009): el tipo de fuente de publicación, la calidad de los autores y del artículo en sí, medida a través de la cantidad de veces que haya sido citado. Para cada una de estas características existen métricas ampliamente aceptadas que pueden aplicarse. En el caso del tipo de fuente de publicación, para aquellas publicaciones realizadas en revistas existen dos índices que se utilizan para estimar su calidad: por un lado el Factor de Impacto (IF, por sus siglas en inglés) generado por la Web Of Knowledge<sup>1</sup> que administra el Institute for Scientific Information (ISI) (Garfield E, 2006) que es parte de la empresa Thomson Reuters; y el índice SJR, SCImago Journal Rank (Gonzalez-Pereira, Guerrero-Bote, & Moya-Anegón, 2009), producido utilizando la base de datos del buscador Scopus<sup>2</sup> de la editorial Elsevier y generado por el grupo de investigación SCImago de la Universidad de Extremadura, España. En ambos casos se trata de métricas que toman las citas que reciben los artículos publicados en una revista y las evalúan tanto en cantidad como en lo referente a la relevancia que tiene la producción que la realiza. Mientras que en caso de que la publicación se realice en un congreso o evento similar existe otro ranking como es

el que genera en la web Computer Research & Education (CORE)<sup>3</sup> en donde a diversas conferencias o congresos se los ubica en uno de los cuatro niveles que tiene establecidos: A+, A, B y C sin proveer mayores detalles sobre el método utilizado para realizar el cálculo.

Para estimar la calidad de la producción de un autor existen otras métricas como pueden ser: el índice H (Hirsch, 2005) y el índice G (Egghe, 2006); lo que hacen éstas es tomar la cantidad de citas recibidas por las diferentes publicaciones del autor y la cantidad de publicaciones para calcular un valor que representa la influencia del mismo. Finalmente, para evaluar la calidad de una colección de publicaciones a través del tiempo se puede utilizar un índice como es el AR (Jin, 2007), que toma la cantidad de citas obtenidas por las mismas y las pondera utilizando ese factor en combinación con la antigüedad de cada uno de los artículos que componen la colección.

## 2 LINEAS DE INVESTIGACION y DESARROLLO

Existen implementaciones de SRI en la web que utilizan diferentes métodos de búsqueda, pero no existen implementaciones de herramientas de este tipo que se apliquen específicamente a bases de datos de documentos científicos en el área de las ciencias de la computación, que además incorporen técnicas de inteligencia artificial o lógica difusa para la mejora de la relevancia de los resultados a presentar al usuario.

La generación de este tipo de Sistemas de Recuperación de Información requiere del desarrollo de diversos componentes, entre los cuales se destaca el algoritmo que se utiliza para establecer un ranking entre los resultados de la consulta a realizar.

<sup>1</sup> Web of Knowledge: <http://wokinfo.com>. Accedido: 25/02/2013.

<sup>2</sup> Scopus: <http://www.scopus.com>. Accedido: 25/02/2013.

<sup>3</sup> CORE: <http://www.core.edu.au>. Accedido: 25/02/2013.

### 3 RESULTADOS OBTENIDOS/ESPERADOS

#### 3.1 Grado de Avance

El presente proyecto ha comenzado sobre la mitad del año 2012. En una primera etapa se analizó el estado del arte, se estudiaron los distintos modelos de SRI y los elementos que deben componerlos. Conjuntamente se seleccionaron las tecnologías que se utilizaron en el desarrollo de un prototipo de meta-buscador, priorizando aquellas que fueran basadas en la filosofía Open Source, como ser: los lenguajes HTML, PHP y SQL, junto al motor de bases de datos MySQL, utilizando como entorno para su implementación al servidor web Apache.

Se han determinado los siguientes componentes a desarrollar del meta-buscador en una primera etapa:

- Módulo para la gestión de las consultas: encargado de adaptar las consultas efectuadas por el usuario para ser utilizadas posteriormente en los buscadores integrados.
- Módulos para la búsqueda en las bases de datos (buscadores): encargado de gestionar la realización de las consultas, adaptadas previamente, sobre los buscadores incorporados. A continuación captura los resultados y los prepara para el próximo componente.
- Módulo para la gestión de los resultados: en primera instancia realiza una pre-selección de los resultados, desechando aquellos que no son relevantes en cuanto a su tipo (informe de una cita sobre el documento, por ejemplo). Posteriormente se procede a la unificación de los resultados en un único listado para la aplicación del algoritmo de ranking.
- Módulo para la mejora de la relevancia de los resultados: en éste se aplican diversas técnicas para que los resultados para que sean de mayor utilidad al usuario.

#### 3.2 Desarrollo del Algoritmo de Ranking

El componente del meta-buscador que realiza el ordenamiento de los resultados obtenidos de las bases de datos, es de los más importantes. En este caso particular, al trabajar con artículos científicos, se requirió que el algoritmo de ranking evaluara las propiedades de los mismos a través de diversas métricas. Entre las características evaluables de los documentos científicos se seleccionaron: la fuente de publicación, la calidad de sus autores y del artículo en sí, medida por la relación entre la antigüedad del documento con la cantidad de veces que ha sido citado.

Para cada una de las propiedades seleccionadas, se determinaron las métricas a considerar para ponderar cada resultado:

- Para la fuente de publicación: se tomaron dos factores para la valoración de este punto, dependiendo si el artículo se publicó en una revista científica o en un congreso del área de conocimiento que corresponda. Para el primer caso, existen a nivel internacional dos métricas: el Factor de Impacto desarrollado por ISI y el índice SJR desarrollado por el grupo de investigación SCImago. En este caso se ha optado por utilizar al segundo ya que presenta diversas ventajas con respecto al primero, como ser (Falagas, Kouranos, Arencibia-Jorge, & Karageorgopoulos, 2008), (Leydesdorff, De Moya-Anegón, & Guerrero-Bote, 2010): es de acceso abierto; en la base de datos de Scopus contiene una mayor cantidad de revistas, incluyendo aquellas que no están escritas en inglés; no sólo hace una evaluación cuantitativa de las citas recibidas por un artículo sino que también lo hace en forma cualitativa, incorporando la calidad de la revista que genera la cita; entre otras. Para el caso de los artículos procedentes de congresos o reuniones científicas se empleó el ranking generado por la Computing Research and Education Association of Australia (CORE). El modelo de

clasificación de este ranking se transformó a un formato numérico para poder operar con él. El valor correspondiente a la fuente de publicación se obtiene mediante la fórmula 1, en caso de haber sido en una revista, y con la fórmula 2, en caso de haber sido en un congreso.

$$\text{fuentePublicacion} = \log_{10}(\text{SJR}) \quad (1)$$

$$\text{fuentePublicacion} = [A^* = 1; A = 0.75; B = 0.5; C = 0.25] \quad (2)$$

- Para la calidad de los autores: en este caso se usó como base el índice H propuesto por (Hirsch, 2005). El mismo representa la cantidad  $x$  de artículos de un autor que han recibido  $x$  citas como mínimo. En caso de que se trate de un artículo con más de un autor se considera la posición que ocupa en el mismo. El cálculo del valor que se utilizó se puede ver en la fórmula 3.

$$\text{autores} = \log_{10}(\sum(\text{indiceH}(\text{autor}_i)/i)) \quad (3)$$

- Antigüedad y citas: en la última métrica se ponderó la calidad de la publicación teniendo en cuenta la antigüedad y la cantidad de citas obtenidas por la misma. Para ello se utilizó como base el índice AR, en la fórmula 4 se puede observar el resultado de la adaptación realizada.

$$\text{calidadPublicacion} = \frac{\text{citasRecibidas}}{\text{antigüedadPublicacion}} \quad (4)$$

- Finalmente se ponderan los tres factores antes expuestos a través de tres constantes que se utilizaron para dar distinto peso a cada elemento, esto se refleja en la fórmula 5. Los valores planteados para  $\alpha$ ,  $\beta$  y  $\gamma$  fueron: 0.5, 0.3 y 0.2 respectivamente.

$$\text{valorFinal} = \alpha * \text{fuentePublicacion} + \beta * \text{autores} + \gamma * \text{calidadPublicacion} \quad (5)$$

Ese valor final fue el que se utilizó para realizar el orden de los resultados antes de presentarlos al usuario.

### 3.3 Trabajos Previstos en la Próxima Etapa

Para el año 2013 se tiene previsto:

- Completar el desarrollo de los componentes del meta-buscador.
- Analizar distintas técnicas inteligentes a incorporar en el meta-buscador para la mejora de la relevancia de los resultados a presentar al usuario final.
- Incorporar elementos dentro del SRI que permitan la expansión de la consulta original realizada por el usuario.
- Optimizar el algoritmo de ranking implementado para la gestión de los resultados.
- Incorporar técnicas de clustering y de lógica difusa para una mejor organización de los resultados.

## 4 FORMACION DE RECURSOS HUMANOS

Este proyecto es parte de las líneas de investigación del “Programa de Investigación en Computación” de la Facultad de Ciencias Exactas Químicas y Naturales de la UNaM, con siete integrantes (todos ellos alumnos, docentes y egresados de la carrera de Licenciatura en Sistemas de Información de la facultad de Ciencias Exactas Químicas y Naturales de la Universidad Nacional de Misiones) de los cuales cuatro están realizando su tesis de grado y dos están realizando un doctorado. Esta línea de investigación vincula al “Programa de Investigación en Computación” del Departamento de Informática de la Facultad de Ciencias Exactas Químicas y Naturales de la Universidad Nacional de Misiones, al Grupo de Investigación Soft Management of Internet and Learning (SMILe) de la Universidad de Castilla-La Mancha, España; y el Departamento de



Bibliotecología de la Universidad Nacional de Mar del Plata, Argentina.

## 5 BIBLIOGRAFIA

- Baeza-Yates, R., & Ribeiro-Neto, B. (1999). *Modern information retrieval* (Vol. 463). ACM press New York.
- Bollen, J., Van de Sompel, H., Hagberg, A., & Chute, R. (2009). A Principal Component Analysis of 39 Scientific Impact Measures. *PLoS ONE*, 4(6).
- Egghe, L. (2006). Theory and practise of the g-index. *Scientometrics*, 69(1), 131-152.
- Falagas, M. E., Kouranos, V. D., Arencibia-Jorge, R., & Karageorgopoulos, D. E. (2008). Comparison of SCImago journal rank indicator with journal impact factor. *The FASEB Journal*, 22(8), 2623-2628.
- Garfield E. (2006). The history and meaning of the journal impact factor. *JAMA*, 295(1), 90-93.
- Gonzalez-Pereira, B., Guerrero-Bote, V., & Moya-Anegón, F. (2009). The SJR indicator: A new indicator of journals' scientific prestige. *arXiv:0912.4141*.
- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, 102(46), 16569-16572.
- Jin, B. (2007). The AR-index: complementing the h-index. *ISSI newsletter*, 3(1), 6.
- Kowalski, G. (1997). *Information Retrieval Systems: Theory and Implementation* (1st ed.). Norwell, MA, USA: Kluwer Academic Publishers.
- Leydesdorff, L., De Moya-Anegón, F., & Guerrero-Bote, V. P. (2010). Journal maps on the basis of Scopus data: A comparison with the Journal Citation Reports of the ISI. *Journal of the American Society for Information Science and Technology*, 61(2), 352-369.
- Olivas, J. A. (2011). *Búsqueda Eficaz de Información en la Web*. La Plata, Buenos Aires, Argentina: Editorial de la Universidad Nacional de La Plata (EDUNLP).
- Pendlebury, D. A. (2009). The use and misuse of journal metrics and other citation indicators. *Archivum Immunologiae et Therapiae Experimentalis*, 57(1), 1-11.
- Salton, G., & McGill, M. (1983). *Introduction to Modern Information Retrieval*. McGraw-Hill, Inc.
- Serrano-Guerrero, J., Romero, F. P., Olivas, J. A., & De la Mata, J. (2009). BUDI: Architecture for fuzzy search in documental repositories. *Mathware & Soft Computing*, 16(1), 71-85.

# Estudio Comparativo y Análisis de Rendimiento de los Lenguajes de Manipulación de Datos en Bases de Datos Orientadas a Objetos y Bases de Datos Objeto-Relacionales

<sup>1</sup>FALAPPA, Marcelo Alejandro, <sup>1</sup>COBO, María Laura, <sup>1</sup>MARTÍNEZ, Diego César, <sup>2</sup>BENEDETTO Marcelo Gabriel, <sup>2</sup>CARABIO, Ana Lía Ramona, <sup>2</sup>ALVEZ Carlos Eduardo,, <sup>2</sup>FERNÁNDEZ, Miguel Antonio, <sup>2</sup>ETCHART Graciela Raquel, <sup>2</sup>CABRERA, Sergio Alberto

<sup>1</sup> Departamento de Ciencias e Ingeniería de la Computación - Universidad Nacional del Sur  
Avenida Alem 1253 - Bahía Blanca ( B8000CPB ) - Tel.: +54(0291)4595135  
{mfalappa, mlc, dcm}@cs.uns.edu.ar

<sup>2</sup> Facultad de Ciencias de la Administración - Universidad Nacional de Entre Ríos  
Monseñor Tavella 1424 – Concordia, Entre Ríos (3200) - Tel.: +54(0345)4231433  
{marben, anacar, caralv, migfer, graetc, sercab}@fcad.uner.edu.ar

## Resumen

El modelo relacional puro tiene una restricción muy fuerte: los tipos de datos deben ser atómicos para satisfacer la primera forma normal, así como las formas más altas. Para tratamiento de objetos complejos existen dos opciones: utilizar lenguajes de manipulación de datos que respeten el paradigma orientado a objetos y manejen objetos persistentes, o bien, utilizar bases de datos objeto-relacionales, las cuales proveen constructores para la manipulación de datos complejos respetando la robustez del modelo relacional.

Por su parte, los lenguajes de programación (LP) han tenido un desarrollo creciente, se han adaptado a las necesidades de diferentes dominios de aplicación, liberando versiones periódicamente y adaptados a estándares. Este crecimiento, generó dentro de los LP, adaptaciones a pautas de diseño e implementación de programas, permitiendo su utilización en el desarrollo de aplicaciones para bases de datos (BD), entre otras. Estas adaptaciones producidas en LP no específicos para BD, poseen la restricción impuesta por el modelo de BD y el modelo del LP, lo que no se produce en lenguajes específicos para BD. Además, los LP poseen características que procuran la producción de software con cualidades como confiabilidad, mantenibilidad y eficiencia, entre otras.

Este proyecto plantea la realización de un estudio comparativo de los lenguajes de manipulación de datos en sistemas de BD, analizando el rendimiento de lenguajes orientados a objetos en función de las cualidades del software que se produce con ellos, y comparándolos con los lenguajes de manipulación de datos en BD objeto-relacionales.

**Palabras Clave:** Bases de Datos, Lenguajes de Manipulación de Datos, Lenguajes de Programación, Lenguajes Orientados a Objetos, Cualidades del Software, Análisis de Rendimiento.

## Contexto

Este proyecto se enmarca dentro del Convenio Específico de Colaboración entre Programas de Investigación y Postgrado, celebrado en el año 2008 entre la Facultad de Ciencias de la Administración de la Universidad Nacional de Entre Ríos y el Departamento de Ciencias e Ingenierías de la Computación de la Universidad Nacional del Sur.

Además, se suscribió un Acuerdo de Colaboración Académico-Científico entre la Facultad de Ciencias de la Administración de la U.N.E.R. y el Instituto de Ciencias e Ingeniería de la Computación (ICIC) del Departamento de Ciencias e Ingeniería de la Computación de la U.N.S. para el desarrollo del presente proyecto de investigación.

Uno de los principales objetivos de este proyecto es la formación de recursos humanos para investigación en la Facultad de Ciencias de la Administración de la UNER, especializados en la línea prioritaria de investigación denominada “*Ingeniería de Software y Lenguajes de Programación*” establecida por Res. 25/11 del C.D. Al ser también ésta una de las principales líneas de investigación del Departamento de Ciencias e Ingeniería de la Computación de la UNS, se justifica la creación de un equipo de investigación inter-universidades que sea contenedor del desarrollo de investigadores de la U.N.E.R. en el área.

## Introducción

Las bases de datos son masivamente utilizadas en las aplicaciones de hoy en día. Detrás de la mayoría de los sistemas informáticos, existe una base de datos así como un sistema de manejo de la misma, que permite el acceso a los datos, brindando seguridad al usuario, recuperación ante fallos, posibilidad de acceso concurrente, y mecanismos de control de concurrencia que garantizan la atomicidad de las transacciones, y la correcta modificación de los datos. Dentro de los diferentes modelos, sin lugar a dudas, el modelo relacional es el modelo que más éxito ha tenido y que más ha perdurado en el tiempo. Probablemente la razón de esto sea debido a la solidez formal del modelo en sí, como así también a la eficiencia de los sistemas de manejo de bases de datos que lo utilizan. Sin embargo, el modelo relacional puro tiene una fuerte limitación derivada de las restricciones impuestas por las formas normales: los campos de las relaciones (tablas) deben ser atómicos. Actualmente, existen dos formas de extender dichos sistemas: mediante bases de datos objeto-relacionales, o mediante la interacción con lenguajes de programación orientados a objetos que permitan la manipulación de objetos persistentes.

En la actualidad, los lenguajes de programación poseen interfaces de desarrollo de aplicaciones y acceso a sistemas de base de datos. La mayoría de las aplicaciones de hoy, poseen interfaces amigables para el usuario que permiten interactuar con la base de datos. Existen varias técnicas para incluir estas interacciones en un lenguaje de programación: una consiste en escribir programas de aplicación en lenguajes de alto nivel, que integran instrucciones que permiten soportar la funcionalidad de una base de datos; o bien utilizar una técnica más dinámica pero más compleja, como las API's (API: *Application Programming Interface*), que son librerías de funciones y procedimientos que pueden ser utilizados por otro software como una capa de abstracción; en el contexto de bases de datos. Estas dos metodologías son las más comunes, pero presentan problemas derivados de las diferencias entre el modelo de la base de datos y el modelo del lenguaje de programación. Este problema disminuye con el uso de una tercera técnica, que es la utilización de un lenguaje de programación específico para construcción de aplicaciones que tienen mucha

interacción con la base de datos, brindando así, compatibilidad entre ambos modelos.

Las bases de datos y los lenguajes de programación han tenido un desarrollo creciente y, en la mayoría de los casos, se han ido adaptando a las necesidades propias de los diferentes dominios de aplicación. Los lenguajes de programación, en particular, son una herramienta de vital importancia para el desarrollo de muchas aplicaciones de software. Éstos poseen características particulares, tales como *simplicidad*, *legibilidad*, *facilidad de escritura*, facilidades para la *auto-documentación*, herramientas para un adecuado diseño de *interfaces*, que procuran la producción de software con cualidades tales como *confiabilidad*, *mantenibilidad* y *eficiencia*, entre otras.

Desde que el software se utiliza para la resolución de tareas complejas y/o críticas, la confiabilidad ha cobrado mayor importancia. El desarrollo de sistemas debe contemplar que los mismos sean tolerantes a fallas, es decir que continúen brindando soporte al usuario aún en presencia de eventos no frecuentes o indeseables, tales como anomalías de hardware o software.

En referencia a la mantenibilidad, no es económicamente factible descartar el software existente y desarrollar aplicaciones de reemplazo desde el principio, dado que el costo de desarrollo de software se ha incrementado y, por lo tanto, las aplicaciones existentes deben modificarse para satisfacer los nuevos requerimientos. En particular, la mayoría de los sistemas de software que utilizan bases de datos, y en especial, las que usan el modelo relacional, han utilizado, reusado y adaptado las bases de datos iniciales más allá de la evolución del sistema.

Por otra parte, la eficiencia ha sido siempre una cualidad deseable de cualquier software. Este requisito afecta tanto al lenguaje de programación, como a la elección del algoritmo a utilizar. Si bien el costo del hardware continúa descendiendo y, al mismo tiempo, su performance continúa en aumento, la necesidad de una ejecución eficiente persiste porque se utilizan computadoras con aplicaciones cada vez más exigentes.

La calidad del software que se desarrolla está directamente relacionada con la calidad que el lenguaje de programación brinda a través de sus características y atributos asociados. Pero también, la calidad

del software es el resultado de ciertos atributos asociados que no tienen relación directa con lo que el software hace, sino con la organización del programa fuente, la documentación y su comportamiento en ejecución.

Las cualidades deseables del software se pueden dividir en cualidades externas e internas. Las primeras son visibles para los usuarios de las aplicaciones, mientras que las internas son aquellas que afectan a los desarrolladores y se relacionan en gran medida con la estructura del software. En general, los usuarios sólo se preocupan por las cualidades externas, aunque son las cualidades internas las que ayudan a los desarrolladores a lograr dichas cualidades externas. En muchos casos, estas cualidades están estrechamente relacionadas y la distinción entre lo interno y externo no es tan fácil de evidenciar.

Entre las cualidades internas y externas del software más representativas, se encuentran:

**Correctitud:** un programa es funcionalmente correcto si se comporta de acuerdo a la especificación de las funciones que debe proporcionar.

**Confiabilidad:** se puede definir la confiabilidad en términos de comportamiento estadístico, como la probabilidad de que el software funcionará como se espera, en un intervalo de tiempo específico.

**Robustez:** un programa es robusto si se comporta de forma razonable, incluso en circunstancias que no se hayan contemplado en la especificación de requerimientos.

**Rendimiento:** en ingeniería de software, a menudo, se equipara rendimiento con *eficiencia*. Un sistema de software es eficiente si utiliza los recursos informáticos de manera económica.

**Facilidad de uso:** un software es fácil de usar, si a los usuarios les resulta fácil operarlo. Esta definición denota la subjetividad de la facilidad de uso. La interface de usuario es un componente importante de dicha cualidad.

**Verificabilidad:** un software es verificable si sus propiedades pueden ser chequeadas con facilidad. El diseño modular, buenas prácticas de codificación y el uso de un lenguaje adecuado contribuyen a la verificabilidad. Suele ser una cualidad interna, aunque a veces se convierte también en una cualidad externa.

**Mantenibilidad:** este término se utiliza comúnmente para hacer referencia a las modificaciones que se realizan en un software después de su lanzamiento inicial.

**Reusabilidad:** es un concepto similar a la capacidad de evolución. En la evolución del software, se modifica un producto para construir una nueva versión del mismo; en la reutilización del software, se producen algunos cambios para construir otro software.

**Portabilidad:** el software es portable si puede funcionar en diferentes ambientes.

**Comprensibilidad:** es una cualidad interna del producto que ayuda en el logro de otras cualidades, como la capacidad de evolución y la posibilidad de verificación.

**Interoperabilidad:** se refiere a la capacidad de un software de coexistir y cooperar con otros sistemas.

**Productividad:** es una cualidad del proceso de producción de software, que mide la eficiencia del proceso y corresponde a la cualidad de rendimiento aplicada al proceso.

**Oportunidad:** es una cualidad relacionada con el proceso que se refiere a la capacidad de ofrecer un producto a tiempo. Históricamente, la oportunidad que ha faltado en los procesos de producción de software llevó a la crisis del mismo, lo cual a su vez condujo al nacimiento de la *Ingeniería de Software*. En la actualidad, muchos de los procesos no culminan en un producto oportuno.

**Visibilidad:** un proceso de desarrollo de software es visible si todos sus pasos y estado actual son claros y están documentados.

De lo expuesto anteriormente, se desprende que en la actualidad la mayoría de las aplicaciones informáticas utilizan bases de datos, existen lenguajes de programación con facilidades para manipular las mismas, se reconocen las cualidades deseables del software, así como también las características que poseen los lenguajes de programación para contribuir al desarrollo de software con dichas cualidades. Uno de los objetivos principales de este proyecto es el estudio y análisis de los diferentes lenguajes de manipulación de datos, y sus efectos en la calidad del software generado por ellos.

## Líneas de investigación y desarrollo

Existen lenguajes de programación que ofrecen soporte para acceder y manipular datos



en bases de datos. El presente proyecto de investigación limitará el estudio de lenguajes de programación orientados a objetos, a aquellos que permitan la posibilidad de manipular objetos persistentes, así como lenguajes de manipulación de datos en bases de datos que, además, provean facilidades para el manejo de interfaces gráficas.

Cada uno de estos lenguajes, resulta más o menos adecuado que otro para producir software con ciertas cualidades, en función de las características que ellos presentan. Para poder determinar claramente si el software que producen posee las cualidades de interés para el presente trabajo, es necesario conocer sus características, su funcionalidad, su estructura; y efectuar mediciones, con la finalidad de comparar los resultados.

En base a los lenguajes seleccionados, las cualidades del software y las características elegidas para el estudio, se podrán conocer aquellos lenguajes que experimenten un mejor comportamiento en aplicaciones orientadas a sistemas informáticos con bases de datos, para luego poder compararlos con sistemas que utilicen bases de datos objeto-relacionales.

## Resultados y objetivos

El objetivo de este trabajo es realizar un estudio comparativo y análisis de rendimiento de lenguajes orientados a objetos, así como lenguajes de manipulación de datos en bases de datos objeto-relacionales. En cada caso, se buscará medir la calidad del software producido con ellos.

### Objetivos Específicos:

-Seleccionar lenguajes de manipulación de datos en bases de datos objeto-relacionales y lenguajes orientados a objetos que manipulen objetos persistentes.

-Realizar un estudio comparativo en base a los lenguajes seleccionados y a las características de los mismos, en función de las cualidades deseables del software que producen.

-Comparar globalmente sistemas desarrollados en lenguajes orientados a objetos que manipulen objetos persistentes con sistemas que utilicen bases de datos objeto-relacionales.

-Analizar los resultados del estudio comparativo efectuado y establecer un diagnóstico del rendimiento de los lenguajes de manipulación de datos estudiados.

## Formación de recursos humanos

Se brindará a los integrantes del proyecto, formación en lo que se refiere a técnicas avanzadas de orientación a objetos, desarrollo en lenguajes puros e híbridos dentro del paradigma, así como también en bases de datos orientadas a objetos y objeto-relacionales.

Los integrantes del proyecto, se desempeñan en cátedras relacionadas directamente con el tema central de la investigación, por lo que este trabajo tendrá impacto directo e inmediato en la docencia.

Se procederá a dirigir becarios de investigación, así como también tesinas finales de grado, dirigidos por el director del proyecto de investigación y/o por los docentes-investigadores del mismo. Para estos casos, también se prevé la presentación a convocatorias de becas ante organismos provinciales y nacionales.

Los integrantes participarán de reuniones científicas y técnicas que permitan actualizar los conocimientos en el tema de interés. También se trabajará en la presentación de trabajos en congresos nacionales e internacionales relacionados con el área del proyecto. Estos trabajos servirán para divulgar los conocimientos obtenidos durante el trabajo de investigación.

Uno de los principales objetivos del proyecto es que el personal docente de la UNER dedicado al mismo avance y/o concluya con sus estudios de posgrado, así como también se incorporen becarios realizando investigaciones en temas afines a la temática del proyecto.

El Director del proyecto, Marcelo Alejandro FALAPPA, es Doctor en Ciencias de la Computación egresado de la Universidad Nacional del Sur (UNS) en el año 1999. Actualmente es Profesor Asociado con Dedicación Exclusiva en el Departamento de Ciencias e Ingeniería de la Computación de la UNS a cargo de la materia curricular *Bases de Datos*, es Investigador Adjunto del CONICET especializado en Actualización de Bases de Datos Deductivas y Teoría de Cambio, y tiene Categoría II en el Programa de Incentivos.

Uno de los codirectores del proyecto, Marcelo Gabriel BENEDETTO, es Magíster en Sistemas de Información, se encuentra realizando los cursos del Doctorado en Informática de la Universidad de Murcia

(España) y cursando la Maestría en Desarrollo Local en la Universidad Nacional de General San Martín; posee categoría IV en el Programa de Incentivos.

La codirectora, Licenciada en Sistemas de Información Ana Lía Ramona CARABIO, posee categoría V en el Programa de Incentivos y se encuentra realizando la Maestría en Redes de la U.N.L.P., restando la presentación de la Tesis; el integrante Miguel Antonio FERNÁNDEZ es categoría IV en el Programa de Incentivos y se encuentra actualmente cursando la Especialización en Gestión de la Innovación y la Vinculación Tecnológica; el integrante Licenciado en Sistemas Sergio Alberto CABRERA se encuentra realizando la Maestría en Sistemas de Información de la UNER, restando sólo la presentación de la Tesis; la integrante María Laura COBO es Doctora en Ciencias de la Computación responsable de la cátedra *Lenguajes Formales y Autómatas*; el integrante Diego César MARTÍNEZ es Doctor en Ciencias de la Computación, responsable de las cátedras *Tecnologías de Programación e Ingeniería de Aplicaciones Web*, categoría III en el Programa de Incentivos; el integrante Carlos Eduardo ALVEZ es Doctor en Ingeniería - Mención Sistemas de Información, egresado de la Universidad Tecnológica Nacional - Facultad Regional Santa Fe en el año 2012, actualmente responsable de las cátedras *Lógica para las ciencias informáticas y Base de Datos*, y tiene Categoría II en el Programa de Incentivos. Finalmente, la Licenciada en Sistemas Graciela Raquel ETCHART se encuentra realizando los cursos del Doctorado en Ciencias de la Computación de la UNS y posee categoría V en el Programa de Incentivos.

## Referencias

GHEZZI, Carlo; JAZAYERI, Mehdi; MANDRIOLI, Dino. *Fundamentals of Software Engineering*. Second Edition. Prentice Hall Inc., 2002.

GHEZZI, Carlo; JAZAYERI, Mehdi. *Programming Language Concepts*. Third Edition. United States of America, John Wiley & Sons Inc., 1997.

ELMASRI, Ramez; NAVATHE, Shamkant B. *Fundamentals of Database Systems*. United States of America, Addison Wesley, 2006.

KUHN, Thomas S. *The Structure of Scientific Revolutions*. Third Edition. Chicago, University of Chicago Press, 1993.

LOUDEN, Kenneth C. *Lenguajes de Programación: Principios y Práctica*. Segunda Edición. México, Thomson Internacional, 2004.

MEYER, Bertrand. *Object-Oriented Software Construction*. Second Edition. 17<sup>th</sup> Printing 2011. United States of America, Prentice-Hall, 1997.

PRATT, Terrence W.; ZELKOWITZ, Marvin V. *Programming Languages. Design and Implementation*. Fourth Edition. United States of America, Prentice-Hall, 2001.

PRESMAN, Roger S. *Ingeniería del Software. Un enfoque práctico*. Sexta Edición. McGraw-Hill, 2005.

SEBESTA, Robert. *Concepts of Programming Languages*. Ninth Edition. United States of America, Addison Wesley, 2009.

SILBERSCHATZ, Abraham; KORTH, Henry F.; SUDARSHAN, S. *Database System Concepts*. Sixth Edition. Mc. Graw Hill, 2010.

SOMMERVILLE, Ian. *Software Engineering*. 9th Edition. Pearson Educación, 2010.

## Aproximación para un método de elicitación y especificación de requerimientos de seguridad para el desarrollo de software

Javier Antúnez, Marisa Panizzi

Facultad de Informática, Ciencias de la Comunicación y Técnicas Especiales, Universidad de Morón  
Cabildo 134 – Morón (CP 1708)  
Tel: 5627-2000

[Javier.antunez@gmail.com](mailto:Javier.antunez@gmail.com), [marisapanizzi@speedy.com.ar](mailto:marisapanizzi@speedy.com.ar)

### 1. Resumen

El propósito de esta línea de investigación consiste en el diseño de un método preliminar para la elicitación y especificación de requerimientos de seguridad de la información. Las consideraciones de seguridad en los proyectos de desarrollo de software suelen ser incorporadas en etapas tardías del desarrollo, con la dificultad de incorporar medidas de protección en elementos ya diseñados (o incluso ya desarrollados), y los altos costos asociados a esta incorporación.

La idea fuerza de este trabajo es enfocarse en la etapa de elicitación de requerimientos.

Se propone un método de elicitación basado en el método ARM (Accelerated Requirements Method) y alineado con la metodología SQUARE-Lite (Security Quality Requirements Engineering - Lite), que considera elementos del contexto de la aplicación y aprovecha el conocimiento de las partes interesadas en la elicitación y en la priorización de los requerimientos.

### 2. Palabras clave

*Seguridad de la información / Requerimientos de seguridad / Elicitación de requerimientos / Especificación de requerimientos / Requerimientos no funcionales (NFR)*

### 3. Contexto

Dadas las características del proyecto, el mismo se encuentra enmarcado en el área sistemas de información-ingeniería de software, más específicamente en la problemática de ingeniería de requerimientos y elicitación y especificación de requerimientos de seguridad de la información. Esta línea de investigación, tuvo su origen en un proyecto

de tesis de grado en desarrollo en la Facultad de Informática, Ciencias de la Comunicación y Técnicas Especiales de la Universidad de Morón.

### 4. Introducción

Mucho ha avanzado la tecnología y la ingeniería del software, las redes están cada vez mas interconectadas y es casi impensable desarrollar negocios sin el uso de tecnologías de información y conexión con Internet.

La evolución de los sistemas de información nos ha llevado desde los esquemas centralizados de hace un algunas décadas atrás, con lenguajes estructurados, a los nuevos sistemas distribuidos desarrollados en múltiples capas con lenguajes orientados a objetos (muchas veces, incluso con distintos lenguajes entre capas). El enfoque de proteger un único sistema o equipo de los esquemas centralizados, no es factible en la actualidad, ya que los sistemas desarrollados desde un tiempo a la fecha, funcionan completamente distribuidos, desplegados sobre múltiples plataformas, con diversos lenguajes de programación y muchas veces con servicios provistos por distintas organizaciones

Este nuevo escenario, genera un aumento en la cantidad de puntos que deben ser cubiertos para asegurar que un problema no afecte el normal desempeño de la organización.

Parte de estos problemas son cubiertos a través de distintas soluciones técnicas y herramientas de seguridad bien conocidas y cuentan con altos niveles de efectividad para su función

(Firewalls, IDS/IPS, VPNs, Antivirus, Antispywares, Sistemas de AAA, PKI, entre otros). La mayoría de estas herramientas actúan a nivel de red y se encuentran implementadas en porcentajes importantes de las organizaciones (en promedio del orden de más del 90% de las organizaciones medianas a grandes). **(RICHARDSON, 2012)**

Debido al nivel de evolución y madurez de las herramientas de protección, y su efectividad para la función que proveen, proliferaron nuevos focos de exposición, desplazando el tipo de ataques desde el nivel de red/infraestructura hacia el nivel de aplicación.

El nivel de madurez relativo a la seguridad de las aplicaciones es mucho menor comparada con la seguridad de otros componentes.

Es común encontrarse con fallas graves de seguridad en los desarrollos, esto se atribuye a que:

- Los requisitos de seguridad, en la mayoría de los casos, no son expresados en forma explícita por el cliente, lo que los convierte en requisitos no funcionales. **(Sindre, 2001)**
- La seguridad del software/servicio se suele considerar como algo “adicional”, que se diseñará e implementará luego que este funcione.
- Los ingenieros de requerimientos no poseen formación adecuada, ni experiencia, como para identificar requisitos funcionales y no funcionales relacionados con la seguridad, especificarlos e incorporarlos en etapas tempranas del proceso de desarrollo.
- Existen malas prácticas de desarrollo llevadas adelante por algunos desarrolladores, muchas veces relacionadas con el bajo índice de madurez de la organización (que no incorpora un marco adecuado de control de los procesos).

Bajo estas condiciones, es difícil poder realizar un aseguramiento adecuado de las condiciones de calidad y seguridad en un entorno de sistemas dado.

Hay estudios que estiman, que el 40% de los errores se detectan en la fase de integración y el 25% en la etapa de beta y post-implementación. **(NIST, 2002).**

Según información del NIST **(NIST, 2002)** resolver un error detectado en fases tardías del proyecto, puede costar entre 10 veces (etapa de integración) y 30 veces (etapa de post implementación) más que si es detectado y resuelto en las fases tempranas.

En este sentido los desarrollos que dejan la seguridad “para después” y no incorporan los requisitos desde la especificación de requerimientos, se ven afectados fuertemente en lo que respecta a tiempos, costos y retrabajos.

Todo esto sin contar las consecuencias y costos de posibles incidentes de seguridad (según estudios recientes los incidentes de seguridad analizados tienen un costo promedio entre USD 100.000 y 300.000, pero se han dado casos de compromiso de la totalidad de la información con costos directos de millones de dólares e indirectos no mensurables -daño en reputación, pérdida de valor de las acciones de las compañías involucradas en el incidente, demandas legales, entre otros-). **(Ponemon Institute LLC, 2012)**

Para tratar la problemática de la seguridad en fases tempranas del proceso de desarrollo de software, hacia el inicio del proyecto, la etapa de elicitación de requerimientos resulta un momento oportuno, ya que sus resultados establecen en gran medida lo que se desarrollará en etapas posteriores. Es decir, los requerimientos constituyen la “materia prima” utilizada para muchas de las decisiones y actividades de fases posteriores del ciclo de vida.



La problemática de la seguridad en las aplicaciones, se viene tratando en distintos papers.

En el análisis de los antecedentes bibliográficos, hemos identificado que existen marcos metodológicos como SQUARE (Mead, Houg, & Stehney, Security Quality Requirements Engineering (SQUARE) Methodology, 2005), SQUARE-Lite (Gayash, Viswanathan, & Mead, 2008), SREP (Mellado, Fernandez-Medina, & Piattini, 2006) y STRIDE (Hernan, Lambert, & Shostack, 2006) para la elicitación, especificación y análisis de requerimientos de seguridad (y en algunos casos análisis de riesgos).

También trabajos que usan como base el modelado basado en árboles de ataque (Schneier, 1999), el comportamiento de un potencial atacante (Lamsweerde, Brohez, De

Landsheer, & Janssens, 2003); el trabajo con metas y catálogos (Saeki & Kaiya, 2009) y restricciones de seguridad (Mouratidis & Giorgini, 2007).

Existe también literatura donde se extienden los casos de uso de UML para analizar aspectos de seguridad (Sindre, 2001), (Firesmith, 2003); (Kabasele Tenday, 2010).

Finalmente se revisaron los estándares que poseen apartados de relevancia, como ser aquellos aspectos relativos a la especificación de requerimientos de software (SRS) en los estándares IEEE std 830:1998, ISO/IEC/IEEE 15288:2008, ISO/IEC/IEEE 12207 y el estándar ISO/IEC/IEEE 29148:2011.

Para la construcción de la solución se propone seguir el esquema conceptual de la **Figura Nro. 1**.

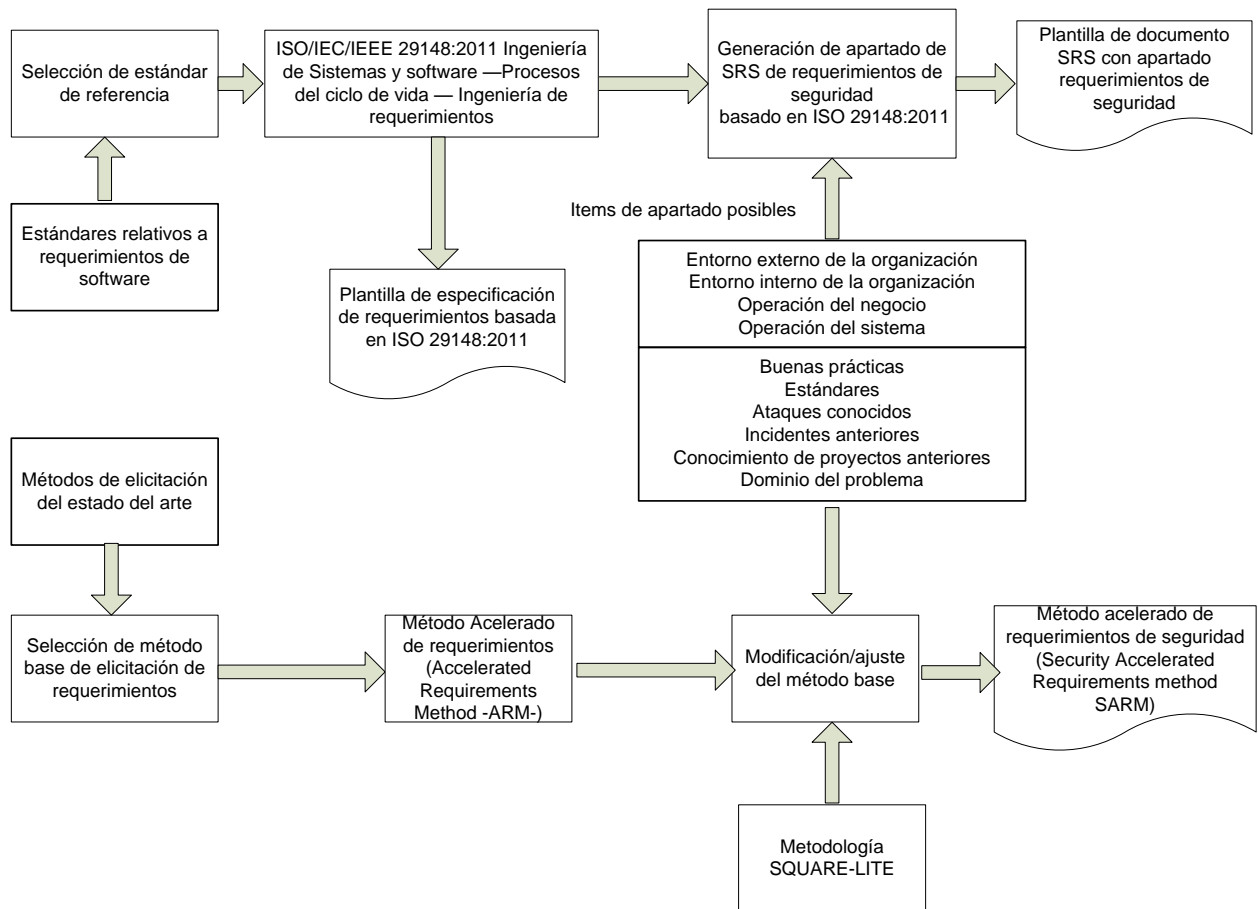


Figura Nro. 1 - Esquema conceptual de construcción de la solución

Trabajando en 2 grandes líneas. La primera busca sentar las bases para la especificación, tomando como base un estándar reconocido.

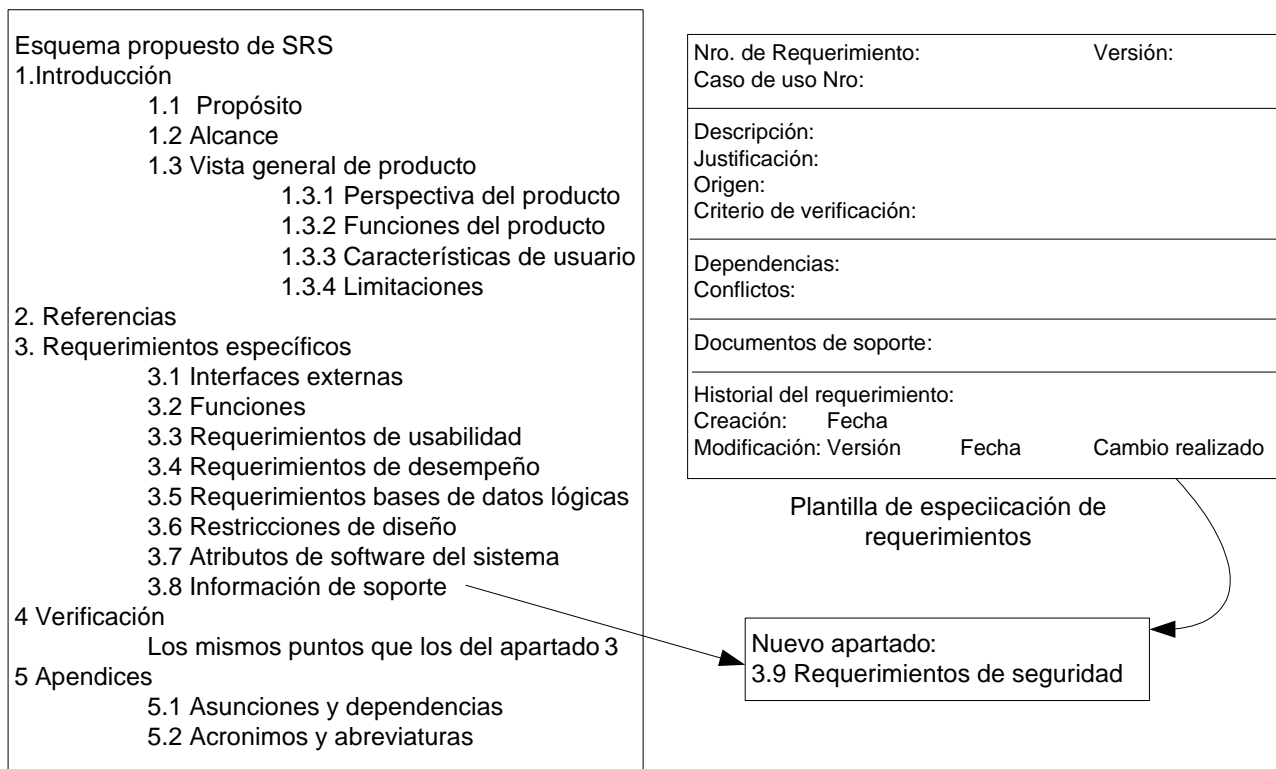
Entre los estándares disponibles relativos a ingeniería o especificación de requerimientos, se seleccionó el estándar ISO/IEC/IEEE 29148:2011, y se intenta crear un nuevo apartado para el documento de especificación de requerimientos de software (SRS), para la especificación de requerimientos de seguridad (ver **Figura Nro. 2**) y en conjunto con esto, una plantilla para la especificación individual de requerimientos que cumpla los requisitos especificados por el mismo estándar.

La segunda línea de actividades, constará de la selección de un método de elicitación existente como base, para construir un método modificado adaptado a la elicitación de

requerimientos obtenidos) buenos resultados en las pruebas realizadas por el equipo del Software Engineering Institute de la Universidad de Carnegie Mellon (**Mead, Requirements Elicitation Case Studies Using IBIS, JAD, and ARM, 2006-2008**).

Se intentará a través de las modificaciones, resolver las limitaciones del método original identificadas por investigaciones previas, principalmente que el resultado de la elicitación eran requerimientos funcionales.

Para conseguir este resultado, se introducirán nuevos pasos y modificaran algunos pasos existentes. Por ejemplo parece oportuno introducir un o varios pasos para realizar la identificación de metas y objetivos de seguridad, asegurarse que las partes interesadas que pueden realizar aportes



**Figura Nro. 2** - Esquema general propuesto para el documento SRS y plantilla de requerimientos

requerimientos de seguridad. Preliminarmente se utilizará como base el método Accelerated Requirements Method (ARM), que demostró gran flexibilidad y (pese a ciertas limitaciones identificadas, principalmente con el tipo de

significativos desde el punto de vista de los requerimientos de seguridad (como por ejemplo los sectores de Seguridad de la Información, Legales, Cumplimiento, Auditoría, etc.).

Finalmente podría modificarse la etapa de cierre para que se requiera que los entregables (documento de Especificación de Requerimientos de Software – SRS), cumplan con la plantilla de requerimientos y del apartado de requerimientos de seguridad.

## 5. Resultados esperados

Se pretende probar el método en un caso real, sobre un proyecto ya finalizado, para analizar los resultados obtenidos originalmente y compararlos con los resultados de la aplicación del método propuesto.

Se espera como resultado, identificar requerimientos que no fueron obtenidos mediante el proceso de análisis de requerimientos tradicional (aplicado en originalmente en el proyecto testigo) e idealmente que todos los requerimientos obtenidos originalmente surjan como resultado de aplicar el método propuesto.

## 6. Bibliografía

Firesmith, D. G. (2003). Security use cases. *Journal of Object Technology*, 53-64.

Hernan, S., Lambert, S. O., & Shostack, A. (2006, November 1). Uncover Security Design Flaws Using The STRIDE Approach. Retrieved Octubre 06, 2012, from MSDN Magazine: <http://msdn.microsoft.com/hi-in/magazine/cc163519%28en-us%29.aspx>

Kabasele Tenday, J.-M. (2010). Using Special use case for security in the software development life cycle. *Information Security Applications: 11th International Workshop WISA 2010* (pp. 122-134.). Springer.

Lamsweerde, A. v., Brohez, S., De Landsheer, R., & Janssens, D. (2003). From System Goals to Intruder Anti-Goals: Attack Generation and Resolution for Security Requirements Engineering. 11th IEEE International Requirements Engineering Conference RE03 - 2nd Workshop on Requirements for High Assurance Systems REHAS 03 (pp. 49-56). Monterrey Bay, CA, USA: Carnegie Mellon - Software Engineering Institute.

Mead, N. R. (2006-2008, 09 22). Requirements Elicitation Case Studies Using IBIS, JAD, and ARM. Retrieved 10 14, 2012, from Build Security In: <https://buildsecurityin.uscert.gov/bsi/articles/best-practices/requirements/532-BSI.html>

Mead, N. R., Houg, E. D., & Stehney, T. R. (2005). Security Quality Requirements Engineering (SQUARE) Methodology. Pittsburgh, PA: Carnegie Mellon University - Software Engineering Institute (CMU-SEI).

Mellado, D., Fernandez-Medina, E., & Piattini, M. (2006). A Common Criteria Based Security Requirements Engineering Process for the Development of Secure Information Systems. In *Varios, Computer Standard and Interfaces* (pp. 244-253). Madrid y Ciudad Real - España: Elsevier.

Mouratidis, H., & Giorgini, P. (2007). SECURE TROPOS: A SECURITY-ORIENTED EXTENSION OF THE TROPOS METHODOLOGY. *World Scientific Publishing*, 17 (2) 285-309.

NIST. (2002). NIST Report "The Economic impacts of inadequate infrastructure for software testing". NIST.

Ponemon Institute LLC. (2012). The impact of cybercrime on Business. Ponemo Institute LLC sponsored by Checkpoint Software Technologies.

RICHARDSON, R. (2012). Computer Security Institute/Federal Bureau of Investigations. COMPUTER CRIME AND SECURITY SURVEY 2010-2011. CSI-FBI.

Saeki, M., & Kaiya, H. (2009). Using Common Criteria as Reusable Knowledge in Security Requirements Elicitation. Tokyo, Japón: Dept. of Computer Science, Tokyo Institute of Technology.

Schneier, B. (1999). Modeling Security Threats. *Dr. Dobb's Journal*, December ISSUE.

Sindre, G. O. (2001). Capturing security requirements trough misuse cases. Retrieved 05 01, 2006, from folk.uio.no: <http://folk.uio.no/nik/2001/21-sindre.pdf>

## Construcción de un Prototipo de Repositorio Temático en el campo de la Ingeniería de Software

María V. Doria<sup>1</sup>, Germán A. Montejano<sup>2,3</sup> & Claudia I. Inchaurredo<sup>1</sup>

<sup>1</sup>Dpto. Sistemas – Facultad de Tecnología y Cs. Aplicadas – Universidad Nacional de Catamarca

Maximio Victoria 55 - C.P: 4700 - San Fernando del Valle de Catamarca

TEL. 03833- 435112

vanesadoria@gmail.com, cinchaurrondo@sistemas.frc.utn.edu.ar

<sup>2</sup>Facultad de Cs. Físico, Matemáticas y Naturales, Universidad Nacional de San Luis, Ejército de los Andes 950 – CP: 5700 – San Luis

<sup>3</sup>Facultad de Ingeniería, Facultad de Ciencias Exactas y Naturales, Universidad Nacional de La Pampa – La Pampa  
gmonte@unsl.edu.ar,

### Resumen

El presente trabajo tiene por objeto mostrar la línea de investigación que se llevará a cabo para la construcción de un prototipo de Repositorio Temático en el campo disciplinar de la Ingeniería de Software que pretende ser implementado en el Dpto. de Sistemas de la Facultad de Tecnología y Ciencias Aplicadas de la Universidad Nacional de Catamarca.

El repositorio de información temático a construir debe permitir el acceso a producción científica de manera libre y abierta, para ello es importante identificar los pasos necesarios para su construcción.

Asimismo para el acceso a la información a través de repositorios de información se deben tener en cuenta las iniciativas que propone el Acceso Abierto (AA).

**Palabras Clave:** Acceso Abierto, repositorio temático.

### Contexto

El desarrollo y crecimiento de las estrategias que forman parte de las iniciativas de AA, representa un beneficio fundamental para la comunicación científica y académica mundial. Estas estrategias están referidas a los repositorios institucionales y temáticos, que almacenan la producción científica de instituciones científicas o académicas, constituyendo la vía más propicia para elevar la visibilidad y el impacto de la investigación. La literatura existente sobre desarrollo y gestión de repositorios temáticos es escasa comparada con la amplitud de información alcanzada por repositorios institucionales.

De acuerdo al contexto mencionado, con esta investigación se busca aportar información sobre los pasos a seguir para el desarrollo y gestión de un repositorio temático en el campo disciplinar de Ingeniería de Software.

La línea de investigación presentada en este trabajo se desarrolla en el marco de una tesis de Maestría en Ingeniería de Software de la Universidad Nacional de San Luis.

### Introducción

La comunicación científica es una de las instancias fundamentales de la actividad investigadora, siendo la publicación el mecanismo formal que regula el proceso por el cual los investigadores realizan sus contribuciones al acervo de conocimientos científicos. De este modo se genera la literatura científica que va configurando la dimensión bibliográfica de las disciplinas, entendidas como sistemas de producción y disseminación de conocimientos [1].

Un repositorio de información, institucional ó temático, representa una alternativa a la publicación y difusión de la literatura científica; si bien no es un canal formal de publicaciones sirve como estrategia para acceder a la información de manera libre y abierta. Libre se refiere a la consulta del documento completo en línea sin costo y abierta esta relacionada con algunos derechos de uso adicionales como el de modificar y distribuir la obra siempre y cuando se cite al o los autores.

Enfocándonos en el objeto de estudio de esta investigación, los repositorios temáticos, existe un consenso general respecto a que fueron los



primeros repositorios en aparecer, pueden describirse como proveedores de servicios que recolectan datos estructurados relacionados con un tema, permitiendo posteriormente al usuario buscar a través de muchas fuentes distribuidas en todo el mundo desde un punto de acceso único [2][3].

Surgieron en el ámbito de disciplinas académicas concretas, tales como, la física, ciencias de la información, ciencias cognitivas, salud, entre otras.

Las características principales de los repositorios temáticos, son las siguientes [4]:

- Permiten almacenar documentos pre-prints (antes de pasar cualquier proceso de arbitraje) o post-prints (documentos revisados por pares y aceptados, publicados o en proceso de publicación) y pueden ser artículos de revistas, comunicaciones en congresos, capítulos para libros o cualquier otra forma de comunicación científica.
- Los documentos disponibles en los repositorios de información se encuentran a texto completo y el acceso es gratuito, libre de cualquier tipo de restricciones.
- Los propios autores son los responsables de la introducción de los documentos en el repositorio, mediante el proceso de autoarchivo.

El funcionamiento de un repositorio temático consiste en que el autor se registre y de esta manera se le otorga un espacio desde el cual incorporar los documentos. En este espacio introducirá los metadatos, descripciones estructuradas y opcionales que están disponibles de forma pública para ayudar a localizar el recurso (documento), es la información no relevante para el usuario final pero de suma importancia para el sistema que administra los documentos. Los metadatos son requeridos a través de formularios, y finalmente depositará el recurso, que según el mecanismo de publicación podrá visualizarse públicamente.

Los metadatos solicitados para el recurso son los que pertenecen a la clase de metadatos descriptivos, en el caso de trabajos científicos son título, palabras clave, resumen, fecha, entre otros que permiten la identificación y su localización a través de la computadora. Estos datos son recolectados por diferentes

proveedores de datos (administradores de contenido) que soportan el protocolo Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). [5]

Las características de este protocolo son [6]:

- Simplicidad: los creadores buscaron una forma simple que se encuentre al alcance de cualquier usuario.
- Normalización: basados en estándares ampliamente utilizados como el protocolo http (Hiper Text Transport Protocol) para la transmisión de datos y órdenes y XML (Extended Markup Lenguaje) para la codificación de metadatos.
- Recolección: existe una entidad que pone a disposición de los interesados información bibliográfica sobre los documentos que almacena.

Por otro lado, los elementos del protocolo son los siguientes [12]:

- RECURSO: Es el documento o producción científica
- ITEM: Colección de propiedades sobre un RECURSO, es decir, metadatos que describen el recurso (título, autor, etc.)
- REGISTRO: Es un ITEM expresado en un formato específico formado por:
  - Encabezamiento: contiene un identificador único (identifica de forma inequívoca un artículo dentro de un repositorio), marca de fecha (fecha de creación, modificación o eliminación del registro con el fin de la recolección selectiva) cero o el conjunto de miembros de la partida con el propósito de la recolección selectiva, un opcional de estado de un atributo con un valor de borrado que indica el retiro de la disponibilidad del formato de metadatos especificado para el elemento.
  - Metadatos: se debe especificar le modelo de metadatos elegido
  - Acerca de: es una parte opcional en la que se pueden definir declaraciones de derechos y declaraciones de procedencias.

El protocolo OAI-PMH recomienda usar el modelo de metadatos Dublin Core sin calificar. [7].

Este modelo se compone de 15 elementos agrupados según:

- Contenido
  1. Title (título): cuando el recurso no se trata de un documento de texto el título no es obvio, como el caso de imágenes, es posible utilizar una frase descriptiva.
  2. Subject (tema): incluye tanto el resumen como las palabras clave del recurso.
  3. Description (descripción): debe dar cuenta del contenido del recurso, por ejemplo puede incluir la introducción, una tabla de contenido o un texto descriptivo del recurso.
  4. Source (fuente): es un elemento relacionado a una Referencia a un recurso de la cual el propio recurso se deriva.
  5. Language (lenguaje): especifica el lenguaje del recurso.
  6. Relation (relación): referida con la relación que puede existir con otros recursos.
  7. Coverage (cobertura): relacionada con incluir la ubicación geográfica y duración temporal del recurso.
- Propiedad Intelectual
  8. Creator (autor): referido a las personas que tienen la propiedad intelectual del recurso.
  9. Publisher (editor): es el encargado de la edición que puede ser una persona o empresa.
  10. Other Agent (otras colaboraciones): se refiere a aquellas personas que sin ser autor/es o editores mantienen una relación directa con la realización del recurso.
  11. Rights (derechos): elemento referido a la propiedad intelectual, derechos de autor y licencias.
- Instanciación (Representación)
  12. Date (fecha): fecha a partir de la cual el objeto está disponible, la forma recomendada se define en la norma ISO 8601: AAAA-MM-DD.
  13. Type (tipo de recurso): es un elemento que sirve para especificar el género del recurso y su propósito.
  14. Format (formato): referido al formato del recurso o dimensiones (tamaño y duración).

15. Identifier (identificador): es un elemento para identificar el recurso por medio de una cadena conforme a un sistema de identificación formal, como un ISBN.

Otro tema importante en el ámbito de los repositorios temáticos es el conocimiento de los aspectos legales referidos a la propiedad intelectual, los derechos de autor y las licencias.

Sobre propiedad intelectual se debe conocer lo establecido en Argentina por la ley de Propiedad Intelectual 11723 [8] y lo establecido por la Organización Mundial de la Propiedad Intelectual (OMPI) [9].

La OMPI también establece los aspectos legales para el derecho de autor y las licencias.

### Líneas de Investigación y desarrollo

Para llevar a cabo la investigación la misma se dividió en fases:

- Fase Exploratoria: se ha analizado el objeto de investigación a partir de la información recopilada que procede no sólo de documentos tradicionales como artículos, tesis y libros sino también, por diversos recursos en línea como blogs, wikis, sitios Web de organizaciones e iniciativas de AA, entre otros. También han sido de especial utilidad otros tipos de documentos como informes o estadísticas nacionales e internacionales, que permitieron profundizar en el estado del arte.

El resultado obtenido es que a pesar que los repositorios temáticos son los primeros en aparecer, la literatura que discute el tema es limitada y se encuentra enfocada más hacia el campo de la Biblioteconomía y la Ciencia de la Información que a la construcción de repositorios temáticos.

- Fase Descriptiva: se han estudiado y definido las directrices de creación de repositorios de información.
- Fase Proyectiva: se ha comenzado el abordaje de la propuesta dentro de lo que se conoce como Investigación Aplicada ya que se propone transformar el conocimiento puro adquirido, en conocimiento utilizable, para ello se construirá un prototipo de repositorio

temático en el campo disciplinar de la Ingeniería de Software.

La construcción del prototipo del repositorio temático en IS se hará conforme a los siguientes pasos:

- Paso 1: Análisis de plataformas de software para la creación de repositorios de información. Se ha comenzado a trabajar en este paso.
- Paso 2: Selección de la plataforma de software adecuada para el repositorio temático en IS.
- Paso 3: Diseño del repositorio temático.
- Paso 4: Prueba del prototipo.
- Paso 5: Presentación del prototipo desarrollado a expertos y potenciales usuarios para conocer su opinión y reconocer las ventajas y desventajas del prototipo creado.

### Formación de Recursos Humanos.

Los estudios realizados en esta línea de investigación pertenecen al trabajo de Tesis de Maestría “*Modelo de un Repositorio en Ingeniería de Software para la Facultad de Tecnología y Cs. Aplicadas de la UNCa*”.

El equipo de trabajo está formado por:

Tesista: Lic. Maria Vanesa Doria.

Director: Mg. Lic. Germán Antonio Montejano.

Co – Director: Mg. Ing. Claudia Inés Inchaurredo

Con este trabajo de tesis se prevé continuar con la formación y fortalecimiento en investigación de la tesista a través de la guía del director y codirector.

Como así también poder interactuar con otros investigadores del país que se encuentran en la misma línea de investigación.

### Referencias.

[1] S. Miguel; El acceso abierto a las publicaciones de la comunidad científica argentina en el campo de la medicina. *Renata*, Vol. 1 N° 2, ISSN 2027-7415, pp 50-65, (2011)

[2] Berlin Declaration on Open Access to

Knowledge in the Sciences and Humanities. (2003) [online] <<http://oa.mpg.de/openaccessberlin/berlindeclaration.html>> [Consulta: 18 de marzo de 2011].

[3] Declaración de Salvador sobre acceso abierto: la perspectiva del mundo en desarrollo. [online] <[http://www.bvs.org.ar/pdf/Salvador-Acceso Abierto.pdf](http://www.bvs.org.ar/pdf/Salvador-Acceso%20Abierto.pdf)> [Consulta: 1 de Noviembre de 2011].

[4] Vásquez, G.; “Repositorios Digitales de Documentos”, Primer Curso sobre Escritura Científica y Repositorios de Acceso Abierto, 6 y 7 de diciembre de 2007

[5] Doria, M., acceso Abierto y Repositorios Temáticos, Editorial Académica Española, pp.9 ISBN 978 – 3-659 -06908 – 6, (2013)

[6] Y. Sarduy Domínguez, P. Urra Gonzalez, Herramientas para la creación de colecciones digitales, *Revista Acimed*, Vol. 14 N° 5, (2006) [online] Disponible: <[http://bvs.sld.cu/revistas/aci/vol14\\_5\\_06/aci19506.htm](http://bvs.sld.cu/revistas/aci/vol14_5_06/aci19506.htm)> [Consulta: 1 de Noviembre de 2011]

[7] Dublin Core [online] Disponible en: <http://dublincore.org/> [Consulta: 23 de Marzo de 2012]

[8] Ley de Propiedad Intelectual, [online] Disponible: <http://www2.mincyt.gov.ar/11723.htm>, [Consulta: 1 de Noviembre de 2011]

[9] Organización Mundial de la Propiedad Intelectual, ¿Qué es La propiedad intelectual?, OMPI N° 450(S). ISBN 92-805-1157-0

## Actualidad del estado de una nueva metodología de gestión y desarrollo de proyectos de software: Metodología Ágil basada en Telecomunicaciones MATE

Marisa Tumino, Juan Bournissen, Claudio Bracalenti, Erik Schlemper, Edgar Streuli y Silvio Kucharsky  
I3@uap.edu.ar  
Universidad Adventista del Plata  
Instituto de Investigación en Informática – I3

### Resumen

El presente trabajo tiene como objetivo primario proponer una metodología ágil que facilite el desarrollo de aplicaciones, para dispositivos móviles, utilizando, como plataforma de desarrollo, medios que permitan el trabajo a distancia y puedan, incluso, soportar desplazamientos de los implicados. El objetivo secundario es el desarrollo de un producto, que facilite la gestión de los proyectos, utilizando la metodología propuesta. Las metodologías ágiles más usuales, y desde donde se han adoptado los principios básicos del desarrollo de aplicaciones para dispositivos móviles, tales como SCRUM y *eXtreme Programming* (XP), sufren limitaciones condicionadas a las localizaciones físicas de los miembros del *stakeholders*<sup>1</sup>.

En virtud de las condiciones demandadas por las nuevas modalidades de desarrollo *Freelance*, donde los miembros de los equipos se encuentran ubicados en diferentes puntos geográficos del mundo, se considera relevante elaborar una propuesta idónea que proporcione respuestas a estas limitaciones.

**Palabras clave:** Teletrabajo, Metodología ágil, Desarrollo semisincrónico distribuido, Virtualización de empresas informáticas, Telefonía móvil.

### Contexto

El presente trabajo es coordinado por la Universidad Adventista del Plata, en el marco de los proyectos de investigación desarrollados

por el Instituto de Investigación en Informática (I3) de la mencionada institución educativa.

### Introducción

Observaciones de las prácticas: Evaluación comparativa de herramientas

Un primer desafío que enfrentó esta iniciativa fue buscar herramientas que asistan a la metodología de trabajo distribuido.

Se han efectuado pruebas sobre un conjunto de productos, tanto pagos como gratuitos, que ofrecen una potencial solución al problema que aborda el trabajo.

Para compartir aplicaciones, escritorios y comunicarse, mediante video conferencia, se probaron las siguientes herramientas: *Adobe Connect*, *TeamViewer*, *HangOut* y *Saros*, tanto en forma individual como combinada, excepto con *Adobe Connect*, obteniendo los resultados que se describen a continuación.

Dentro de las dificultades encontradas durante las pruebas, cabe señalar las siguientes:

1. *Adobe Connect* es una herramienta orientada a conferencias con la que se puede compartir escritorio, escribir en pizarra, organizar usuarios y administrar la comunicación entre los participantes. Esta herramienta resulta útil para emplearse en exposiciones o eventos generales donde se presenta una ponencia con la participación de un auditorio remoto. Asimismo permite grabar las conversaciones y almacenar el título de los distintos bloques en que se puede fraccionar la ponencia. Esta herramienta funciona con un servidor centralizado de Adobe, lo que implica que cada integrante de la reunión mantiene una conexión con el servidor para funcionar direccionando todo el tráfico de red al mismo.

<sup>1</sup> **Stakeholder:** quienes pueden afectar o son afectados por las actividades de una empresa. (R. E. Freeman – “*Management: A Stakeholder Approach*” - Pitman, 1984)



1.1. Ancho de banda: tanto de subida como de bajada, debe mantenerse el mismo ancho, con lo que la conexión a internet debe asegurar una gran capacidad de ambos lados, y no como generalmente ofrecen los *Internet Service Provider* (ISP) con gran velocidad de bajada y poca velocidad de subida. De *download* se requiere un mínimo de 300 kbps y lo ideal sería obtener la misma velocidad de *upload* aunque con 100 kbps funciona correctamente en caso de no pretender transmisión de video.

## 1.2. Comunicación

1.2.1. Retardo: durante las pruebas la señal sonora del *stream* tuvo un retraso promedio de diez segundos, lo que permitió una comunicación fluida mientras hubo un único expositor unidireccional pero, dado el desfase, no fue apropiado cuando se debieron compartir ideas entre varios participantes.

1.2.2. Cortes: los cortes observados en la comunicación se debieron al insuficiente ancho de banda disponible. Por su parte los ISP no proveen un ancho de banda constante.

Estas consideraciones hacen de esta herramienta un medio no recomendable para ser utilizado en el trabajo diario del equipo de desarrollo distribuido donde los participantes requieren un flujo continuo y dinámico de comunicación bidireccional.

2. *Team Viewer* es una aplicación utilizada para mantener videoconferencias entre múltiples personas, compartiendo el escritorio del organizador de la reunión. La naturaleza de la imagen compartida puede ajustarse para máximo rendimiento o máxima calidad. En las pruebas realizadas con las videoconferencias, se pudo apreciar eventualmente leves retrasos y/o desacoples entre imagen y sonido, con tendencia a pérdidas parciales, distorsiones, reverberancias y acoples del sonido, oscilando los resultados desde un buen desempeño hasta la incapacidad de comunicarse en forma comprensible.

La herramienta funciona muy bien para uso compartido de escritorios, permitiendo a los diferentes participantes tomar el control del ordenador del expositor, aunque no simultáneamente, lo que permite el desarrollo en parejas en forma distribuida.

3. *HangOut* consiste en una funcionalidad de *google+*, la red social de Google, y actúa en

forma similar a *google talk*, diferenciándose en que *HangOut* permite realizar conferencias entre múltiples personas. Además permite compartir documentos, aplicaciones (propias de *google+* o *apps* específicas) y escritorios, aunque con una calidad inferior a la que ofrece *TeamViewer* y sin la posibilidad de obtener el control remoto. En pruebas realizadas se ha podido observar que *HangOut* responde con retrasos y algunas pérdidas cuando el tráfico de red aumenta.

4. *Saros* nace como un *plugin* (Plug-in) para el entorno de desarrollo Eclipse, que permite el desarrollo colaborativo de aplicaciones. Los cambios efectuados por uno de los integrantes del equipo se reflejan en las versiones locales de los desarrolladores. Esta herramienta permite la transferencia de un proyecto completo pero se recomienda utilizar una herramienta de versionado de software a los efectos de mantener disponible todas las versiones de los archivos que forman parte del historial del proyecto.

Pruebas combinadas: *TeamViewer*, *HangOut* y *Saros*

En este punto se señala que el objetivo fundamental de efectuar pruebas con la combinación de herramientas, fue probar su rendimiento ante una demanda de uso más exigente de las capacidades de la red.

En primera instancia se realizaron pruebas con *TeamViewer* y *HangOut* para luego evaluar el desempeño conjunto de *HangOut* y *Saros*, ambas herramientas corriendo al mismo tiempo en dos computadoras distintas. Una de ellas fue una máquina de escritorio con altos recursos, en lo que a hardware se refiere, la otra fue una *netbook* con limitaciones en tanto a sus capacidades gráficas y de procesamiento.

Se procedió a probar la comunicación por medio de voz ip en las circunstancias mencionadas y como fruto de las experiencias puede destacarse que resulta más recomendable la utilización de *Saros*, para compartir el código, y *HangOut*, que tiene la ventaja de indicar si las personas involucradas se encuentran conectados a *Google+*, *Google talk* o *Gmail*, para compartir la imagen y la voz. Por su parte se ha prescindido del *Adobe Connect*

por no reunir los requerimientos mínimos que exige esta modalidad de trabajo. Bajo estas condiciones la programación en parejas opera con eficiencia dado que la tecnología permite una comunicación fluida con la posibilidad de compartir pantallas según diferentes opciones.

Evaluación comparativa entre metodología ágil centralizada y la distribuida

A partir del análisis de estas experiencias, y con la visión enfocada en una nueva metodología ágil distribuida, se procedió intencionalmente a efectuar una evaluación comparativa del desempeño de los desarrolladores en ambas modalidades metodológicas: centralizada y distribuida.

Para tal efecto, uno de los desarrolladores del equipo integró simultáneamente dos *stakeholders* utilizando en uno de ellos la modalidad centralizada de metodología ágil y en el otro la modalidad distribuida que se propone en el presente trabajo. El propósito de esta práctica fue analizar comparativamente ambas modalidades a los efectos de extraer conclusiones conducentes a mejorar esta propuesta. Del análisis surgieron las siguientes apreciaciones:

1. *Ronda*: las reuniones diarias (*Rondas*) que se llevan a cabo por internet son más breves, precisas y organizadas que las que se mantienen en forma presencial, puesto que suelen ser más concretas que en el caso de reuniones físicas donde se da lugar a distracciones y charlas informales no pertinentes. Adoptar las prácticas de reunión diaria o *rondas* es una tarea extremadamente importante al comunicarse de forma remota.

2. Espacio físico: el trabajar desde el domicilio personal puede representar tanto una ventaja como una desventaja, dependiendo de las circunstancias. En una oficina de desarrollo por lo general se cuenta con mejor infraestructura, orden e higiene, dado que comúnmente en los llamados *software factory* se cuenta con personal dedicado a esta tarea. En el caso de los desarrolladores *freelance* estas condiciones no siempre están garantizadas, lo que suele ocasionar pérdidas de concentración. El trabajo desde el domicilio personal conlleva comodidades que por lo general superan a las

que se encuentran en las oficinas de desarrollo y que pueden influir positivamente en el confort del desarrollador.

3. Indumentaria: en las oficinas los trabajadores suelen verse obligados a utilizar indumentaria formal, lo que puede generar incomodidad en algunos casos y dificultar la concentración en períodos prolongados de tiempo.

4. Aprovechamiento de las comunicaciones: en una oficina la comunicación con otros miembros del equipo surge de forma espontánea y ágil. En el caso de la comunicación vía internet se requiere mayor coordinación, lo que provoca una pérdida en el dinamismo o espontaneidad. La comunicación de voz es esencial a la hora de programar de a pares en forma distribuida y *HangOut* cubre este requisito. Por su parte, la compartición de código demanda una realimentación instantánea de los pares y el uso de *Saros* facilita esta modalidad al mostrar inmediatamente los cambios efectuados por el par desarrollador.

5. *Debug*: el proceso de *debugging* utilizado en la metodología ágil distribuida es más rápido que el que se da en la metodología propuesta por *eXtreme Programming* (XP). Al operar cada desarrollador con un equipo propio, pueden ejecutar más pruebas en la misma unidad de tiempo o encarar la búsqueda del error desde perspectivas diferentes, lo que potencia el rendimiento.

6. Capacidad de los desarrolladores: el aprendizaje de un lenguaje o de una tecnología nueva depende de la capacidad autodidacta de cada desarrollador. Si bien en la metodología XP un desarrollador más experimentado puede enseñar al mismo tiempo que desarrolla, el otro miembro del equipo está más concentrado en *ver lo que el otro hace* que en *verificar si lo hace bien*. En la metodología distribuida se propone trabajar de a pares pero cada uno con su equipo personal. Esto implica una ventaja a la hora de visualizar más directamente en su propia pantalla las acciones ejecutadas por su par y, simultáneamente, disponer del equipo para efectuar otras tareas paralelas, como por ejemplo ejecutar pruebas o buscar información mediante foros u otros medios disponibles. El hecho de que cada desarrollador cuente con su

propio equipo brinda la capacidad de investigar mientras el par continúa trabajando, ya sea evaluando el código, probándolo o incluso desarrollándolo.

## Líneas de investigación y desarrollo

Los temas que se desarrollan se encuadran dentro de las siguientes líneas de investigación del I3:

1. Ingeniería del software: (a) Ingeniería del software orientada al web, (b) Métodos avanzados de producción de software y (c) Usabilidad y accesibilidad de aplicaciones y sistemas web
2. Dispositivos Móviles en Aplicaciones Empresariales.

## Resultados y objetivos

A partir de los análisis de las diferentes bibliografías consultadas y de las experiencias de desarrollo de los diferentes miembros del equipo de investigación, se proponen las siguientes pautas de trabajo distribuido para la Metodología Ágil basada en Telecomunicaciones (MATE):

1. El ciclo productivo de un entregable debe estar comprendido entre 15 y 60 días. Este producto representará una Entrega de Valor Agregado (EVA) al entregable anterior.
2. El producto se entrega debidamente testeado, por lo que cada ciclo productivo debe contemplar el tiempo necesario para ejecutar todos los tipos de pruebas, incluyendo las pruebas beta.
3. El trabajo se deberá desarrollar a distancia en modalidad semi-sincrónica, con al menos 4 horas diarias de trabajo de programación en pareja, sincrónico, y el resto de testeo individual, asincrónico.
4. El mejor lugar desde donde el cliente, en caso de existir, puede participar, es su propio ámbito de trabajo.
5. Para una mejor organización del trabajo se definen los siguientes roles, constituyendo todos ellos el *stakeholders* en el sentido aceptado del vocablo en el campo de conocimientos de las ciencias informáticas:

- a. El dueño de la idea.
- b. El facilitador.

c. El oráculo.

d. Los desarrolladores.

El dueño de la idea constituye una interfaz que vincula al “cliente” con el “equipo”.

El facilitador por su parte es el líder de equipo de desarrollo quien comienza y termina las reuniones virtuales, determina el orden del día, enuncia los puntos acordados, fija posiciones con el dueño de la idea, gestiona los recursos y acuerda la extensión de cada EVA en función de lo propuesto por los desarrolladores.

El oráculo<sup>2</sup> gestiona toda la información concerniente al proyecto, tal que cualquier integrante del *stakeholders* tenga un rápido acceso a la misma desde cualquier punto.

Los desarrolladores determinan la extensión de las tareas y tienen a su cargo el desarrollo del código y su correspondiente testeo y depuración.

Cada integrante del *stakeholders* puede ejecutar más de un rol al mismo tiempo. La idea es que estos roles puedan ser desempeñados desde cada uno de los ámbitos de los integrantes del equipo, incluyendo al cliente (de existir).

### Mecánica de trabajo propuesta

Una vez concebida una idea, a partir de la cual se pretende construir una aplicación, se constituye un equipo de análisis formado por los desarrolladores conducidos por el dueño de la idea. Se trabaja sobre el análisis del producto, intentando enunciar las funcionalidades deseadas y haciendo abstracción de su dificultad. Se elabora un documento de precedencia de funciones, en virtud del potencial de generación de utilidades de cada una, acordado con el cliente o el dueño de la idea. Los miembros del equipo de desarrollo dividen las funciones en tareas estimables, en términos de tiempo y esfuerzo, y se estipulan los plazos para la realización de las mismas. Este procedimiento se realiza una única vez por cada EVA.

Semanalmente, preferentemente todos los lunes, se realizan reuniones donde intervienen todos los involucrados, con excepción del

---

**2 Oráculo (RAE)** Persona a quien todos escuchan con respeto y veneración por su mucha sabiduría y doctrina.

cliente o dueño de la idea. En estas instancias el facilitador conduce la revisión de la asignación semanal de tareas.

En la *ronda* diaria, y en forma breve, cada par de desarrolladores y el facilitador registran el reporte de avances y de dificultades. En esta oportunidad se revisan las tareas pendientes. En caso de que la distancia física imponga una diferencia horaria tal que impida la reunión simultánea de todos los pares, el facilitador mantendrá reuniones dedicadas con cada grupo de pares que gocen de husos horarios compatibles.

La jornada de trabajo no debe superar las 8 hs. diarias. Los pares productivos deben poder trabajar al menos 4 hs en forma sincrónica por lo que sus locaciones geográficas y costumbres deben permitirlo. Se trabaja en forma colaborativa y simultánea por captura de escritorio. El tiempo de trabajo independiente se dedica a la revisión de código, prueba y depuración.

#### Aplicación de la MATE

Las recomendaciones propuestas para la nueva MATE se están aplicando actualmente en el desarrollo de una aplicación de gestión de la propia metodología, similar a la *Kunagi*, que asista a la administración del tablero, con las tareas que conforman los sprint, y que ostente la particularidad de poder ejecutarse en todo dispositivo móvil con pantalla táctil. Esta aplicación dispondrá además de los gráficos de avance, a modo de asistentes visuales del trabajo concluido y del que se encuentra pendiente. La interfaz gráfica facilitará el arrastre de las notas dentro del tablero con la facultad de actualizar los estados de cada tarea.

#### Formación de recursos humanos

En esta línea de investigación participan 3 docentes y 3 alumnos de la carrera de Sistemas de Información de la Universidad Adventista del Plata.

#### Referencias

Ableson, F. (2011). *Android: Guía para Desarrolladores* (2º Ed.) - Madrid: Anaya Multimedia.

Blanco, P., Camarero, J., Fumero, A., Warterski, A., Rodríguez, P. (2009). *Metodología de desarrollo ágil para sistemas móviles: Introducción al desarrollo con Android y el iPhone*. Universidad Politécnica de Madrid. Documento recuperado de [http://www.adamwesterski.com/wp-content/files/docsCursos/Agile\\_doc\\_TemasAnv.pdf](http://www.adamwesterski.com/wp-content/files/docsCursos/Agile_doc_TemasAnv.pdf)

Carvajal Riola, J. C. (2008). *Metodologías ágiles: herramientas y modelo de desarrollo para aplicaciones Java EE como metodología empresarial*. Tesis final de Master.

Fantasia, J., Ferrochio, D., Maldonado, C., Martinez, E., Trujillo H. (2006). Desarrollo de una metodología utilizable en la construcción de aplicaciones de tecnología Móvil. Recuperado de [http://sedici.unlp.edu.ar/bitstream/handle/10915/20785/Documento\\_completo.pdf?sequence=1](http://sedici.unlp.edu.ar/bitstream/handle/10915/20785/Documento_completo.pdf?sequence=1)

Gamma E., Helm, R., Johnson, R. y Vlissides, J. (2003). *Patrones de Diseño*. Madrid: Pearson Educación.

Letelier, P. y Penadés, M. C. (2006). *Metodologías ágiles para el desarrollo de software: eXtreme Programming (XP)*. Universidad Politécnica de Valencia. Documento recuperado de <http://www.willydev.net/descargas/masyxp.pdf>

Ray, J. y Johnson, S. (2010). *Desarrollo de Aplicaciones para Iphone*. Madrid: Anaya Multimedia.

Scalone, F. (2006). Estudio comparativo de los modelos y estándares de calidad del software. Tesis de Maestría. Recuperado de <http://laboratorios.fi.uba.ar/lsi/scalone-tesis-maestria-ingenieria-en-calidad.pdf>

Wake, W. C. (2000). *Extreme Programming Explored*. Recuperado de <http://faculty.ksu.edu.sa/MubarakRashed/Books/Extreme.Programming.Explored.pdf>



## Análisis metodológico para la utilización de Process Mining como tecnología de optimización y respaldo de la implementación de procesos de negocio bajo el marco de BPM

Virginia María Magliano<sup>1</sup>, Mg. Patricia Bazán<sup>2</sup>, Lic. José Martínez Garro<sup>3</sup>

<sup>1,3</sup> Facultad de Informática UNLP, <sup>2</sup> LINTI Facultad de Informática UNLP

[vir.magliano@gmail.com](mailto:vir.magliano@gmail.com), [pbaz@ada.info.unlp.edu.ar](mailto:pbaz@ada.info.unlp.edu.ar), [josemartinezgarro@gmail.com](mailto:josemartinezgarro@gmail.com)

### Resumen

Hoy en día las organizaciones usan tecnologías de información para soportar sus procesos de negocio siendo BPM (*Business Process Management* o Gestión de procesos de negocio) la tecnología pionera.

Para asistir la optimización del ciclo de vida de BPM, es necesario contar con una tecnología específica que se encuentre centrada en el proceso y no en los datos como la mayoría de los enfoques tradicionales. Es aquí cuando entra en juego *Process Mining* (*Minería de procesos*), una poderosa tecnología para administrar procesos operacionales no triviales. Los algoritmos y técnicas emergentes hacen posible analizar datos de eventos complejos y alinear los procesos con la información para adecuarse a los requerimientos de cliente como cumplimiento, eficiencia entre otros.

En este artículo se enuncia un enfoque para la aplicación de *Process Mining* sobre un BPMS (*Sistema de administración de procesos de negocio*) y de cómo *Process Mining* puede asistir a los procesos de negocio en todo su ciclo de vida. [1] [2] [3] [11]

**Palabras clave:** *BPM* (*Business Process Management*), *Process Mining* (*Minería de procesos*), *BAM* (*Business Activity Monitoring*) *BPMS*, *XES* (*Extensible Event Stream*), *BI* (*Business Intelligence*).

### Contexto

El presente es un trabajo de fin de carrera de Licenciatura en Sistemas de la Facultad de Informática de la UNLP, de la alumna Virginia María Magliano, dirigida por la Mg. Patricia Bazán.

### Introducción

La mayoría de las herramientas de *Business Intelligence* (BI) utilizan datos de eventos para soportar la toma de decisiones. Bajo el paraguas de BI se encuentran otras tecnologías, como BAM (Monitoreo de actividades de negocio), CPM (Gestión del

rendimiento corporativo), CPI (Mejora continua de procesos), and BPI (Inteligencia de procesos de negocio) que permiten realizar reportes y *dashboards* (*tableros de mando*), pero solo se focalizan en los datos y no en el proceso de inicio a fin.

Por otro lado los BPMS utilizan modelos de proceso para analizar procesos operacionales. Por lo general estos modelos se encuentran desconectados de los datos de eventos reales, y por lo tanto los resultados pueden ser no confiables ya que los mismos se basan en un modelo idealizado y no en los hechos observados. *Process Mining* busca cubrir la brecha entre BPM y BI combinando datos de eventos y modelos de proceso. [3] [2] [9] [10]

La idea básica de *Process Mining* es extraer conocimiento de logs de eventos extraídos de distintos sistemas de información, para esto posee un conjunto de técnicas que se agrupan según su funcionalidad en:

- Técnicas de descubrimiento de proceso
- Técnicas de chequeo de concordancia del proceso con la realidad
- Técnicas de extensión y mejora del proceso

#### a. Análisis ciclo de vida de procesos en BPM y vinculación con Process Mining

El ciclo de vida de procesos en BPM está compuesto por cuatro etapas principales:

- Diagnóstico y obtención de requerimientos
- (re) Diseño
- Implementación
- Ejecución/Monitoreo

En la fase de diseño de procesos se elabora un modelo representativo de los requerimientos.

Este modelo se transforma en un sistema ejecutable en la etapa de configuración/implementación.

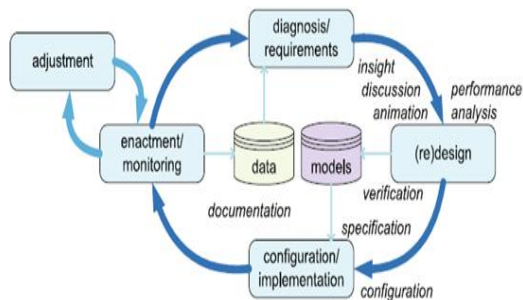
Luego comienza la fase de monitoreo, en esta etapa los procesos se ejecutan y mientras

se monitorean por administradores para ver si algún cambio es necesario.

Algunos de los cambios son tomados en la etapa de ajuste. En esta etapa no se crea nuevo software ni se rediseña el existente, solo se utilizan controles predefinidos para adaptar o reconfigurar el proceso.

En la etapa de diagnóstico y requerimientos se evalúa el proceso y se monitorea requerimientos emergentes debido a cambios en el entorno del proceso.

Una pobre performance o la imposición de nuevas demandas del medio ambiente pueden generar una nueva iteración en el ciclo de vida de BPM, comenzando en la fase de rediseño.



**Figura 1. Ciclo de vida de un proceso de negocio [3]**

Como puede verse en la **Figura 1** el modelo de proceso juega un papel fundamental en las etapas de (re)diseño y configuración/implementación, mientras que los datos tienen un rol dominante en las etapas de monitoreo y diagnóstico y obtención de requerimientos.

Hasta hace un tiempo atrás, no había conexión entre los datos producidos por la ejecución de los procesos y el proceso de diseño actual. Además, el ciclo de BPM solo se reiniciaba cuando había un cambio externo muy importante.

*Process Mining* aplica sus técnicas basándose en estos dos pilares: datos y modelos.

Para la **etapa de diagnóstico y obtención de requerimientos** *Process Mining* brinda técnicas que permiten hacer un sondeo sobre el proceso y obtener información general del mismo tal como: el *log inspector* y el *dotted chart*.

Para la **etapa de diseño** brinda técnicas con distintos grados de complejidad de descubrimiento de patrones que permiten descubrir un modelo de proceso.

Para la **etapa de configuración/ implementación** *Process Mining* posee técnicas que permiten extender el modelo y agregar las perspectivas organizacionales, de tiempo y de caso.

Para la **etapa de ejecución y monitoreo** brinda distintos tipos de técnicas como diagramas de transiciones, modelos *fuzzy* y técnicas de análisis BPM que permiten monitorear el proceso, hacer predicciones, analizar la performance, detectar cuellos de botella, entre otras tareas. [2][3][8]

## b. Enfoque para aplicación de *Process Mining* sobre un BPMS

El enfoque propuesto por *Process Mining* para obtener un modelo completamente integrado cubriendo todos los aspectos relevantes del proceso (**Figura 2**), es aplicable a cualquier organización donde no se tenga siquiera un proceso, o en alguna donde se tenga uno y el mismo se desea extender y mejorar. El enfoque consiste de cinco pasos:

- **Paso 1:** obtener el log de eventos. Aquí es cuando se extraen los datos que conformarán el log de una variedad de sistemas de información, esta acción debe ser llevada a cabo mediante una herramienta generadora de log de eventos. En este paso es fundamental tener un panorama completo del proceso y un conocimiento de la ubicación de los datos.
- **Paso 2:** crear o descubrir un modelo de proceso, este paso puede realizarse utilizando alguna de las técnicas de descubrimiento de procesos provistas por *Process Mining* o bien ser creado convencionalmente como por ejemplo por los analistas del proceso.
- **Paso 3:** conectar eventos en el *log* a actividades en el modelo: este paso es esencial para proyectar información en los modelos y agregar perspectivas. Utilizando la técnica de *replay* se conectan los eventos y las actividades en el modelo.
- **Paso 4:** extender el modelo:
  - **Agregar la perspectiva organizacional:** es posible analizar la red de trabajo social y subsecuentemente identificar entidades organizacionales que conectan actividades en grupos de recursos
  - **Agregar la perspectiva tiempo:** los *timesteps* y frecuencias se pueden usar para aprender distribuciones de probabilidad que describen adecuadamente los tiempos de espera y tiempos de servicios y probabilidad de ruteo.

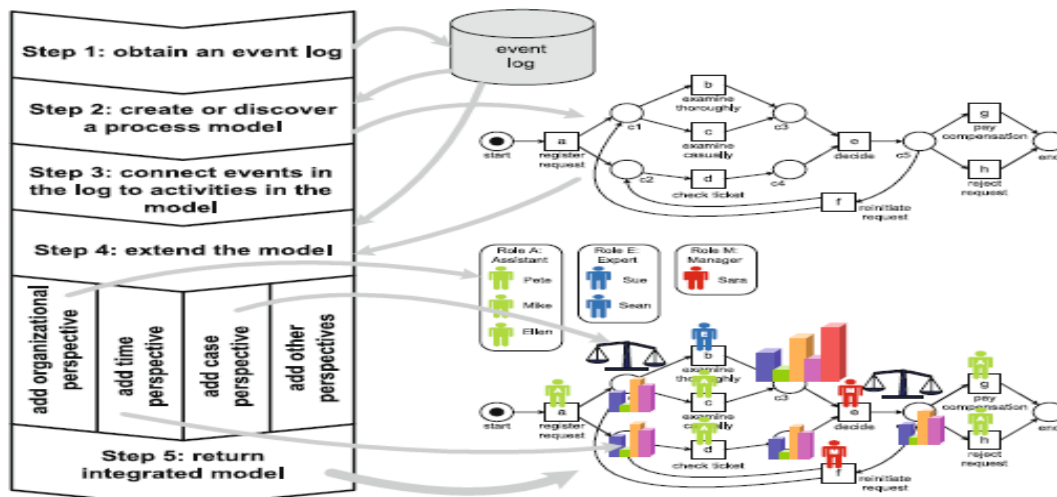


Figura 2 Enfoque para obtener un modelo completamente integrado que cubre la perspectiva organizacional, de tiempo y de caso. [3]

- **Agregar la perspectiva de caso:** por medio de Minería de Decisión se pueden usar los atributos en el log. Esto muestra como los datos son relevantes y deben ser incluidos en el modelo.
- **Agregar otra perspectiva:** dependiendo de la información en el log otras perspectivas se pueden agregar al modelo. Por ejemplo, información y riesgos y costos se pueden agregar al modelo.
- **Paso5:** devolver el modelo integrado, listo para ser analizado con las distintas técnicas de Process Mining.

El modelo integrado resultante provee una vista holística del proceso. Esto provee nuevos puntos de vista y puede generar varias ideas para el mejoramiento del proceso. [2] [3]

Si quisiéramos implementar un proceso en un BPMS desde el principio, obteniendo el log de eventos de los distintos sistemas de información en los que el proceso puede haber dejado un "rastros" y mediante la aplicación del enfoque propuesto de *Process Mining*, se obtendrá un panorama completo del proceso ayudando a la implementación del mismo.

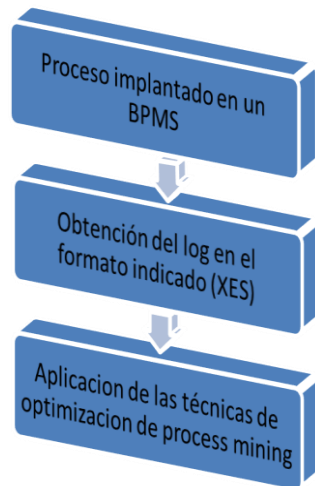
Sin embargo, si quisiéramos aplicar el enfoque a un proceso ya implantado en un BPMS e utilizar las técnicas de *Process Mining* para la optimización del mismo, el enfoque sería distinto por varias razones:

- Ya se dispone de un modelo de proceso.

- Como el proceso ya se encuentra en producción, el mismo genera un historial en la base de datos del BPMS mismo, y este estaría compuesto por la instancia de proceso y las variables de la misma.
- La conexión de los eventos con las actividades del modelo se realiza innatamente por el BPMS, la base de datos del mismo almacena las instancias de procesos ejecutadas registrando para cada evento a que actividad del modelo corresponde.
- Un BPMS ya contempla las cuatro perspectivas: La perspectiva organizacional ya se encuentra relacionada al modelo porque cada actividad o *lane* debe tener un actor. En cuanto a la perspectiva de tiempo un BPMS siempre registra con *timestamps* la fecha de ocurrencia de todos los eventos. La perspectiva de caso está cubierta por las variables de proceso y de actividad que difieren de un caso a otro.

Se puede concluir entonces que un proceso implementado en un BPMS es un buen escenario para la aplicación de *Process Mining* ya que el procesamiento que se debe realizar para aplicar las técnicas es mínimo.

En definitiva, la problemática se reduce a construir el log de eventos en el formato indicado para poder ser importado en una herramienta de *Process Mining* y aplicarle las distintas técnicas de optimización, chequeo de concordancia y performance que nos brinda *Process Mining*.



**Figura 3 Procedimiento para aplicar las técnicas de *Process Mining* a un proceso implantado en un BPMS.**

Como se muestra en la **Figura 3** el procedimiento para aplicar las técnicas de optimización de *Process Mining* a un proceso implantado en un BPMS es el siguiente:

- En una primera etapa se implementó el proceso en un BPMS con o sin asistencia de las técnicas de *Process Mining*, y ahora se quiere utilizar la misma para medir la performance, monitorear y optimizar el proceso.
- En la segunda etapa se debe tener un conocimiento de la ubicación de los datos en la base de datos del BPMS. Luego mediante una herramienta de generación de logs de eventos en formato XES extraen los datos y se genera el log.
- Una vez generado el log mediante una herramienta de *Process Mining* se puede importar el mismo y se le pueden aplicar las técnicas de optimización y monitoreo.

Como conclusión nos podemos encontrar con dos escenarios para la aplicación de *Process Mining* sobre un BPMS:

- Primer escenario: No se encuentra modelado el proceso para la cual se recomienda aplicar el primer enfoque mencionado de manera de obtener una vista general del proceso y luego implementar el proceso en el BPMS.
- Segundo escenario: el modelo ya se encuentra implementado y corriendo y se desea utilizar las técnicas como método de optimización, para

chequear la credibilidad del proceso, analizar la performance del mismo, y monitorearlo.[1][2][3][8]

## LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Las líneas de investigación que desprende esta investigación abarcan en primer lugar la capacidad de utilizar *Process Mining* como herramienta que de soporte tanto en la implementación como en la optimización de procesos en el marco de un BPMS. Por otro lado queda como línea de desarrollo encontrar alguna forma de generar el log en formato XES a partir de los datos del BPMS.

## RESULTADOS Y OBJETIVOS

La metodología de *Process Mining* propuesta fue aplicada en procesos en producción existentes tales como el citado en [6] y [7], con resultados altamente positivos. En la actualidad el objetivo radica principalmente en la utilización de las herramientas de *Process Mining* (tecnología completamente centrada en el proceso) como herramienta de soporte y optimización de los BPMS. Otro foco principal radica en la automatización de la integración entre BPM y *Process Mining* para de esta forma sistematizar la obtención de información que se requiere para la aplicación de esta metodología.

## FORMACIÓN DE RECURSOS HUMANOS

BPM brinda un nuevo enfoque en la ciencia de los servicios de información, requiriendo una capacitación y formación de recursos humanos relacionados con esta tecnología.

El presente trabajo se enmarca en una línea de investigación en *Process Mining* y BPM donde se están formando alumnos para desarrollar su tesina e interactuar con investigadores formados con el objeto de incorporar herramientas de soporte de esta línea de trabajo para solucionar problemas reales.

## REFERENCIAS

- [1] "Business Process Management: concepts, Languages, Architectures". Mathias Weske Springer-Verlag Berlin Heidelberg 2007
- [2] "Process Mining Conformance and Extension". O. by Anne Rozinat. - Eindhoven: Technische Universiteit Eindhoven, 2010. - Proefschrift.



[3] “Process Mining Discovery, Conformance and Enhancement of Business Processes”. Wil M.P. van der Aalst Department Mathematics & Computer Science Eindhoven University of Technology Den Dolech 2 5612 AZ Eindhoven The Netherlands. 2011.

Facultad de Informática UNLP , LINTI Facultad de Informática

[4] “Using Process Mining to Bridge the Gap between BI and BPM”. Wil M.P. van der Aalst Department Mathematics & Computer Science Eindhoven University of Technology Den Dolech 2 5612 AZ Eindhoven The Netherlands. 2011.

[5] “Process Mining: The next step in Business Process Management”. Wil M.P. van der Aalst Department Mathematics & Computer Science Eindhoven University of Technology Den Dolech 2 5612 AZ Eindhoven The Netherlands.

[6] “Business Process Mining: An Industrial Application”. W.M.P. van der Aalst, H.A. Reijers, A.J.M.M. Weijters, B.F. van Dongen, A.K. Alves de Medeiros, M. Song; and H.M.W. Verbeek Department of Technology Management, Eindhoven University of Technology, P.O. Box 513, NL-5600 MB, Eindhoven, The Netherlands. [w.m.p.v.d.aalst@tm.tue.nl](mailto:w.m.p.v.d.aalst@tm.tue.nl) Dept. of Industrial Engineering, Pohang University of Science and Technology, San 1 Hyoja-Dong, Nam-gu, Pohang, 790-784, South Korea.

[7] “Modelado y análisis de un proceso quirúrgico mediante técnicas de minería de procesos”. 4th International Conference on Industrial Engineering and Industrial Management XIV Congreso de Ingeniería de Organización Donostia- San Sebastián , September 8th -10th 2010

[8] “Process Mining: Overview and Opportunities”. Wil Van der Aalst, Eindhoven University of Technology

[9] “Process Mining Project Methodology: Developing a General Approach to Apply Process Mining in Practice”. T.H.C. VAN DER HEIJDEN. BSc Industrial Engineering — TU/e 2011 Student identity number 0611037.

[10] “Business Process Mining: From Theory to Practice”. Turner CJ, Tiwari A, Olaiya R, Xu Y. Business Process Management Journal 2012 Vol. 18 Iss:3, pp.493-512

[11] “Mejora metodológica de procesos de negocio mediante el uso intensivo de BI y KM”. Lic. José Martínez Garro1, Mg. Patricia Bazán2

# Técnicas y Herramientas para Desarrollo de Sitios Web Accesibles

Adriana Martín<sup>1,2</sup>, Gabriela Gaetán<sup>1</sup>, Viviana Saldaño<sup>1</sup>, Gabriela Miranda<sup>1</sup>, Sabrina Pastrana<sup>1</sup>, Diego Vilte<sup>1</sup>, Estefanía Gómez Vega<sup>1</sup>

<sup>1</sup>Unidad Académica Caleta Olivia  
Universidad Nacional de la Patagonia Austral (UNPA)  
Acceso Norte Ruta 3, (9011) Caleta Olivia, Santa Cruz, Argentina

<sup>2</sup>Grupo de Investigación en Ingeniería de Software (GIISCo)  
Departamento de Ingeniería de Sistemas  
Facultad de Informática  
Universidad Nacional del Comahue (UNComa)  
Buenos Aires 1400, (8300) Neuquén, Argentina

e-mails: adrianaelba.martin@gmail.com;  
ggaetan / vivianas / gmiranda @uaco.unpa.edu.ar; sabrina.pastrana@gmail.com;  
dvilte773@yahoo.com.ar; chefito\_anime@hotmail.com

## Resumen

En Junio de 2011, la República Argentina dio sanción a la ley 26.653, denominada “Guía de Accesibilidad para Sitios Web del Sector Público Nacional”, la cual adhiere a las Directrices de Accesibilidad al Contenido Web, versión 1.0 (WCAG 1.0). Sin embargo, aunque dictaminado por la legislación vigente, las instituciones y organizaciones argentinas tienen aún un largo camino por recorrer para ofrecer sitios Web más accesibles a los ciudadanos. Es desalentador comprobar, que en general, los líderes al frente de los equipos de desarrollo Web desconocen y/o desestiman la relevancia de la Accesibilidad. Se necesitan enfoques más proactivos que despierten la conciencia de Accesibilidad y propicien la incorporación de este atributo de calidad entre los top 10 del desarrollo de productos TIC.

Enmarcados por esta realidad, nuestro Proyecto de Investigación tiene por objetivo desarrollar técnicas y herramientas que propicien la inclusión de los requerimientos de Accesibilidad en las prácticas de diseño,

como así también las evaluación y re-diseño de portales, sitios y aplicaciones Web.

**Palabras clave:** *Accesibilidad Web, Diseño temprano de RNF, Técnicas y Herramientas de Evaluación y Re-Diseño IU.*

## 1. Contexto

Nuestra línea de investigación se encuentra enmarcada en el Proyecto de Investigación N° 29/B144, denominado “Diseño y Evaluación de Portales Web” dirigido por la Dra. Adriana Martín y codirigido por la Mg. Gabriela Gaetán, del Departamento de Ciencias Exactas y Naturales, Unidad Académica Caleta Olivia, Universidad Nacional de la Patagonia Austral. Período: 2012-2014. El presente Proyecto tiene sus propios objetivos de investigación pero funda su trabajo en los resultados ya alcanzados en el área de mejora de la Accesibilidad de sitios Web aplicable a catálogos de componentes para Sistemas de Información Geográficos (SIG), por el Proyecto de Investigación N° 29/B107, denominado

“Mejora del Proceso de Selección de Componentes para Sistemas de Información Geográficos” dirigido por la Dra. Alejandra Cechich y la Mg. Gabriela Gaetán, del Departamento de Ciencias Exactas y Naturales, Unidad Académica Caleta Olivia, Universidad Nacional de la Patagonia Austral. Período: 2010-2012. Con Informe de Avance Aprobado e Informe Final Entregado.

Desde el comienzo de las actividades de investigación, este equipo de investigación mantiene una estrecha relación con el grupo de investigación GIISCo perteneciente al Departamento de Ingeniería de Sistemas de la Facultad de Informática de la UNComa, principalmente por medio del soporte en asesoría externa brindado por la Dra. Alejandra Cechich, docente-investigador Categoría I del Programa de Incentivos.

## 2. Introducción

La Accesibilidad permite el acceso universal a la Web para todos los usuarios independientemente de sus posibles limitaciones, tanto personales (algún tipo de discapacidad, edad, nivel cultural, etc.), como derivadas del entorno (iluminación, ruido, etc.) o de la tecnología empleada (dispositivo, velocidad de conexión, software, etc.).

En los últimos años, la Accesibilidad Web se ha convertido en varios países del mundo en una preocupación porque atañe directamente a la posibilidad de acceso de los ciudadanos a la información, comunicación y servicios (públicos y privados) ofrecidos a través de la Web. Es así como el consorcio W3C ha desarrollado recomendaciones, denominadas Directrices de Accesibilidad al Contenido Web, versión 1.0 (WCAG 1.0) [19] y versión 2.0 [20], que son recomendaciones de Accesibilidad consideradas como normas de facto y citadas como referencia obligada en la mayoría de las legislaciones sobre Tecnologías de la Información y Comunicación (TICs) de todo el mundo. Por ejemplo, la Section 508 de Estados Unidos [18], el PAS 78 del Reino

Unido [16], la Legislación Italiana [17], y la ley 26.653 denominada “Guía de Accesibilidad para Sitios Web del Sector Público Nacional”<sup>1</sup> de Argentina, entre otras.

Si bien existen varias técnicas y/o herramientas que asisten a la producción y evaluación de contenido accesible, la oferta se debilita cuando se trata de propuestas que permitan aplicar estos instrumentos de Accesibilidad al terreno de una implementación concreta y para mejorar la experiencia de los usuarios Web en los diferentes dominios de aplicación.

A más de doce años en que el consorcio W3C lanzó las recomendaciones WCAG 1.0, el balance sobre el estado de la WWW desde la perspectiva de accesibilidad, es poco alentador. A pesar de los esfuerzos invertidos por diversos organismos y la producción propuesta por los grupos de investigación, la conformidad a la Accesibilidad sigue siendo una propiedad difícil de alcanzar por los productos desarrollados para la Web. En la Web coexisten aplicaciones con diferentes grados de complejidad que van desde aplicaciones de hipertexto Web (presencias Web básicas) hasta aplicaciones Web propiamente dichas (sistemas Web complejos) con escalas intermedias en las denominadas aplicaciones de Software Web.

En este contexto, los expertos en Accesibilidad deben no sólo proveer buenas prácticas de diseño, evaluación y/o re-diseño accesible, sino que también, propuestas concretas para transferir las mismas al desarrollo de los diferentes productos requeridos por la Web.

Si bien existen numerosas fuentes de información referidas a la Accesibilidad Web como así también herramientas específicas para ayudar a evaluar la Accesibilidad de los sitios Web [6][1][2], aún siguen siendo escasas las propuestas que atienden al diseño [3][4][21][14][15] para dar respuesta temprana a los problemas de Accesibilidad que normalmente surgen en la interfaces de

<sup>1</sup> Ver Ley 26.653

[http://www.infoleg.gov.ar/infolegInternet/\\_anexos/175000-179999/175694/norma.htm](http://www.infoleg.gov.ar/infolegInternet/_anexos/175000-179999/175694/norma.htm)

usuario (IU) cuando se desarrollan aplicaciones Web complejas.

En el desarrollo de aplicaciones Web accesibles, el método denominado "Aspect-Oriented Web Accessibility Design" (AO-WAD), permite considerar propiedades de Accesibilidad en etapas tempranas del desarrollo de aplicaciones para la Web [5][12][9]. La propuesta se materializa en un proceso que combina herramientas de diseño conceptual conocidas y probadas (en la Ingeniería del Software y la Ingeniería Web) y conceptos del paradigma Orientado a Aspectos. La elección de este paradigma permite tratar de manera adecuada las características propias de la naturaleza de la Accesibilidad como requerimiento genérico y no-funcional. El hilo conductor de AO-WAD es el proceso de Desarrollo Dirigido por Modelos dentro del cual se consideran los "concerns" de Accesibilidad desde los requerimientos y a través del diseño a la implementación de la aplicación Web con interfaces de usuario "accesibles". AO-WAD provee soporte al desarrollo de aplicaciones Web, es decir, un sistema con plataforma de despliegue en la Web que combina las características de las aplicaciones de hipermedia Web y de las aplicaciones software Web.

### 3. Líneas de Investigación y Desarrollo

El perfil de nuestro grupo de investigación se define en base a las actividades vinculadas a Investigación y Transferencia en temáticas relacionadas con la Ingeniería de Software y en particular, con la Ingeniería Web.

Actualmente y en el marco del presente Proyecto de Investigación, estamos abocados a extender el método AO-WAD para mejorar el tratamiento temprano de los requerimientos de Accesibilidad Web mediante el diseño y la evaluación y en esta dirección estamos abordando los siguientes tópicos:

- Personalización del método AO-WAD en distintos dominios de aplicación y/o

contextos de navegación.

- Procesos de evaluación y monitoreo del cumplimiento de estándares de accesibilidad en sitios Web del Sector Público Nacional en Argentina.
- Transferencia e implementación del método AO-WAD integrado.

### 4. Resultados y Objetivos

El objetivo del proyecto actual de investigación es desarrollar técnicas y herramientas que mejoren las prácticas de diseño y evaluación de portales, sitios y aplicaciones Web, enfocándonos en características particulares concernientes a factores de calidad, tales como la Usabilidad y la Accesibilidad.

En esta dirección, nos encontramos actualmente trabajando sobre la extensión del método AO-WAD para que pueda ser utilizado en conjunto con métodos tradicionales de la Ingeniería Web y de esta manera contribuir desde el diseño, rediseño y/o evaluación a mejorar el acceso de los usuarios al conocimiento y servicios desplegados en la WWW.

Para alcanzar este objetivo direccional, nuestro Proyecto de Investigación ha fijado los siguientes objetivos operacionales:

- Generalizar los elementos y las relaciones definidas en el método AO-WAD.
- Diseñar recomendaciones que faciliten el uso del método AO-WAD en métodos de la Ingeniería Web.
- Diseñar recomendaciones para su transferencia y uso en la industria.
- Diseñar e implementar herramientas de soporte al método integrado.
- Personalizar según dominios de casos de estudio y/o contextos de navegación.
- Validar mediante casos de estudio.

Nuestro grupo de investigación ha estado trabajando fuertemente en desarrollo de productos de calidad para la Web, orientados específicamente a mejorar la Accesibilidad. Como base de este trabajo se tomó el



enfoque AO-WAD propuesto en trabajos previos [12][5][9]; y los resultados de las actividades de investigación y desarrollo actuales han quedado plasmados en las siguientes contribuciones científicas [7][8][11][10][13].

## 5. Formación de Recursos Humanos

Nuestro grupo de investigación reúne a 5(cinco) investigadores, entre docentes y alumnos de postgrado de la UNPA; a partir de 2013 se incorporan 2(dos) becarios alumnos de la carrera de grado Ingeniería en Sistemas de la UNPA y se obtuvo una tercera beca de postgrado para 1(un) integrante, Ing. Gabriela Miranda, que es alumna de la Maestría en Informática y Sistemas en la UNPA. La directora y codirectora de nuestro grupo de investigación, Dra. Adriana Martín y Mg. Gabriela Gaetán respectivamente, son Categoría III en el Programa de Incentivos y el grupo cuenta además con 1(un) docente-investigador, Ing. Viviana Saldaño, con Categoría IV. Los integrantes que conforman nuestro grupo de investigación han finalizado en 2012 sus carreras de postgrado o se encuentran haciendo carreras de postgrado, estos últimos ya sea en etapa próxima a terminar o con un grado de avance significativo.

Durante el 2012, 1(un) investigador defendió su tesis de Doctorado en Ciencias Informáticas en la UNLP y 1(un) investigador defendió su tesis de Maestría en Ingeniería de Software en la UNLP. En marzo de 2013, 1(un) investigador defenderá su tesis de Maestría en Ingeniería de Software en la UNLP.

También durante 2012, 2(dos) integrantes, inscriptos en la Maestría en Informática y Sistemas en la UNPA, han reunido todos los créditos y ambos han avanzado en el desarrollo de la Tutoría de Investigación requerida por el plan de estudios de la Maestría UNPA y se encuentran en la etapa de definición de sus temas de tesis.

Tesis de Grado dirigidas y aprobadas en 2012: 3(tres). Tesis de Postgrado y Grado con dirección en curso desde 2012: Postgrado (2 tesista), Grado (1 tesista).

## 6. Agradecimientos

A la UNPA por el soporte al Proyecto de Investigación en curso N° 29/B144 "Diseño y Evaluación de Portales Web".

## 7. Referencias

- [1] Brajnik, G. *Barrier Walkthrough - Heuristic evaluation guided by accessibility barriers*. <http://users.dimi.uniud.it/~giorgio.brajnik/projects/bw/bw.html>
- [2] Brajnik, G. *Beyond Conformance: The Role of Accessibility Evaluation Methods*. In: S. Hartmann et al. (Eds.) WISE 2008. LNCS vol. 5176, pp. 63–80. Springer-Verlag Berlin Heidelberg (2008)
- [3] Casteleyn, S., Van Woensel, W., Houben, G-J. *A Semantics-based Aspect-Oriented Approach to Adaptation in Web Engineering*. In HT (2007) [doi.acm.org/10.1145/1286240.1286297](https://doi.org/10.1145/1286240.1286297)
- [4] Centeno, V., Kloos, C., Gaedke, M., Nussbaumer, M. *Web Composition with WCAG in Mind*. W4A (2005) [doi:10.1145/1061811.1061819](https://doi.org/10.1145/1061811.1061819)
- [5] Martín, A., Cechich, A., Rossi, G. *Accessibility at Early Stages: Insights from the Designer Perspective*. In 8th International Cross-Disciplinary Conference on Web Accessibility, (Hyderabad, Andhra Pradesh, India, 2011), ACM, 9, [doi:10.1145/1969289.1969302](https://doi.org/10.1145/1969289.1969302)
- [6] Martín, A., Cechich, A., Rossi, G. *Comparing Approaches to Web Accessibility Assessment*. In: Handbook of Research on Web Information Systems Quality, pp. 181-205. Information Science Reference, Hershey NY (2008)

- [7] Martin, A., Gaetan, G., Saldaño, V., Miranda, G., Molina, S., Pastrana, S. *Diseño y Evaluación Tempranos para Priorizar la Accesibilidad en la WWW*. Actas del XVI Workshop de Investigadores en Ciencias de la Computación (2012)
- [8] Martin, A., Gaetan, G., Saldaño, V., Miranda, G., Molina, S., Pastrana, S. *Propiciando el Diseño y la Evaluación de Sitios para una WWW Accesible*. 2º Encuentro de Investigadores de la Patagonia Austral, Universidad Nacional de la Patagonia Austral, Unidad Académica San Julián. (2012)
- [9] Martín, A., Mazalu, R., Cechich, A. *Supporting an Aspect-Oriented Approach to Web Accessibility Design*. IEEE ICSEA'10, pp. 20-25. (2010)
- [10] Martin, A., Miranda, G., Saldaño, V., Gaetan, G. *AO-WAD: A Generalized Approach for Accessible Design within the Development of Web-Based Systems*. The Seventh International Conference On Software Engineering Advances, ICSEA 2012. Lisbon, Portugal. ISBN: 978-1-61208-230-1. (2012)
- [11] Martin, A., Miranda, G., Saldaño, V., Gaetan, G. *AO-WAD: A Proposal for Accessible Design within Web Engineering Approaches*. Actas del XVIII Congreso Argentino en Ciencias de la Computación, Universidad Nacional del Sur, Bahía Blanca. (2012)
- [12] Martín, A., Rossi, G., Cechich, A., Gordillo, S. *Engineering Accessible Web Applications: An Aspect-Oriented Approach*. World Wide Web Journal, Vol. 13, Springer. pp. 419-440. (2010)
- [13] Miranda, G., Martin, A., Gaetan, G. *Mejora de la Accesibilidad Web mediante el Uso de Agentes Inteligentes. Una Revisión del Estado del Arte*. ICT UNPA (en revisión).
- [14] Moreno, L., Martinez, P., Ruiz, B. *A MDD Approach for Modeling Web Accessibility*. IWOST (2008) doi:10.1.1.163.9478
- [15] Plessers P., Casteleyn S., Yesilada Y., De Troyer O., Stevens R., Harper S., Goble C. *Accessibility: A Web Engineering Approach*. WWW (2005) doi:10.1145/1060745.1060799
- [16] Publicly Available Specification: A Guide to Good Practice in Commissioning Accessible Websites (2006) en <http://www.hoboweb.co.uk/seo-blog/pas-78/>
- [17] Stanca Law (2004) en [http://www.pubbliaccesso.it/biblioteca/documentazione/guidelines\\_study/index.htm](http://www.pubbliaccesso.it/biblioteca/documentazione/guidelines_study/index.htm)
- [18] US Government Electronic and Information Technology Accessibility Standards (2003) en <http://www.section508.gov/index.cfm?fuseAction=stdsdoc>
- [19] Web Content Accessibility Guidelines 1.0 (WCAG 1.0) en <http://www.w3.org/TR/WCAG10/>
- [20] Web Content Accessibility Guidelines 2.0 (WCAG 2.0) en <http://www.w3.org/TR/WCAG/>
- [21] Zimmermann, G., Vanderheiden, G. *Accessible Design and Testing in the Application Development Process: Considerations for an Integrated Approach*. Universal Access in the Information Society 7(1-2), 117-128 (2008).

## Trazabilidad de Versiones en Ingeniería de Requisitos

Andrea F. Vera<sup>(1)</sup>, Graciela D. S. Hadad<sup>(1)</sup>, Jorge H. Doorn<sup>(1,2)</sup>

<sup>(1)</sup>DIIT, Universidad Nacional de La Matanza, Florencio Varela 1903, San Justo

<sup>(2)</sup>INTIA, Dto de Computación y Sistemas, Facultad de Ciencias Exactas, UNCPBA, Pinto 399, Tandil  
{[avera\\_ghadad](mailto:avera_ghadad@ing.unlam.edu.ar)}@[ing.unlam.edu.ar](mailto:ing.unlam.edu.ar), [jdoorn@exa.unicen.edu.ar](mailto:jdoorn@exa.unicen.edu.ar)

### Resumen

La trazabilidad persiste como un aspecto relevante y difícil de tratar en la práctica de la Ingeniería de Software. El presente proyecto considera tratar en forma combinada la trazabilidad y el manejo de versiones en el caso de modelos de requisitos. Es un hecho observable que las trazas fluyen esencialmente del proceso del negocio hacia los requisitos, mientras que los cambios de versiones suelen tener origen en fenómenos casi independientes de este flujo. Esta ortogonalidad puede ser muy bien descripta mediante un modelo conjunto de trazas y versionado. Este modelo debe ser acompañado de guías de producción de trazas para modelos basados en lenguaje natural. La idea primordial es que el modelo de trazas esté desacoplado de los modelos del problema y que la producción de trazas se realice simultáneamente con el modelado de requisitos y en forma automática o semi-automática. Por otro lado, se manejan los cambios en los modelos de requisitos generando concurrentemente cambios en las trazas, versionando luego las trazas.

**Palabras clave:** gestión de requisitos, modelos de requisitos, trazabilidad, sistema de versionado.

### Contexto

Esta propuesta es parte del proyecto de investigación "Adaptabilidad y Completitud en Procesos de Requisitos" del Departamento de Ingeniería e Investigaciones Tecnológicas de la UNLaM.

### Introducción

La importancia del adecuado

conocimiento de los requisitos de un sistema de software y su impacto sobre todo el proceso de desarrollo ha sido estudiada a lo largo de varias décadas. Muchas y muy importantes contribuciones se han producido como consecuencia de una gran cantidad de esfuerzos de investigación, de actividades de normalización y de aplicación práctica de modelos, normas y heurísticas disponibles.

El proceso de la ingeniería de requisitos debería incluirse en los modelos de procesos de software. Por ejemplo, el proceso de requisitos propuesto por Leite et al. [Leite 04] consiste en, primero, la construcción de un glosario denominado Léxico Extendido del Lenguaje (LEL) y de un conjunto de escenarios actuales (EA) que representan situaciones observables en el universo de discurso; segundo, construir escenarios que representan situaciones proyectadas con el nuevo sistema del software, denominados escenarios futuros (EF); y finalmente, definir las especificaciones de requisitos del sistema de software (ER) en base al conocimiento adquirido y modelado en las etapas previas.

Dada la evolución de los requisitos, es indispensable gestionar los mismos [Leite 97]. Ello involucra básicamente la administración de las dependencias entre los mismos y la administración de las vinculaciones entre el documento de requisitos y otros documentos, modelos y componentes del software, como también hacia sus orígenes [Palmer 97] [Kotonya 98] [Davis 99]. La trazabilidad de los requisitos es un tópico de la gestión de requisitos que se encarga de mantener la evolución de estos a través del ciclo de vida del software en ambas direcciones: hacia adelante en las siguientes etapas del proceso de desarrollo y mantenimiento, y hacia atrás hasta sus orígenes. La trazabilidad de requisitos no solo concierne con la gestión de cambios

sino también ayuda en la verificación y validación de requisitos, en el control del proceso de software [Davis 99] [Palmer 97] y en la detección de conflictos, asegura la consistencia entre decisiones tempranas y tardías, y permite controlar la alocaación de requisitos, entre otras facilidades.

A pesar de los múltiples propósitos que cubre la trazabilidad de requisitos, a menudo se aplica parcialmente en la práctica. Esto no solo se debe a los altos costos de producción y mantenimiento de las trazas, sino también al esfuerzo de definir las trazas necesarias.

Los problemas en la trazabilidad, que aún persisten se deben principalmente a la falta de herramientas automáticas o semi-automáticas para identificar y mantener trazas [Cleland-Huang 03] [De Lucía 07]. Dado que la recolección y mantenimiento de la información de rastreo es de muy alto costo, se deben tener políticas que indiquen qué tipo de rastreos se realizarán y cómo se mantendrá dicha información.

El trabajo de Huffman Hayes et al. [Huffman Hayes 07] se basa principalmente en la recuperación automática de trazas para generar una matriz de trazabilidad de requisitos. Se utiliza una herramienta que produce vínculos candidatos entre dos artefactos textuales, permitiendo que manual-mente se eliminen vínculos y se agreguen nuevos. Antonioli et al. [Antonioli 02] y Marcus & Maletic [Marcus 03] han presentado propuestas similares, permitiendo la generación automática de trazas desde artefactos textuales hacia código fuente, mientras que la propuesta de Cleland-Huang et al. [Cleland-Huang 05] genera automáticamente vínculos entre artefactos gráficos (Grafo de Interdependencia de Softgoals y diagramas UML). En todas estas propuestas las trazas se producen desde artefactos ya existentes.

Estos autores proponen la generación de vínculos en tiempo de ejecución, es decir, el vínculo se crea cuando se necesita recuperar una traza. Esta modalidad evita el constante mantenimiento de vínculos debido a la evolución, aunque afecta la velocidad de recuperación de trazas.

## Líneas de investigación y desarrollo

En este proyecto se detallará un mecanismo de generación de trazas que acompañe a cada actividad del proceso de requisitos junto con un mecanismo de versionado de trazas, el cual permitirá mantener la historia de los cambios. Para ello, se refinará el modelo de trazas desarrollado inicialmente [Doorn 10].

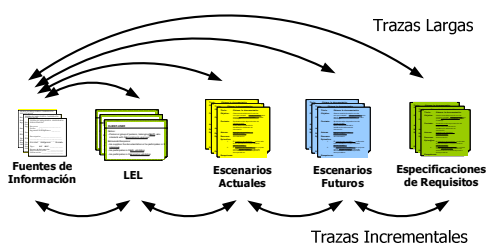
Este modelo de trazas apoya el proceso de requisitos de Leite et al. [Leite 04]. Las ER sirven de ancla para la pre y post trazabilidad, permitiendo el rastreo de los requisitos hacia sus orígenes (vinculando las ER con los EF, estos con los EA y con los símbolos del LEL) y hacia adelante a artefactos externos al proceso de requisitos (modelos de diseño, código, casos de prueba, entre otros).

Un tema relevante es definir y seleccionar el tipo de traza a utilizar. Existen dos tipos de trazas posibles: trazas incrementales y trazas largas (Figura 1). Una traza larga puede ir directamente de una fuente de información (FI) a un EA (abarca la FI, el símbolo del LEL y el EA), mientras que una traza incremental para llegar desde una FI a un EA, debe vincularse desde la FI al símbolo del LEL, y desde éste vincularse al EA.

El proceso de requisitos utilizado [Leite 04] admite el uso de cualquiera de estos tipos de trazas. Una desventaja que tienen las trazas largas es su mantenimiento,



debido a la gran información que contienen. Por ejemplo, una traza que vincula una FI con un EF va a contener información de los vínculos de la FI, del símbolo del LEL, del EA y del EF en cuestión. Por otro lado, su ventaja está dada por el acceso directo y rápido que presenta de un elemento trazable a otro elemento. Obviamente, estas bondades se invierten al usar trazas incrementales.



**Figura 1. Tipos de Trazas**

En el campo del versionado, existe una gran variedad de estrategias [Sommerville 05], que abarcan desde la forma de generación de versiones hasta el mantenimiento y control de cada cambio. Por un lado, se debe decidir el nivel de especificidad de las versiones: por cada elemento individual del modelo, por grupos de elementos, por el modelo íntegro, e incluso por un conjunto consistente y cohesivo de modelos. Es decir, se debe determinar el tipo de ítem a versionar. La estrategia debe identificar la frecuencia de emisión de una versión: desde una nueva versión por cada cambio efectuado, o el acumulado de sucesivos cambios hasta la decisión de una nueva versión del ítem. La estrategia debe determinar cómo guardar cada cambio realizado sobre cada elemento, y para ello existen varias formas de almacenamiento de cambios:

- versión vieja +  $\Delta$  cambio
- versión vieja y versión nueva
- versión nueva -  $\Delta$  cambio

El obstáculo de la primera opción es la dificultad en la legibilidad del estado actual pero aporta una buena visibilidad de los cambios. La segunda opción tiene buena legibilidad de ambos estados pero sin reconocer los pasos transitados. La tercera opción presenta el estado actual sin problemas y también buena visibilidad de los cambios, yendo hacia atrás pero sin saber hasta donde retroceder. Debe tenerse en cuenta también que llevar una historia de cambios para reconstruir versiones hacia atrás o hacia delante es una opción costosa por la complejidad que ello involucra. La segunda opción puede mejorarse llevando una historia simple de cambios, sin requerir a partir de ella reconstrucciones de versiones. Es decir, que la traza permite reconocer el contexto y los motivos del cambio sin la complejidad del registro de todos los detalles que involucra el mismo. Debe analizarse la relación beneficio-costo de cada opción para cada proyecto en particular.

En el caso de modelos en lenguaje natural, la segunda opción de bajo costo comparativo es recomendable, porque las trazas se pueden manejar en forma simple. Una traza puede informar la causa de un cambio, quién y cuándo se realizó y los efectos producidos, por ejemplo: lista de símbolos del LEL, de EA, de EF y/o de requisitos afectados (ER).

## Resultados y Objetivos

El objetivo propuesto es definir un modelo de trazas que permita seguir la traza de los requisitos a lo largo de todo el proceso de requisitos y que, a su vez, sea extensible a todo el ciclo de vida del software y proveer un modelo de

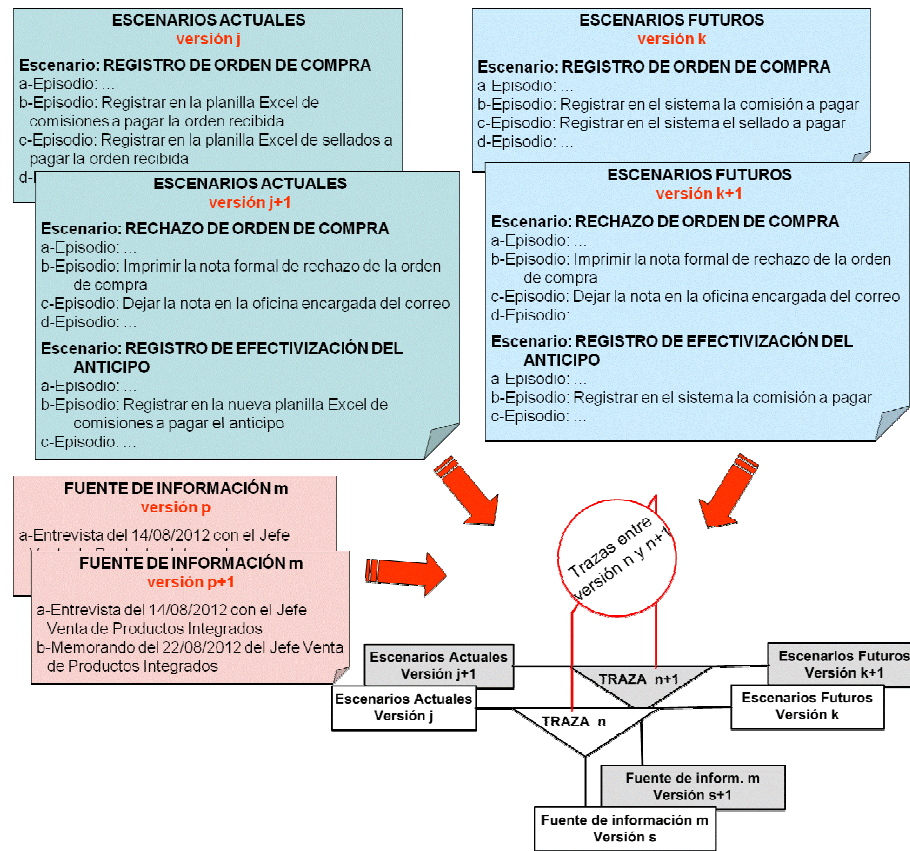


Figura 2. Modelo de Versionado de Trazas

versionado. Como aspecto distintivo de otras estrategias de trazabilidad, la propuesta presentada involucra un modelo de trazas independiente de los modelos de requisitos. Los objetivos específicos son los siguientes:

- Refinar el modelo de trazas.
- Refinar el proceso de captura de trazas.
- Diseñar en detalle el proceso de navegación con trazas.
- Refinar el concepto de trazado de versiones adoptado.

Se ha definido un esquema del modelo de versionado de trazas, que se presenta en la Figura 2. La Traza contiene información acerca de los elementos modelables del proceso de requisitos (ítems) que se relacionan. La Traza entre Versiones

relaciona dos Trazas, donde una Traza se ha generado a partir de la otra por un cambio en uno o más de los ítems que relacionaba; por ende, la Traza entre Versiones contiene la causa o motivo que ha generado la creación de la nueva versión de Traza. Se ha definido en forma preliminar el contenido de las entidades y atributos del modelo de trazas y versionado:

**Ítem**

<id tipo ítem, id ítem, referencia ítem>

**Traza**

<id traza, {id tipo ítem, id ítem}<sub>2</sub><sup>N</sup>, motivo>

**Traza entre versiones**

<id traza N, id traza N+1, motivo>

siendo: id tipo ítem = {símbolo LEL, EA, EF, ER, FI}

Se planifica estudiar en detalle cada uno de los sub-procesos involucrados en el proceso de requisitos desde el punto de vista de las trazas para proveer mecanismos automáticos o semi-automáticos de registro de las mismas.

Algunas de las tareas de investigación ya realizadas consistieron en la definición de guías para producir trazas en cada actividad del sub-proceso de extracción de requisitos desde los EF.

## Formación de Recursos Humanos

El presente proyecto es parte de la tesis doctoral "Modelado del registro de trazas en la Ingeniería de Requisitos" que está desarrollando la Ing. Andrea F. Vera en la UNLP.

## Referencias

- [Antoniol 02] Antoniol G., Canfora G., Casazza G., De Lucia A., Merlo E., "Recovering Traceability Links between Code and Documentation", *IEEE Transactions on Software Engineering*, 28(10):970-983, 2002.
- [Cleland-Huang 03] Cleland-Huang J, Chang CK, Christensen M, "Event-based traceability for managing evolutionary change", *IEEE Trans. Softw. Eng.*, 29(9):796-810, 2003.
- [Cleland-Huang 05] Cleland-Huang J., Settini R., BenKhadra O., Berezhanskaya E., Christina S., "Goal-Centric Traceability for Managing Non-Functional Requirements", *International Conference on Software Engineering*, EEUU, pp. 362-371, 2005.
- [Davis 99] Davis A., Leffingwell D., "Making Requirements Management Work For You", *Crosstalk, The Journal of Defense Software Engineering*, 12(4), 1999.
- [De Lucía 07] De Lucia A., Fasano F., Oliveto R., Tortora G., "Recovering Traceability Links in Software Artifact Management Systems using Information Retrieval Methods", *ACM Transactions on Software Engineering and Methodology*, 16(4):13.1-13.50, 2007.
- [Doorn 10] Doorn JH, Hadad GDS, Vera AF, "Consolidación de Requisitos: Gestión de Requisitos", *Anuario de Investigaciones: Resúmenes Extendidos 2010*, Osvaldo Spósito y Andrés Dmitruk (eds), Universidad Nacional de La Matanza, San Justo, ISBN: 978-987-1635-55-9, pp.19-26, 2012.
- [Huffman Hayes 07] Huffman Hayes J., Dekhtyar A., Karthikeyan Sundaram S., Ashlee Holbrook E., Sravanthi Vadlamudi, April A., "REquirements TRacing On target (RETRO): improving software maintenance through traceability recovery", *Innovations in Systems and Software Engineering*, 3(3):193-202, 2007.
- [Kotonya 98] Kotonya, G., Sommerville, I.: *Requirements Engineering: Processes and Techniques*, John Wiley & Sons, 1998.
- [Leite 97] Leite, J.C.S.P., "Software Evolution, The Requirements Engineering View", keynote address en *anales de XXVI JAIIO - SoST'97 Simposio en Tecnología de Software*, Buenos Aires, Argentina, pp.21-23, 1997.
- [Leite 04] Leite, J.C.S.P., Doorn, J.H., Kaplan, G.N., Hadad, G.D.S., Ridao, M.N., "Defining System Context using Scenarios", en el libro "Perspectives on Software Requirements", Kluwer Academic Publishers, EEUU, ISBN: 1-4020-7625-8, cap. 8, pp.169-199, 2004.
- [Marcus 03] A. Marcus, J. Maletic, "Recovering Documentation-to-Source Code Traceability Links using Latent Semantic Indexing", *Twenty-Fifth International Conference on Software Engineering (ICSE)*, pp. 125-135, 2003.
- [Palmer 97] Palmer J.D., "Traceability", en: *Software Requirements Engineering*, R.H. Thayer, M. Dorfman (eds.), *IEEE Computer Society Press*, 2º ed., pp. 364-374, 1997.
- [Sommerville 05] Sommerville I, "Ingeniería del Software", *Pearson Educación*, 7º ed., 2005.

## Evaluación de metodologías y técnicas para la generación de pruebas aplicadas un caso real a partir de sus especificaciones funcionales

**Silvia Sanchez Zuaín, Lilia Palomo y Beatriz Fernández Reuter**

Facultad de Matemática Aplicada - Universidad Católica de Santiago del Estero

*silvia.sanchez@ucse.edu.ar - lilia.palom@ucse.edu.ar - bfreuter@gmail.com*

### RESUMEN

Las pruebas son una parte importante del proceso de verificar si el software satisface las expectativas del usuario y deben presentarse a lo largo de todo el ciclo de vida del desarrollo, como un aspecto decisivo en el control de calidad del producto final. Como parte de las mismas están las pruebas funcionales, para las cuales se requiere una buena planificación que consiste en definir los aspectos a chequear y la forma de comprobar su correcto funcionamiento, punto en el cual adquieren sentido los casos de prueba. También es necesario contar con una metodología y herramientas adecuadas que permitan validar los requisitos funcionales, corregir los errores y mejorar su calidad.

Este trabajo tiene como objetivo presentar una visión global de distintas metodologías para la generación de pruebas a partir de la definición de requisitos funcionales, aplicadas a un caso real, a fin de determinar su efectividad en la detección de errores de sistemas de forma temprana.

**Palabras clave:** Pruebas del sistema, metodologías de pruebas tempranas, generación de pruebas, casos de prueba, requisitos funcionales, casos de uso.

### CONTEXTO

La Facultad de Matemática Aplicada, impulsa el desarrollo de un conjunto de

proyectos que apuntan a incentivar la investigación desde las cátedras promoviendo la interacción vertical y horizontal. Esto se centra en posibilitar la participación de los docentes en investigación y obtener resultados transferibles a las actividades de las cátedras.

Es por esto, que desde las cátedras de Programación I, Sistemas de Información II e Ingeniería del Software, en el año 2011 surge el proyecto “Generación de pruebas de sistemas a partir de las especificaciones funcionales aplicadas a un caso real” con el objetivo de promover la investigación aplicada, formación de recursos humanos e innovación de los contenidos de las cátedras y en las prácticas profesionales.

El presente artículo es una línea de investigación que surge del proyecto mencionado anteriormente, ante la necesidad de determinar cual o cuales metodologías utilizar para la generación de pruebas en un caso real, compararlas y así determinar su efectividad en la detección de errores de sistemas en etapas tempranas.

### INTRODUCCIÓN

El desarrollo de software implica una serie de actividades de producción donde las posibilidades de que aparezcan fallos humanos son grandes. Los errores pueden aparecer desde el primer momento del



proceso, a causa de objetivos especificados de forma errónea, o bien en pasos posteriores del diseño y desarrollo. Por lo tanto, el desarrollo de software debe ir acompañado de una actividad que garantice su calidad. Un aspecto crucial en el control de calidad del desarrollo de software son las pruebas, cuyo objetivo, consiste en asegurar que el sistema hace lo que el cliente quiere que haga, es decir, verificar que el software cumple con sus requisitos.

Es importante comprender que los errores pueden insertarse en un requerimiento, en el diseño, en un componente del código, en la documentación, o en cualquier momento durante el desarrollo o el mantenimiento.

Como corregir los errores en un sistema tiene un costo elevado a medida que el proceso de desarrollo de software avanza en sus etapas, es fundamental detectarlos en la fase más temprana posible. Una forma de hacerlo es generando casos de prueba en paralelo con el desarrollo del software y donde estos casos de prueba definidos, simulen interacciones del actor con el sistema para verificar que el sistema hace lo que se espera de él [2].

Otra manera de detectar de forma temprana los errores del sistema es a partir de sus especificaciones funcionales. Algunas investigaciones proponen diversas metodologías que permiten la generación de casos de prueba a partir de la definición de casos de uso, tal y como lo plantea Correa y Giandini [1], que genera casos de prueba, a partir de cada caso de uso; y casos de pruebas de datos, a partir de los datos necesarios para la ejecución de un caso de uso. Además, se vale de diagramas de actividades de UML, derivados del modelo de casos de uso, para especificar el comportamiento de los casos de prueba; y de una transformación modelo a

texto para obtener como resultado final de este proceso, las pruebas. Otro caso es el de Heumann et al [5] que desarrolla un método que parte de los casos de uso para generar los casos de prueba, identificando en cada uno los posibles escenarios, o caminos de ejecución, y definiendo los valores a probar con cada caso de prueba. Por último, obtiene una lista de casos de prueba, con los valores que se deben probar y los resultados esperados para cada caso.

En los últimos años se ve una tendencia hacia la automatización del proceso de generación de casos de prueba a través de diferentes métodos y herramientas que parten de casos de uso escritos en lenguaje no formal y aplican diferentes métodos para obtener objetivos de prueba [3], o bien, de casos de usos bien definidos que se toman de entrada para ejecutar algoritmos que generan diagramas de actividad de los que se derivan los casos de prueba [4] [7] [8] [9] [10].

Estas metodologías que se mencionan precedentemente indican lo que una prueba debe hacer, al ejecutar un escenario posible de un caso de uso, qué información hay que suministrarle y qué información va a devolver. Por lo tanto, lo primero que es necesario saber para poder implementar las pruebas es tener en cuenta los puntos comunes de las propuestas de generación de pruebas del sistema [6], de los que se se consideran entre otros: a- El objetivo de estas propuestas es obtener un conjunto completo de pruebas del sistema que permitan garantizar que el sistema software cumple con la especificación funcional dada, lo cual permite asegurar su calidad. b- Todas parten de los requisitos funcionales del sistema y la mayoría permiten comenzar a desarrollar los casos de prueba del sistema en cuanto se

comience a disponer de los requisitos funcionales.

Del análisis de lo expuesto surge el planteo de la conveniencia sobre la aplicación de algunas de estas metodologías para determinar su alcance, es decir, que tan efectivas son para detectar errores, que tan fácil de realizar son, si es conveniente utilizar una técnica manual o automatizada, que nivel de cobertura tienen, entre otros aspectos.

### **LINEAS DE INVESTIGACIÓN Y DESARROLLO**

Las líneas de investigación del presente trabajo tienen como eje central las metodologías y herramientas para las pruebas tempranas del software, y en particular la generación de casos de pruebas. Estas líneas de investigación en pruebas del software se resumen a continuación:

- Investigación de las actuales formas usadas para derivar casos de prueba funcional.
- Análisis y comparación de metodologías y herramientas existentes que soportan los procesos de pruebas.
- Estudio del proceso de aplicación de las metodologías de generación de casos de prueba para un caso real.
- Evaluación de la cobertura y efectividad de las pruebas y diseño de pruebas.

El equipo de investigadoras, se propone desarrollar proyectos compatibles con esta línea de investigación y diseñar una estrategia para su desarrollo. Con ello, se quiere dar impulso, por ejemplo, a pequeños proyectos o actividades que puedan ser desarrollados como trabajos de grado y/o monografías de especialización, que servirán para confirmar o rechazar hipótesis, y con el fin de alimentar futuros proyectos.

### **OBJETIVOS Y RESULTADOS ESPERADOS**

El objetivo general de este trabajo es analizar y comparar las diferentes metodologías para la generación de pruebas tempranas a partir de la especificación de requisitos funcionales, para determinar cuál o cuáles de ellas aplicar.

Para llevar a cabo este objetivo primero se deberá seleccionar un caso real al cuál se aplicarán dichas metodologías. Luego, determinar las metodologías que se utilizarán y los criterios con los cuales se realizará la comparación de las mismas. Por último, generar casos de prueba para comprobar si el sistema definido en el caso real, implementa el comportamiento especificado en sus requisitos funcionales.

De esta investigación se espera que sus resultados se incorporen en los contenidos de las cátedras relacionadas y al espacio curricular correspondiente.

Como resultados indirectos se espera la consolidación del grupo de investigación, la formación de nuevos investigadores y la motivación y entrenamiento en investigación de los estudiantes de grado.

### **FORMACIÓN DE RECURSOS HUMANOS**

El equipo de trabajo está integrado por 3 docentes de la carrera de Ingeniería en Informática, 2 adjuntos y 1 Jefe de trabajos prácticos, todos con dedicación semi-exclusiva.

El grupo hace difusión y formación de recursos humanos desde las cátedras: Programación I, Sistemas de Información II e Ingeniería de Software.

Se considera de gran interés la incorporación de becarios para motivar a los alumnos de la

carrera de Ingeniería en Informática a realizar su trabajo final de grado en el área de este proyecto.

Además, se prevé que una de las docentes que integran el equipo, desarrolle su tesis de Maestría en el marco de este trabajo.

### REFERENCIAS

- [1] Natalia Correa, Roxana Giandini (2011). Generación Automática de Casos de Prueba a partir de Casos de Uso: Una Propuesta Basada en MDD/MDT. 40JAIIO - ASSE 2011 - ISSN: 1850-2792 - Página 168
- [2] Javier J. Gutiérrez, María J. Escalona, Manuel Mejías y Jesús Torres (2006). Hacia una propuesta de Pruebas Tempranas del Sistema. XV Jornadas de Ingeniería del Software y Bases de Datos JISBD 2006 José Riquelme - Pere Botella (Eds) .CIMNE, Barcelona, 2006.
- [3] Javier J. Gutiérrez, María J. Escalona, Manuel Mejías, Jesús Torres, Arturo Torres-Zenteno. Generación automática de objetivos de prueba a partir de casos de uso mediante partición de categorías y variables operacionales.
- [4] J. J. Gutiérrez, M. J. Escalona, M. Mejías, J. Torres. Derivation of test objectives automatically.
- [5] J. Heumann, "Generating Test Cases From Use Cases," The rational edge, <http://download.boulder.ibm.com/ibmdl/pub/software/dw/rationaledge/jun01/GeneratingTestCasesFromUseCasesJune01.pdf>, 2001.
- [6] J. Gutiérrez, M. J. Escalona, M. Mejías, "Analysis of Proposals to Generate System Test Cases From System Requirements," in CAiSE'05 Forum, Porto, Portugal, 2005.
- [7] Natalia Correa, Roxana Giandini (2012). Casos de Prueba del Sistema Generados en el Contexto MDD/MDT. 41 JAIIO - ASSE 2012 - ISSN: 1850-2792 - Pág. 91-105
- [8] Marc-Florian Wendland, Ina Schieferdecker and Alain Vouffo-Feudjio. 2011. Requirements-driven testing with behavior trees. 2011 Fourth International Conference on Software Testing, Verification and Validation Workshops.
- [9] Yasmine Ibrahim Salem y Riham Hassan. 2011. Requirement-Based Test Case Generation and Prioritization. 978-1-61284-185-4/111 - 2011 IEEE
- [10] Bill Hasling, Helmut Goetz, Klaus Beetz. 2008. Model Based Testing of System Requirements using UML Use Case Models. 2008 International Conference on Software Testing, Verification, and Validation

## Utilizando contratos JML para optimizar diseños orientado a objetos siguiendo MDA

Marcelo Uva, Mariana Frutos, Ariel Gonzalez, Ariel Arsaute, Marcela Daniele, Paola Martellotto, Fabio Zorzan.

Departamento de Computación, Facultad de Ciencias Exactas, Físico-Químicas y Naturales,  
Universidad Nacional de Río Cuarto.

Ruta 36 Km. 601 -CP 5800 - Río Cuarto - Córdoba - Argentina Tel. (0358) 4676235  
{uva, mfrutos, agonzales, aarsaute, marcela,Paola, fzorzan}@dc.exa.unrc.edu.ar

### Resumen

Model Driven Architecture (MDA) define un proceso de construcción de software basado en la producción y transformación de modelos. En Ingeniería de Software, refactorización es la técnica que reestructura código de una aplicación, alterando su estructura interna sin modificar su comportamiento externo. Por otro lado, Java Modeling Language (JML) es un lenguaje para especificar programas Java, utiliza precondiciones, postcondiciones e invariantes de la lógica de Hoare. Este trabajo plantea una técnica basada en MDA que posibilita la construcción de una herramienta automática que tomará como entrada código de una aplicación Java y realizará optimizaciones en su diseño basándose en reglas de refactorio. La técnica requiere contar con los contratos JML de los métodos de las clases involucradas. El principal aporte de este trabajo es la utilización de contratos JML para asegurar que el comportamiento de un módulo se mantiene sin cambios, luego de la aplicación de reglas de refactorio.

**Palabras clave:** MDA, Refactoring, JML.

### Contexto

La línea de investigación presentada en este trabajo se desarrolla en el marco del proyecto “Ingeniería de Software: Automatización de Procesos de Desarrollo de Software”, perteneciente a los Proyectos y Programas de Investigación (PPI) de la secretaría de Ciencia y Técnica de la Universidad Nacional de Río Cuarto.

### Introducción

Model Driven Architecture (MDA) surge a partir de las tecnologías estandarizadas por la Object Management Group (OMG)[1]. MDA define un proceso de construcción de software basado en la producción y transformación de modelos. Los principios sobre los cuales se fundamenta MDA son: abstracción, automatización y estandarización. El proceso central de MDA es la transformación de modelos. La idea subyacente es producir nuevos modelos a partir de uno dado, de manera tal de preservar propiedades en un modelo abstracto independiente de los cambios producidos en la tecnología. MDA establece cuatro niveles de abstracción: CIM (Computation Independent Model), PIM (Platform Independent Model), PSM (Platform Specific Model), y la aplicación final.

En Ingeniería de Software se define el término *refactorización* (“Refactoring”)[2] a la técnica que posibilita reestructurar el código fuente de una aplicación de software, alterando su estructura interna sin modificar su comportamiento externo. De acuerdo a la jerarquía de abstracción establecida por MDA, la *refactorización* define transformaciones sobre el código fuente de una aplicación, es decir, en el nivel menos abstracto de los cuatro.

La refactorización se realiza a menudo como parte del proceso de desarrollo de software: los desarrolladores alternan la inserción de nuevas funcionalidades y casos de prueba con la refactorización del código para mejorar su consistencia interna y claridad. El proceso de



refactorización consiste en la aplicación de una serie de reglas de transformación que preservan comportamiento. Cada transformación puede producir una reestructuración significativa. Posterior a la aplicación de cada regla, el diseñador/programador provee una batería de casos de prueba para “asegurar” que el comportamiento no es alterado. Por otro lado, Java Modeling Language (JML) [3] es un lenguaje de especificación para programas Java, que utiliza precondiciones, postcondiciones e invariantes de la lógica de Hoare, siguiendo el paradigma de diseño por contrato [4]. JML provee una semántica formal para describir el comportamiento de un módulo de Java, dejando de lado todas las posibles ambigüedades con respecto a las intenciones iniciales del diseñador.

### **Refactorización de modelos**

La refactorización (“Refactoring”) es una técnica de ingeniería de software que permite reestructurar código fuente, alterando su estructura interna sin modificar su comportamiento externo. En MDA, la refactorización establece las reglas que permiten el mapeo de un modelo en otro dentro de un mismo nivel de abstracción.

Martin Fowler [2] define refactorización al área que se especializa en el análisis y el diseño orientado a objetos, entre otros. La refactorización define cambios estructurales internos en el software para mejorar su comprensión con un costo inferior al de modificar el comportamiento observable del sistema. Este proceso es logrado mediante la aplicación de un conjunto de reglas de refactorización.

El último punto de las ventajas citadas por Fowler, establece implícitamente al proceso de refactorización como un procedimiento de optimización guiado por heurísticas. Es por ello que para este trabajo se ha acotado el amplio conjunto de reglas definidas en [2] a un subconjunto posible de automatizar. Es decir, la corrección de la aplicación de las reglas ya no dependerá de las habilidades de un desarrollador particular.

La refactorización forma parte del proceso de fabricación de un software: los desarrolladores alternan la inserción de nuevas funcionalidades y

casos de prueba con la refactorización del código para mejorar su consistencia interna y su claridad. Los casos de prueba deben ser superados por el módulo, antes y después de la aplicación de cada regla de refactorización. Esto provee una “cierta” seguridad acerca de la no alteración del comportamiento del código, aunque “el testing de los programas puede ser muy efectivo para mostrar la presencia de errores, sin embargo resulta inadecuado para mostrar su ausencia” Edsger Dijkstra.

En este trabajo, se propone una alternativa basada en contratos JML para asegurar que los componentes involucrados no modifiquen su comportamiento.

### **Líneas de investigación y desarrollo**

En la literatura existente sobre el tema se pueden encontrar diferentes mecanismos orientados a evolucionar o mejorar diseños aplicando técnicas de refactorización. Opdyke [5] presentó un conjunto de transformaciones básicas para poder agregar, eliminar y mover elementos entre clases, con la finalidad de optimizar diseños de aplicaciones desarrolladas en C++. Para ello definió una serie de condiciones iniciales para asegurar uno de los principios fundamentales de toda técnica de refactorización, el mantenimiento del comportamiento del programa.

En este trabajo se plantea una técnica fundada en los principios de la filosofía MDA a partir de la cual se sientan las bases para la construcción de una herramienta automática, que tomará como entrada el código fuente de una aplicación orientada a objetos implementada en Java y realizará una serie de optimizaciones en su diseño aplicando reglas de refactorización. Las reglas de transformación establecidas en este trabajo son definidas a nivel PSM dentro de la jerarquía MDA, ya que se ha trabajado principalmente con diagramas de clases en UML [6] implementados en el lenguaje Java. La técnica requiere los contratos JML de cada uno de los métodos involucrados. Esto permite contar con la semántica formal de cada componente, eliminando todo tipo de ambigüedad con respecto a las intenciones iniciales del diseñador.

En la figura 1 se presenta la arquitectura genérica de la técnica descrita en este trabajo y sobre la cual los autores están desarrollando un prototipo.

La arquitectura genérica esta compuesta por un conjunto de módulos independientes. Cada módulo posee una interfaz bien definida. El mecanismo toma como entrada un programa Java anotado con sus contratos JML. Para la generación del modelo XMI equivalente, se utiliza la herramienta ArgoUML[7], que permiten realizar ingeniería inversa y luego exportar el diagrama de clases generado a una especificación XMI o XML Metadata Interchange (XML de Intercambio de Metadatos) [8].

Posteriormente, el módulo “Motor de la transformación” en conjunto con el módulo “Reglas de Transformación” interactúan buscando en el modelo XMI algún patrón para la aplicación de una de las reglas de refactorización. Luego, de la aplicación de cada regla, los contratos JML deberán seguir verificándose. Una vez finalizado el proceso de transformación se producirá un nuevo modelo XMI el cual será exportado finalmente a código fuente Java mediante la aplicación de técnicas de ingeniería directa. Este proceso permite asegurar que la semántica del programa original se mantiene sin cambios, ya que todas las modificaciones realizadas no afectarán el cumplimiento de los contratos preexistentes.

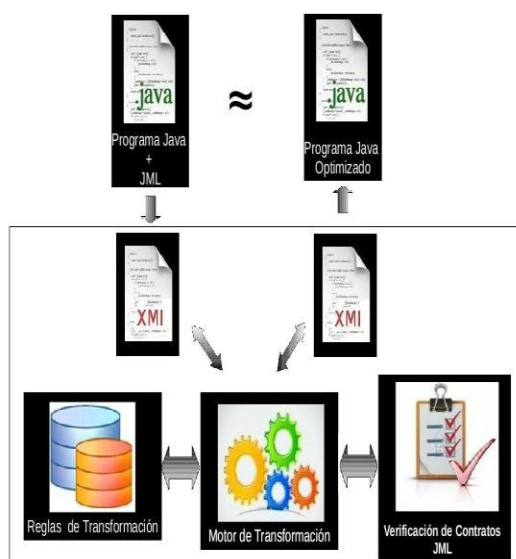


Fig. 1. Arquitectura genérica descrita en la técnica

Este trabajo está basado un conjunto amplio de reglas de refactorización, algunas de ellas, vinculadas a patrones de diseño. Recordemos que un patrón de diseño [9] proporciona una solución efectiva y reusable a un problema recurrente de diseño. Por ejemplo, las reglas de refactorización “Extract Composite” y “Replace Constructor with Factory Method”, están relacionados a los patrones “Composite” y “Factory Method”. Las mejoras producidas en el diseño luego de la aplicación de estas reglas, están ligadas íntimamente a las ventajas de la aplicación de cada uno de estos patrones.

El módulo “Reglas de transformación” ha sido concebido como un componente que pueda crecer paulatinamente a medida que nuevas reglas se vayan incorporando. Cada una de éstas deberá establecer el patrón de búsqueda a detectar y las condiciones iniciales y finales para su aplicación. Se debe tener en cuenta, que no todas las reglas de refactorización son posibles de automatizar. Un gran número de reglas de refactorización plantean heurísticas para la mejora de un diseño, cuya aplicación correcta sólo dependerá de las habilidades del diseñador/programador. Es por ello que para la incorporación de nuevas reglas, es necesario un estudio profundo acerca de las condiciones previas a la aplicación de la misma y la factibilidad o no para su automatización.

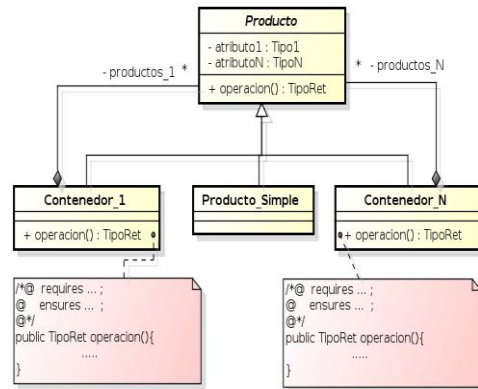
A modo de ejemplo se presenta la regla “*Extract Composite*” [10].

### Regla Extract Composite

La regla *Extract Composite* establece una refactorización combinando la regla anterior y el patrón de diseño Composite.

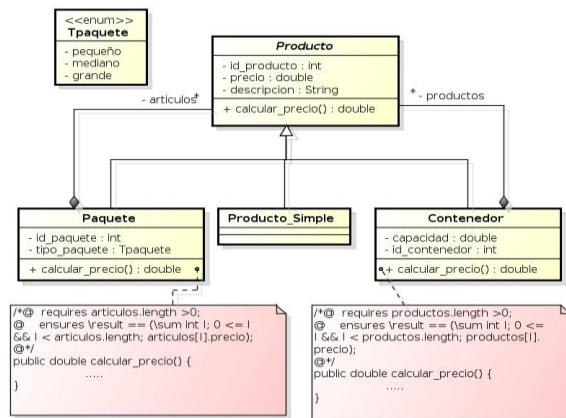
En aquellos modelos en donde los objetos instanciados de una subclase están compuestos o relacionados con objetos de la misma clase, existe alta probabilidad de que se encuentre duplicación de código. Esta regla plantea un escenario particular en donde se produce la duplicación de métodos (o bien duplicación del “comportamiento” de métodos).

No es necesario que los nombres de los métodos sean iguales, sí los perfiles deberían coincidir (o al menos uno de ellos debería estar incluido en el otro).



**Fig. 2. Diseño genérico previo a la aplicación de la regla Extract Composite**

En figura nro. 2 presentamos la estructura genérica de la regla. La aplicación de la regla deberá inicialmente buscar aquellos métodos u operaciones en las subclases “hermanas” que posean la misma signatura, o bien que una incluya a la del otro método. El caso más común es encontrar métodos con el mismo nombre y signatura.



**Fig. 3. Diseño genérico posterior a la aplicación de la regla Extract Composite**

Luego, se aplica un algoritmo de sustitución para equiparar nombres de variables en ambos contratos JML. En la figura 2, los roles *productos\_1* y *productos\_N*, se reemplaza en la figura 3 por el rol *productos*. Una vez culminado el proceso de sustitución se verifican si los contratos de las operaciones iniciales se implican mutuamente.

Bastará con que uno de los dos contratos implique al otro, como para reemplazar el contrato genérico por el más débil. Finalmente, se creará una superclase intermedia (en la Figura 3 “*Producto\_Compuesto*”) que contendrá el método en común con los contratos (previa sustitución) del método con el contrato más débil. En el caso de que los contratos se impliquen mutuamente, se podrá optar por cualquiera de las dos (previas sustituciones).

### Resultados y Objetivos

El cambio continuo de los requerimientos, la incorporación de nuevas funcionalidades, la minimización de los tiempos de respuesta junto con la mejora y aprovechamiento de los recursos son características comunes presentes en todo proyecto de desarrollo de software.

Los sistemas actuales requieren de la ejecución de actividades orientadas a elevar los niveles de calidad del software. Dentro de éstas se encuentran las tareas de refactorización que permiten la reorganización y reestructuración de entidades/funcionalidades adaptándolas a los nuevos requerimientos y diseños arquitecturales. Todas estas tareas deben asegurar el mantenimiento del comportamiento original. En este trabajo se ha presentado una técnica para optimizar diseños de aplicaciones Java de manera automática utilizando reglas de refactoro. La técnica está basada en la filosofía MDA. Se establecen transformaciones a nivel PIM (entre diagramas de clases) y a nivel PSM (entre aplicaciones java concretas). Uno de los aportes de este trabajo es el uso de contratos JML para asegurar que el comportamiento de los módulos se mantiene igual luego del proceso automático de refactorización.

La arquitectura genérica propuesta está conformada por módulos independientes con interfaces bien definidas. El módulo “Reglas de Transformación” almacena las reglas de transformación. Cada una de ellas establece un patrón específico para su correcta aplicación. La incorporación de nuevas reglas se realiza de manera incremental logrando automatizar poco a poco el proceso de optimización. A medida que el módulo

aumente su tamaño (cantidad de reglas) se obtendrán mejores modelos y en consecuencia, mejores aplicaciones.

Actualmente los autores de este trabajo están desarrollando un prototipo de la herramienta que permite la automatización parcial de la propuesta presentada en este trabajo. Para ello se han utilizado las herramientas ArgoUML[7] y ESC/Java2[11].

### Formación de Recursos Humanos

Los temas abordados en esta línea de investigación brindan un fuerte aporte al proceso de perfeccionamiento continuo de los autores de este trabajo, que se desempeñan como docentes de las carreras de computación que se dictan en la Universidad Nacional de Río Cuarto y participan en asignaturas relacionadas a dichos temas.

### Referencias

- [1] Miller, J., Mukerji, J., MDA Guide Version 1.0.1 Document number omg/2003-06-01, <http://www.omg.com/mda>, 2003.
- [2] Fowler, M. Refactoring Improving The Design of Existing Code. Addison Wesley Longman, Inc. 1999.
- [3] Gary T. Leavens, Albert L. Baker, and Clyde Ruby. [JML: A Notation for Detailed Design](#). In Haim Kilov, Bernhard Rumpe, and Ian Simmonds (editors), [Behavioral Specifications of Businesses and Systems](#), chapter 12, pages 175-188. Copyright [Kluwer](#), 1999
- [4] Meyer, Bertrand: *Applying "Design by Contract"*, in Computer (IEEE), pp. 40–51. 1992.
- [5] Opdyke, W. Refactoring Object-Oriented Frameworks. Tesis doctoral, Universidad de Illinois, Urbana-Champaign.1992.
- [6] Grady Booch, Ivar Jacobson & Jim Rumbaugh (2000) [OMG Unified Modeling Language Specification](#). First Edition: March 2000.
- [7] ArgoUML. Open Source Software Engineering Tools. [argouml.tigris.org](http://argouml.tigris.org) Tigris.org.
- [8] XML Metadata Interchange Specification OMG Formally Released Versions of XML. Version 2.0.1 formal/05. <http://www.omg.org/spec/XMI/ISO/19503/PDF/>
- [9] Erich Gamma, Richard Helm, Ralph Johnson y John Vlissides. *Design Patterns: Elements of Reusable Object-Oriented Software*. Addison-Wesley, 1995.
- [10] Joshua Kerievsky. Refactoring to Patterns. Addison-Wesley Professional. Part of the Addison-Wesley Signature Series (Fowler) series. 2005.
- [11] Patrice Chalin, Joseph R. Kiniry, Gary T. Leavens, and Erik Poll *Beyond Assertions: Advanced Specification and Verification with JML and ESC/Java2* . <http://www.eecs.ucf.edu/~leavens/JML/fmco.pdf>



## Implementación de una red social académica universitaria

Ana M. Piccin, Sergio O. Aguilera, Alejandro Oliveros

Facultad de Ingeniería y Tecnología Informática

Universidad de Belgrano

F. Lacroze 1947 – Piso 1

1426 Buenos Aires

Teléfonos: 4772-4010 int. 150 / 4511-4716

[ana.piccin@comunidad.ub.edu.ar](mailto:ana.piccin@comunidad.ub.edu.ar); {sergio.aguilera, alejandro.oliveros}@ub.edu.ar

### Resumen

Se presenta el proyecto de implementación de una *red social académica* (RSA), con el propósito de alentar prácticas pedagógicas participativas y favorecer el protagonismo de los alumnos en las actividades académicas y en su propio aprendizaje. Se enuncia un concepto de *redes sociales académicas* como caso particular de *redes sociales* (RS). A partir de una breve caracterización del sistema de información utilizado recientemente en la UB, se plantea la implementación de la RSA *ConexiónUB* en términos de transformación cultural, lo que será determinante para el diseño del plan. Se detallan las características críticas que se deberán tener en cuenta para el diseño. En base a conclusiones extraídas de experiencias previas, se destaca la relevancia de la categorización de los usuarios para determinar las etapas de implementación y el plan de capacitación. Se plantean y justifican los ejes que guían el diseño del plan de trabajo. Las conclusiones que se obtengan aportarán al conocimiento de buenas prácticas en la implementación de RSA.

**Palabras clave:** Cambio cultural, Construcción colaborativa del conocimiento, Educación superior, Redes sociales (RS), Redes sociales académicas (RSA).

### Contexto

Este proyecto surge como respuesta al requerimiento de la Universidad de Belgrano (UB) de implementar el sitio *ConexiónUB* (<http://conexion.ub.edu.ar>) como espacio para la comunicación entre docentes y alumnos en toda la institución. *ConexiónUB* es una *red social académica* (RSA) que actualmente funciona en escuelas de enseñanza media.

La UB ha decidido implementar una RSA para dar nuevas formas a la relación entre docentes y alumnos, para promover espacios ágiles de comunicación e intercambio y para favorecer que los alumnos encuentren protagonismo en el aprendizaje. Al momento de este requerimiento, se encuentra en desarrollo un trabajo de investigación destinado al estudio de las prácticas de diseño, desarrollo e implementación de *redes sociales académicas* (RSA). [Oli 2012] y está en

sus comienzos una tesina dedicada a la movilidad de RSA. Mediante la aplicación de *ConexiónUB*, se ha cumplido la primera etapa de ejecución de un proyecto de capacitación docente, en la aplicación de RSA en la enseñanza, en diez escuelas medias de Gran La Plata y zona sur del Gran Buenos Aires.

## Introducción

“Cuando las redes de computadoras vinculan tanto a las personas como a las máquinas, se transforman en redes sociales”. [Wel 1996]

**Redes sociales.** Las innovaciones tecnológicas aparecen como soluciones a problemas planteados por la sociedad y, en la medida en que son adoptadas, estas innovaciones modifican paulatinamente a la propia sociedad que las ha construido. De modo semejante, la sociedad se apropió de las TIC para fortalecer sus lazos y éstas, a su vez, modificaron las formas en que esos lazos (preexistentes) se construyen y se mantienen.

Entendemos *como red social (RS)* a un sitio Web equipado con herramientas de comunicación, que permite la administración de lazos interpersonales y el intercambio de contenidos entre sus integrantes. El elenco de herramientas incluidas en el sitio está orientado a favorecer la conexión e interacción entre todos sus participantes. A lo largo del ciclo de vida de la RS, las relaciones entre sus integrantes se expanden. Potencialmente, todos podrían estar comunicados entre sí, ya sea por tener conocidos en común, como por tener intereses similares. Para el desarrollo de estas relaciones, las RS ofrecen a los usuarios funcionalidades específicas, como la generación de grupos y

herramientas de comunicación sincrónicas y asincrónicas que facilitan la organización de las comunicaciones según las preferencias de los actores.

El valor de una RS excede la colección de herramientas que ofrece a los usuarios. Su identidad está asociada a características de sus participantes, a las reglas que se establecen entre ellos, a los principios que se reflejan en estas reglas, al contexto en el que se desarrolla y el tipo de contenidos que circulan.

**Redes sociales académicas.** Una *red social académica (RSA)* es un caso particular de *red social (RS)*.

Las diferencias específicas de una RSA con respecto a una RS se localizarán en particularidades derivadas de las características de sus *stakeholders* (docentes, alumnos, directivos, personal de apoyo, etc.), con el tipo de actividad (comunicación entre docentes y alumnos con fines educativos) y con el contexto en el que se desarrollan (normativas institucionales, cultura institucional) y, notablemente, en los contenidos que circulan.

La funcionalidad y organización que ofrecen las RSA, promueven y soportan intercambios entre los usuarios que culminan en la construcción colaborativa de conocimiento. Para gestionar el conocimiento, las RSA ofrecen herramientas sobre las que se pueden diseñar estrategias de planeamiento, gestión y control. En el sistema de gestión del conocimiento intervienen: las políticas institucionales, generalmente fijadas por la institución educativa; las estrategias, implementadas sobre las herramientas provistas en la red; y los actores de la RSA, que funcionan como agentes de control y custodios de la integridad e identidad de la red. En algunos casos, estos agentes de control

son específicos. En otros, es la comunidad en su totalidad la que decide el valor de los contenidos, su inclusión o exclusión de la RSA y la que mantiene la identidad expresiva discursiva de la red.

Existen distintos tipos de RSA, que están determinados por las particularidades que asumen las características específicas de la RSA y su articulación en un dado contexto institucional.

**Un cambio cultural.** La adopción de nuevas tecnologías implica la transición hacia nuevas formas culturales [Doo 1999]. En nuestro caso, se trata de llevar a los usuarios desde un escenario estático a un escenario dinámico.

Denominamos *escenario estático* al sistema con el que se encuentran familiarizados los usuarios. El escenario estático se materializa en un repositorio en el que los docentes acumulan los documentos básicos de las asignaturas (por ejemplo: programas, cronogramas, guías de trabajos prácticos) y otros materiales que desean compartir con sus alumnos. Los alumnos acceden a este repositorio sólo para seleccionar material y bajarlo a sus computadoras. No existen otras funcionalidades significativas. Por esta razón, muchos docentes decidieron comunicarse además con sus alumnos mediante el uso de herramientas comunes en la Web, como el *correo electrónico*, el *chat*, los grupos de interés y el acceso compartido a contenidos.

Denominamos *escenario dinámico* a la RSA que se proyecta implementar. Como se describió anteriormente, una RSA ofrece una amplia variedad de formas de comunicación e intercambio, de posibilidades de compartir conocimientos e incluso habilita y favorece su producción colaborativa. En este tipo de ambientes, cada docente

deberá decidir con qué estrategia organizará tu tarea educativa, de qué manera la aplicará en sus clases, qué tipo de espacios de trabajo ofrecerá a sus alumnos.

La utilización de las RSA en la educación modificará las formas de la vinculación y de los intercambios entre docentes y alumnos de la institución. La potencia y flexibilidad de las RSA promoverán prácticas docentes participativas y favorecerán el protagonismo de los alumnos en las actividades académicas y en su propio aprendizaje. [Ert2009] [Cas2009] Podemos prever que esto impactará sobre las estrategias docentes, sobre la calidad de los contenidos, así como sobre las dinámicas de construcción del conocimiento.

Por esta razón, se decidió seguir una estrategia de dos etapas básicas. En primer lugar considerar, como punto de partida para el plan de implementación, la adaptación de las funcionalidades de la RSA *ConexiónUB* a la cultura de comunicación preexistente en la institución, centrada en el uso del repositorio con el agregado de herramientas Web 2.0 seleccionadas de la Internet. Una vez logrado este objetivo, será más sencillo, en una segunda etapa, liberar las demás propiedades de la *red social académica* (RSA).

**Los usuarios.** Para la implantación de una RSA, es crítico asegurar la participación de los usuarios y su adhesión a la dinámica de trabajo impuesta por la propia RSA. La experiencia ganada en capacitación docente y el análisis de la experiencia previa en la UB, nos han señalado la importancia de la variedad de perfiles de los docentes que utilizarán el sistema. En

nuestro caso, podemos prever que la categorización de los usuarios, estará fuertemente asociada con sus competencias informáticas. Como más significativas podemos citar: la idoneidad en el uso de herramientas informáticas básicas, la idoneidad en el uso de la Internet como fuente de información, la experiencia en el uso de herramientas Web 2.0 para la comunicación y participaciones anteriores en RS. Estas características estarían relacionadas, aunque no siempre, con el grupo etario del usuario y con el tipo de disciplina en la que se desempeñe el docente. Por estas consideraciones, se ha decidido comenzar con una la implementación piloto en la Facultad de Ingeniería y Tecnología Informática (FITI), donde se supone que hay mayor afinidad tecnológica en usuarios docentes, para luego extenderla al resto de la universidad.

En cuanto a los usuarios alumnos, podemos considerar que son nativos digitales, con lo que el cambio cultural no sería significativo en el sentido de adopción de la RSA como ámbito de comunicación.

No prevemos por ahora situaciones significativas vinculadas con la incorporación de otros actores.

**Aportes de este proyecto al conocimiento sobre las RSA.** Este proyecto de implementación nos enfrentará con situaciones específicas y generalizables de las RSA. Las conclusiones que se obtengan serán útiles para definir criterios y diseñar estrategias de implementación en otras instituciones de educación superior. El trabajo realizado hasta ahora nos permite anticipar nuevas líneas de investigación. Podemos enunciar, por ejemplo, los

estudios relacionados con la identidad de la RSA como: hasta dónde podrá extenderse la RSA sin dejar de ser la misma; quiénes podrán, o no deberán ser sus integrantes, para no desvirtuarla o cuál el valor de la RSA como modelo informático de las relaciones interpersonales de la institución.

**Plan de trabajo.** La experiencia en formación docente nos permite suponer, además, que la exposición programada y gradual a herramientas seleccionadas, facilitará que docentes y alumnos se apropien de ellas y las apliquen en las distintas actividades académicas [Lev2007]. El plan de trabajo tiene en cuenta estas premisas y propone una implementación incremental.

Las etapas que conforman el plan son las siguientes.

1. Revisión bibliográfica.
2. Caracterización del contexto informático existente.
3. Categorización de perfiles del usuario docente.
4. Adecuación del sistema básico
5. Diseño de instrumentos para la recolección de datos
6. Formación de los usuarios.
7. Implementación de la RSA.
8. Análisis de los datos recogidos.
9. Conclusiones y generalización.

Actualmente el proyecto se encuentra en el proceso de adecuación del sistema. Ya se ha realizado la caracterización del sistema preexistente. Se está diseñando el plan de capacitación.



## Líneas de investigación y desarrollo.

1. Las RSA como modelos institucionales. Alcances y limitaciones.
2. Dinámica de utilización de las RSA. Sus implicancias sobre el concepto de “ciclo de vida” en el desarrollo de software.
3. Utilización (o reutilización) individual e institucional del conocimiento producido en las RSA. Problemas de seguridad, diseño, almacenamiento y acceso.
4. Problemática de la propiedad intelectual de los productos desarrollados y del conocimiento producido en la utilización de RSA.
5. Seguridad en las RSA. Sus *stakeholders* y consecuencias en el proceso de desarrollo.
6. Buenas prácticas en el desarrollo e implementación de RSA.
7. Los ajustes a las metodologías de la Ingeniería de Software para asegurar que las RSA sirvan a sus propósitos.

## Resultados y Objetivos

### Objetivo principal

Estudiar los efectos en distintos sectores institucionales durante la implementación de la RSA *ConexiónUB* en la Universidad de Belgrano. Obtener conclusiones sobre buenas prácticas en la implementación de RSA en el ámbito universitario.

### Objetivos Específicos.

- Identificar el elenco de funcionalidades básicas con la que se liberará la versión inicial de la red social académica.

- Implementar la RSA en su versión inicial.
- Implementar la RSA en su forma final.
- Generar repertorio de buenas prácticas de construcción.
- Redactar el informe final

## Formación de Recursos Humanos

El equipo está integrado por los tres firmantes del *paper*. Además se cuenta con el apoyo de alumnos avanzados de la carrera de Ingeniería Informática.

Se encuentra en curso el tratamiento de una tesis de grado dedicada a temas de movilidad de las RSA.

## Referencias bibliográficas

- [Cas2009] Castaño Garrido, C. (Coord) (2009) *Web 2.0: El uso de la Web en la sociedad del conocimiento: investigación e implicaciones educativas*, Universidad Metropolitana, Caracas, Venezuela, ISBN: 978-980-247-161-4.
- [Doo 1999] Dooley, Kim E. (1999) “Towards a Holistic Model for the Diffusion of Educational Technologies: An Integrative Review of Educational Innovation Studies”, *Educational Technology & Society* 2(4) 1999 ISSN 1436-4522.
- [Ert2009] Ertl, Bernhard (2009) “Conceptual and procedural knowledge construction in computer supported collaborative learning” en *CSCL'09 Proceedings of the 9th international conference on Computer supported collaborative learning - Volume 1* International Society of the Learning Sciences ISBN: 978-1-4092-8598-4.
- [Fer2011] Fernández, Diego (2011): “Sitios Web 2.0 aplicados al ámbito académico”, *Tesis de grado de Ingeniería de Sistemas*, Universidad de Belgrano, defendida el 9/08/2011. Tutora Lic. Ana M. Piccin.
- [Lev2007] Levin-Peled, Rachel et al (2007) “Promoting collaborative learning in higher education: design principles for hybrid

courses” en *CSCL'07: Proceedings of the 8th international conference on Computer supported collaborative learning*, Publicado por International Society of the Learning Sciences ISBN: 978-0-6151-5436-7.

[Oli 2012] Oliveros, Alejandro y Piccin, Ana M. (2012) “Web 2.0 para el ámbito académico: Implicancias en la Ingeniería de Software”, Wicc 2012- XIV Workshop de Investigadores en Ciencias de la Computación, 25-27 de abril de 2012, Posadas, Misiones, Argentina.

[Wel 1996] Wellman, Barry (1966) “For a Social Network Analysis of Computer Networks. A Sociological Perspective on Collaborative work And Virtual Community”. *SIGCPR/SIGMIS '96*, Denver Colorado US.

## Modelos y Tecnologías en Gobierno Electrónico

Luis Reynoso, Claudio Vaucheret, Guillermo Grosso, Marcelo Amaolo, Jorge Szneck, Daniel Dolz,  
Maximiliano Klemen, Marcelo Kazalukian

Grupo de Investigación en Gobierno Electrónico

Departamento de Teoría de la Computación, Programación e Ing. de Computadoras – Facultad de Informática  
Universidad Nacional del Comahue

Buenos Aires 1400, (8300) Neuquén

Contactos: {Claudio.Vaucheret, Luis.Reynoso, Guillermo.Grosso}@fai.uncoma.edu.ar

### Resumen

Una de las demandas más importantes de este siglo es la necesidad de acciones de gobierno eficaces promoviendo iniciativas de modernización para lograr un Estado más simple, conectado y eficiente. En ese proceso los logros en Gobierno Electrónico son claves, los cuales incorporan Tecnologías de la Información y Comunicaciones (TICs) en organizaciones del Estado con el objetivo de mejorar la gestión interna y externa. En dicha mejora de gestión es fundamental la interacción, coordinación y sinergia entre la Sociedad Civil, el sector Privado y el Estado.

Por ello es importante el estudio de tecnologías empleadas y requeridas por organizaciones públicas al definir sus propios objetivos en gobierno electrónico, adaptando sus procesos para brindar mejores servicios y con mayor alcance. Luego, certificar que estos objetivos y servicios estén integrados. No menor es el estudio sobre el desarrollo de nuevas TICs para lograr un estado más accesible e igualitario, sus mecanismos de participación y construcción de consenso, el estudio de interoperabilidad semántica de sus arquitecturas de software y servicios web, y la calidad de sus productos y modelos como la satisfacción de los actores involucrados. Dicho estudio global permitirá la especificación y desarrollo de consistentes modelos y

tecnologías aplicados al gobierno electrónico.

Palabras Clave: Gobierno Electrónico, Tecnologías de la Información y la Comunicación, Sociedad de la Información, Gestión Pública, Interoperabilidad, Voto Electrónico, Servicios Web.

### Contexto

La temática presentada se inserta en el contexto de los siguientes proyectos y convenios:

- Proyecto Métodos y Tecnologías para Gobierno Electrónico (Res. Falf 063/12).
- Convenio firmado entre la Facultad de Informática y la Provincia del Neuquén (14/05/2011) para desarrollo de actividades formativas, pasantías y específicas de la disciplina en ciencias de la computación.

### Introducción

El Gobierno Electrónico es el término que se utiliza para referir al uso de las TICs empleadas como herramientas para lograr un mejor gobierno. Bajo esta definición, el gobierno electrónico se distingue como la ruta que habilita el uso de las tecnologías, especialmente internet, para lograr buenos gobiernos, pues integra a las personas, los ciudadanos, los procesos, la información y la tecnología al servicio de las iniciativas y proyectos

gubernamentales. Los beneficios esperados en el sector público pueden resumirse en el incremento en la eficiencia de las operaciones del gobierno, fortalecimiento de la democracia, mejora de la transparencia de los actos y la provisión de mejores servicios para los ciudadanos, las empresas, el tercer sector y para dentro del propio gobierno.

Así, el uso de las TICs para promover el gobierno más eficiente y efectivo, facilita el acceso a los servicios que el gobierno ya brinda utilizando sus enormes potencialidades comunicacionales, brindando un acceso público a la información, y haciendo más cercanas a los ciudadanos las políticas y acciones del gobierno. El Gobierno Electrónico va mucho más allá de la utilización de otro medio de otorgar servicios a los ciudadanos y empresas, ya que se ha tornado en una oportunidad histórica para la reforma y transformación de sus actividades, habilitando esquemas antes utópicos, tales como la gobernanza participativa [2] y la comunicación directa para mejorar todos los servicios que debe ofrecer a su ciudadanía, transformando el gobierno centrado en el ciudadano.

El éxito del gobierno electrónico radica justamente en gestionar de una manera distinta la información, en compartir los datos e información para el uso público de una manera más racional y controlada, en cambiar la forma en que los funcionarios y empleados públicos interactúan entre ellos y con el público en general. Como un proceso que atraviesa todos los estratos y áreas del gobierno, el uso de la tecnología se comprende como un compromiso de cambio hacia nuevas prácticas, más que la incorporación de más computadoras y más infraestructura de redes de comunicación. Hacer procesos más eficientes con el uso de la

tecnología no constituye gobierno electrónico sino cambia la forma en la que los ciudadanos se pueden vincular con el gobierno, la mera acumulación de servicios informatizados no constituye más que la eficientización de procesos, que no logran nuevos servicios, y que no muestran eficacia en la aplicación de procedimientos y políticas.

Existen varios modelos que reconocen diferentes niveles de madurez en el uso de las TICs hacia la constitución de gobierno electrónico, distinguiéndose entre todos, aquellos modelos con mayores niveles de madurez, identificados como Gobierno Integrado y Gobierno Transformado, con libre disponibilidad de información segura y distinguida, servida a quienes requieren de su procesamiento y conocimiento. La base de estos esquemas de gobierno “continuo” de información compartida, es justamente la necesidad de que todas las agencias de gobierno sean capaces de compartir adecuadamente su información.

La *compartición de información* o *information sharing* es la clave requerida para un gobierno en red en el que no es posible distinguir sus áreas. Por ejemplo, compartir la información relacionada a un servicio entre distintas jurisdicciones (ya sean municipales, provinciales o incluso nacional, y entre entidades del mismo nivel) de forma completamente transparente para el ciudadano. Las iniciativas de compartir información gubernamental intentan “destrabar” los datos ubicados de manera fragmentada distribuidos en múltiples agencias o áreas, transformándolos en información integrada en plataformas o infraestructuras de datos accesible a usuarios autorizados para su consumo directo o su posterior procesamiento.

Así, la compartición de información atraviesa niveles intra- / inter-organizacionales, implicando diversidad



de actores y problemas no solo tecnológicos y culturales, los cuales tradicionalmente, han implicado que la mayoría de las iniciativas de compartir información han requerido importantes esfuerzos, en tiempo y costos.

## Líneas de Investigación y Desarrollo

Para abordar la problemática planteada en la sección anterior, el proyecto distingue dos líneas de investigación:

- ***Línea 1: Modelos y Tecnologías para Interoperabilidad, Transparencia & Legitimidad en Gobierno Electrónico:***

En materia de gobierno electrónico tanto la comunicación intergubernamental como la del gobierno con la sociedad en su totalidad adquieren gran relevancia. Se denomina *interoperabilidad* a la habilidad de dos o más sistemas o componentes para intercambiar información y utilizar la información intercambiada. La interoperabilidad facilita el flujo de información y, cuando los procesos e información están directamente accesibles desde aquellos involucrados que los requieren y, se provee suficiente información para comprender tal información, se puede indicar que la interoperabilidad contribuye a la *transparencia*, un saliente aspecto de una Buena Gobernanza señalado por Graham [2].

En relación a la línea de interoperabilidad el proyecto prestará especial énfasis en el estudio de niveles de madurez y calidad de aplicaciones interoperables de gobiernos provinciales. En esta misma línea de proyecto se incluye también la política provincial de integrabilidad [5], dentro de la cual se han difundido recientes trabajos desarrollados por miembros del presente grupo [9].

El proyecto de investigación también profundizará sobre compartición de información de Infraestructura de Datos Espaciales (IDEs) de Gobierno, las cuales integran un conjunto de organismos que intercambian información espacial a partir de la construcción colectiva de una infraestructura común, utilizando estándares abiertos (ej. WMS, Wep Map Service; WFS, Web Feature Service; WCS, Web Coverage Service, etc) para distribuir sus productos y servicios [6]. La interacción entre organismos es un componente imprescindible para lograr sinergia interinstitucional aprovechando las mejores prácticas de cada organismo y la actualización de la información [4].

El razonamiento acerca de estas temáticas y problemáticas de interoperabilidad, transparencia y compartición de información en conjunto contribuirá al fortalecer investigación relacionada al *desarrollo sostenible* y *ciudades inteligentes*. En una ciudad inteligente el capital social y humano y la infraestructura de comunicación moderna (TICs) y tradicionales (redes de transporte) confluyen hacia un desarrollo económico sustentable y una alta calidad de vida, con una administración sabia de recursos a través de la participación.

El análisis de redes e infraestructuras requiere de la aplicación de nuevos métodos heurísticos cuando las técnicas y algoritmos tradicionales no dan una solución satisfactoria. Se prevee la aplicación de algoritmos metaheurísticos para su aplicación en problemas relacionados a ciudades inteligentes.

Por otro lado, no se puede dejar de tener en presente, en relación a la investigación a abordar, el estudio de software de aplicación de redes sociales, el cual se ha tornado un instrumento importante en la formación de opinión y comunicación de la sociedad y las mismas incluso se están tornando en importantes objetos de

estudio como insumo para la toma de decisiones de los gobiernos [4]. En este tipo de estudios se debe resaltar la importancia que están adquiriendo estudios sobre Inteligencia Colectiva [7], Aplicaciones Web Sociales [8], Análisis de Redes Sociales [9] y Sociología Computacional. El proyecto prevé incluir estudios y difusión científica asociada a dichas áreas de estudio.

Una piedra angular de la noción de gobierno y democracia es la participación de distintos actores en la toma de decisiones de la sociedad. También es la única manera para que la gente pueda influir en las decisiones y políticas que afectan en sus vidas cotidianas. Hoy en día las tecnologías web 2.0 permiten a los ciudadanos nuevos mecanismos de expresión, las cuales deberían ser tenidas en cuenta en las políticas gubernamentales. Esta temática está en relación con la próxima línea de investigación.

▪ **Línea 2 Voto Electrónico:**

Desde otro aspecto debemos tener presente que todo buen gobierno trata de garantizar y propiciar mecanismos legítimos de participación. Uno de estos aspectos se relaciona con el tema de Voto Electrónico, el cual está en estrecha relación con el principio de “*legitimidad y Voz*” de una Buena Gobernanza [2]. Tal principio se refiere a mejoras en mecanismos de participación en la toma de decisiones así como favorecer y propiciar mecanismos de participación y consenso de manera constructiva. El desarrollo de modelos de software, transparentes y seguros, de Voto electrónico, será un importante medio facilitador para la implantación de nuevas prácticas de gobierno, y permitirá la adopción de mecanismos más ágiles y modernos en la toma de decisiones.

La realización de elecciones de autoridades del estado (o de otras institución social) a través del uso de las TICs es un aspecto concreto de los distintos procesos del Gobierno Electrónico. Este fenómeno, denominado por algunos autores como democracia digital, presenta muchas aristas que merecen un estudio particular.

El Voto Electrónico [1] se puede tomar como un concepto amplio que implica la realización de actos electorales utilizando las TICs, abarcando a todo el subconjunto de aspectos que hacen a los sistemas electorales: confección de padrones, mapas electorales, gestión de la logística electoral y sus resultados, etc. El Voto Electrónico posee también aspectos tecnológicos y aspectos sociológicos que deberían ser considerados. Cuando nos referimos a los aspectos tecnológicos podemos mencionar conceptos tales como votación remota, urnas electrónicas, auditabilidad del software, protocolos de seguridad, etc. En cuanto a los aspectos sociológicos, se puede mencionar la credibilidad y confianza de parte de la ciudadanía a su empleo, lo que conlleva tener que estudiar el grado de madurez de una determinada sociedad para que sea capaz de adoptar algún nivel de esta tecnología e ir analizando su evolución.

Podemos decir que existe un patrón evolutivo en la implementación del voto electrónico que requiere estudio y consideración. En las diferentes aproximaciones abordadas en la implementación de voto electrónico hay una serie de cuestiones recurrentes que deben resolverse, muchos de ellos relacionados con los aspectos de confiabilidad y seguridad, lo que se traduce en temas tales como privacidad o confidencialidad, anonimato, autenticidad, integridad, etc. El éxito de cualquier estrategia de votación electrónica va a radicar en la aceptación por parte de la

comunidad a la que está destinada. Y esto a su vez depende de un sinnúmero de factores, no solo los tecnológicos.

## Resultados y Objetivos

Los miembros del proyecto reúnen experiencia previa comprobable en relación a la temática del proyecto, los mismos han desempeñado cargos directivos de informática en áreas de gobierno tanto provincial como nacional, entre los que se destacan: Directivo de Oficinas de Tecnologías de Información Provincial (Decreto Prov. Neuquén 504/2004), Directivo de Sistemas en Direcciones Provinciales y Ministerios (Decretos Prov. Neuquén 157/2008, 137/2011, 2365/2005), Gestión Universitaria (Ord. C.S. UNComa 721/2010, Res. C.D. FaI-UNComa 04/2010), etc.

Las actividades informáticas más relevantes en materia de gobierno electrónico realizadas por miembros del proyecto de investigación incluyen:

- Trabajo conjunto con el gobierno de la provincia de Neuquén y Rio Negro.
- Especificación de Métricas para aplicaciones interoperables.
- Definición de Arquitecturas de integrabilidad de servicios sobre el protocolo HTTP.
- Difusión y marcos teóricos para Infraestructura de Datos Espaciales.

Además, si bien es reciente la creación del grupo se han enviado tres trabajos a congresos internacionales, contrastando con la comunidad científica resultados preliminares producto de investigación temprana.

## Formación de Recursos Humanos

El grupo reúne aproximadamente a 14 investigadores, entre los que se cuentan docentes y alumnos de UNComa y un asesor externo de UNU/IIST. Varios de los docentes-investigadores del grupo han

terminado o se encuentran próximos a terminar carreras de postgrado. Se cuenta actualmente con 1 doctor, 2 magisters, 1 doctorando, 2 estudiantes de maestrías.

## Referencias

- [1] Mahmud Aleuy (2007), La Votación Electrónica, Modernización del Régimen Electoral Chileno, 1ra. Edición. pp.. 221-240, ISBN: 978-956-8425-05-7, PNUD.
- [2] Graham, J., Amos, B., Plumptre, T. (2003) Principles for Good Governance in the 21st Century. Policy Brief No.15
- [3] Institute of Electrical and Electronics Engineers (1990). IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries. New York, NY.
- [4] Reynoso, L., Alvarez, M. (2012) Estructuras de Redes Sociales y Globales en Procesos Mesosociales y Macrosociales. VII Congreso Internacional de Tecnologías para la Educación y el Conocimiento. Tecnologías Emergentes. Universidad de Educación a Distancia, España- (<http://www.uned.es/infoedu/CIE-2012/index.htm>).
- [5] Laffite, R. (2008). Documento de Integrabilidad, Directiva Nro 1-001 GE2008. Normas y Procedimientos para Formulación, Desarrollo e Implementación de Aplicaciones Informáticas. Sec.G. P y C. E. Provincia de Neuquén.
- [6] Álvarez, M., Gallego, D., Zerpa, C. (2012). Las IDE y el Gobierno Electrónico: esbozando perspectivas futuras. En Bernabé - Poveda, M.A. y López - Vázquez, C.M. Fundamentos de las Infraestructuras de Datos Espaciales. UPM - Press; Colección Científica. Madrid (España).
- [7] Segaran, T. (2008). Programming Collective Intelligence: Building Smart Web 2.0 Applications. O' Reilly.
- [8] Bell, G. (2009) Building Social Web Applications: Establishing Community at the Heart of Your Site. O' Reilly.
- [9] Tsvetovat, M; Kouznetsov, A. (2009). Social Network Analysis for Startups: Finding connections on the Social Web . O' Reilly.
- [10] Sznek, J., Vaucheret, C., (2011). Arquitectura para la Integrabilidad de Servicios sobre el protocolo HTTP basado en PKI. 5° Simposio de Informática en el Estado SIE (40° JAIIO). Córdoba, Argentina. 2011. pp. 1-14

## El modelado de requerimientos en las metodologías ágiles

Silvia Rivadeneira<sup>1</sup>, Gabriela Vilanova<sup>2</sup>, María Miranda<sup>2</sup>, Diana Cruz<sup>1</sup>  
Departamento Ciencias Exactas y Naturales

<sup>1</sup>Unidad Académica Río Turbio, <sup>2</sup>Unidad Académica Caleta Olivia,  
Universidad Nacional de la Patagonia Austral  
[grivadeneira@uart.unpa.edu.ar](mailto:grivadeneira@uart.unpa.edu.ar), [vilanova@uolsinectis.com.ar](mailto:vilanova@uolsinectis.com.ar), [dianalrcruz@gmail.com](mailto:dianalrcruz@gmail.com)

### Resumen

Las metodologías ágiles surgidas a fines de los '90, no han sido muy tenidas en cuenta por el mundo académico, quizás recién a mediados de la década anterior se comienzan a encontrar trabajos de investigación al respecto.

Nos enfocamos en uno de los trabajos realizados en el marco de la investigación en el que se estudiaron seis metodologías ágiles para analizar que prácticas y artefactos implementaban en el proceso de desarrollo de proyectos de software, específicamente en las actividades de ingeniería de requerimientos. Aquí se rescatan las metodologías XP, SCRUM y APM como las únicas que abarcan estas etapas, solas o en combinación con otras tradicionales o actuales.

**Palabras clave:** modelado de requerimientos, gestión de requerimientos, metodologías ágiles

### Contexto

La Universidad Nacional de la Patagonia Austral (UNPA) está ubicada en la provincia de Santa Cruz y, está compuesta por cuatro unidades académicas dispuestas en las localidades de: Caleta Olivia, San Julián, Río Gallegos y Río Turbio.

Este artículo pertenece al Proyecto de Investigación 29/B134 denominado "Modelado de Requerimientos y Diseño de Sistemas Complejos" radicado en la Unidad Académica Caleta Olivia (UACO). El proyecto es del tipo 1 de Ciencia y Tecnología y responde a temáticas relacionadas con carrera de pre-grado, grado y postgrado dictadas en UNPA y comprende actividades de investigación básica y aplicada.

El proyecto de investigación es financiado parcialmente por UNPA.

El grupo de investigación conformado es nuevo en la UNPA y sus integrantes residen en varias localidades patagónicas, tales como Comodoro Rivadavia, Caleta Olivia, Río Turbio y Punta Arenas (Chile).

### Introducción

Las actividades de elicitación, documentación y verificación de requerimientos forman parte de la ingeniería de requerimientos y permiten conocer la realidad de una organización, pero es la inestabilidad o cambio de estos

requerimientos lo que complejiza el desarrollo de un proyecto, por lo que, aún hoy, buscando una gestión más adecuada de esos cambios, se sigue investigando sobre los métodos, técnicas y estrategias que permitan alcanzar resultados positivos en el desarrollo de software.

En este artículo se analizan las metodologías ágiles más utilizadas enfocándose en la etapa de gestión de requerimientos dentro del ciclo de vida del software, donde se explicitan las prácticas, técnicas y artefactos que se utilizan en la mencionada etapa. Para ello se indagó en la literatura existente para conocer cada una de las metodologías involucradas, logrando una primera selección y comparación de las mismas.

#### 1. Acercándonos a las metodologías ágiles

Las denominadas metodologías ágiles surgen en la década de los '90 y a comienzos del 2001 se reúnen sus representantes para discutir nuevos métodos, crear *The Agile Alliance* [2] y producir el Manifiesto por el Desarrollo Ágil del Software [1]. Si bien, el manifiesto ágil define una serie de principios y valores que guían al equipo del proyecto de software, comúnmente no se cuenta con especificaciones concretas relacionadas con aspectos metodológicos clave cubiertos [13]. En [11] se expresa que existen cuatro metodologías (XP, Scrum, DSDM<sup>1</sup> y FDD<sup>2</sup>) que se destacan por proveer una guía concreta, en su especificación, respecto de los procesos, prácticas, roles y productos. Pero en lo que respecta a las etapas de inicio del proyecto y especificación de requerimientos [9] XP y Scrum son las que proporcionan procesos, prácticas, actividades y artefactos específicos, así como la gestión del proyecto, y en este caso también podríamos agregar APM<sup>3</sup>.

Entre las características del enfoque ágil podemos expresar que impulsan generalmente una gestión de proyectos que promueve el trabajo en equipo, la organización y responsabilidad propia, un grupo de buenas prácticas de ingeniería de software que brindan una entrega rápida de software de alta calidad, y un enfoque de negocios que alinea el desarrollo con las necesidades del cliente y los objetivos de la empresa [10]. El ciclo de desarrollo que aplican las metodologías ágiles es iterativo e

<sup>1</sup> Dynamic System Development Methods

<sup>2</sup> Feature-Driven Development

<sup>3</sup> Agile Project Management



incremental, lo que permite entregar el software en pequeñas partes denominadas incrementos. Cada iteración se puede considerar como un sub-proyecto en el que las actividades de análisis de requerimientos, diseño, implementación y testing son llevadas a cabo con el fin de producir un subconjunto del sistema final. El proceso se repite varias veces produciendo un nuevo incremento en cada ciclo hasta que se elabora el producto completo. Aunque todas las metodologías ágiles adoptan este ciclo, cada una presenta sus propias características [16].

Entre las metodologías ágiles que cubren las primeras etapas de un proyecto de software podemos describir:

- **XP** [17] el proceso está constituido por seis fases (exploración, planificación, iteraciones, producción, mantenimiento, cierre de proyecto). Los roles son reducidos, un cliente, programadores, verificadores, consultores técnicos, un consejero, un rastreador y el Gran Jefe. Algunas de sus prácticas son: entregas pequeñas y frecuentes, diseño simple, prueba continua, refactorización continua, programación de a pares, integración continua, todo el equipo radica en el mismo lugar.
- **Scrum** [18] está indicado para proyectos en entornos complejos, donde se necesitan rápidos resultados y los requerimientos son altamente cambiantes o poco definidos. Es adaptativo, ágil, auto-organizado y con pocos tiempos muertos [12]. Fue concebido para utilizar en combinación con otras metodologías. Su proceso se caracteriza por sprints o iteraciones de un mes. El resultado del sprint es un incremento ejecutable que se muestra al cliente. Otra característica son las reuniones diarias que no llevan más de 15 minutos y su objetivo es coordinar e integrar el producto a entregar. El proceso se compone de 4 fases (pre-juego – planeamiento, pre-juego – montaje, juego y post-juego), 6 roles (scrum master, propietario, equipo, cliente, gestor y usuario), dispone de prácticas y herramientas para la gestión de sus fases tales como: product backlog, sprint backlog, estimación de esfuerzos, gráfico burn-down, gráfico burn-up, planning poker [9].
- **APM** [19] el proceso se compone de 5 fases (previsión, especulación, exploración, adaptación y cierre). Los roles que podemos encontrar son: patrocinador ejecutivo, gestor de proyectos, gestor de producto, ingeniero jefe, gestores, equipo de cliente, equipo de proyecto, proveedores y gobierno. Proporciona prácticas para cada

fase, tales como: caja de producto, hoja de datos de proyecto, listado de características de producto, gestión de carga de trabajo, reuniones diarias, decisiones consensuadas y participativas, revisión de adaptación del producto, proyecto y equipo. Al igual que Scrum es un marco de trabajo, y permite la combinación con otros enfoques de gestión de proyectos más formales.

**2. El proceso de requerimientos y las prácticas ágiles**

En la búsqueda por mejorar el proceso de gestión de requerimientos nuestro trabajo se centró en las actividades que se realizan en esa etapa y para ello analizamos que prácticas ágiles y artefactos suelen aplicarse en las metodologías antes mencionadas.

Historia de Usuario	
Número: 1	Nombre: Ingreso o supresión de Roles
Usuario: Administrador	
Modificación de Historia Número:	Iteración Asignada:
Prioridad en Negocio: Baja (Alta / Medio / Baja)	Puntos Estimados:
Riesgo en Desarrollo: (Alto / Medio / Bajo)	Puntos Reales:
Descripción: En la administración del sistema tendrá la opción de administrar usuarios, al ingresar a esta opción se desplegará un listado de los usuarios, los usuarios van a tener la opción de asignar roles, el administrador hace clic sobre esta opción relacionada con el usuario y el sistema le despliega el listado de roles disponibles para que el administrador seleccione los adecuados para ese usuario. Una vez el usuario administrador del sistema de la opción de guardar, el sistema pide confirmación y luego procederá a almacenar los cambios.	
Observaciones:	

Figura 1. Historia de usuario.

XP demanda comunicación oral tanto para los requerimientos como para el diseño [12]. Como puede observarse en la Figura 1, las historias de cliente, tarjetas de historia (story cards), o historias de usuario son tarjetas simples en papel donde se describen breves requerimientos similares a las tarjetas CRC, con una granularidad de 10 a 20 días. Se utilizan para estimar prioridades, alcances y tiempo de realización. Si existen discrepancias se utiliza la más optimista. Otros artefactos son lista de tareas en papel o pizarra y gráficos visibles en pared [9]. Una práctica que los equipos están utilizando es SDD<sup>4</sup> [15] como metodología para gestionar los requerimientos en proyectos ágiles.

Product Backlog	
	Fecha:
	Estimado:
Tipo: Nuevo ___ Mejora ___ Arreglo ___	Fuente:
Descripción:	
Notas:	

Tabla 1. Product Backlog.

Scrum no se ocupa de los detalles relacionados con el modelado de requerimientos, pero claramente el artefacto Product Backlog recoge los requerimientos del sistema y sus estimaciones. Los desarrolladores no pueden implementar los requerimientos si no están en el Product Backlog (ver Tabla 1). En cada iteración, el equipo de desarrollo comienza con la lista priorizada de

<sup>4</sup> Storytest-Driven Development

requerimientos almacenada en el Sprint Backlog. Luego, cuando los requerimientos surgen y evolucionan debido al entorno dinámico del negocio o una solicitud de cambio de las partes interesadas, la modificación de las prioridades de los requerimientos en el product backlog se produce sobre la base actual del valor de los requerimientos de negocio [24].

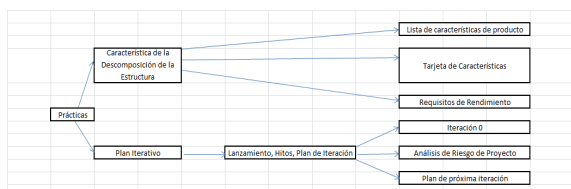


Figura 2. Prácticas de la Fase de Previsión.

En APM [19] la fase de previsión define la visión del producto, los objetivos del proyecto, la comunidad del proyecto y el equipo que trabajará. En la fase de especulación el equipo se ocupa de generar el plan de entregas basado en funcionalidades utilizando información de la especificación del producto, plataforma de arquitectura, recursos, análisis de riesgo, niveles de defecto, restricciones de negocio y fechas objetivo. Las prácticas ágiles que se utilizan se muestran en las Figuras 2 y 3.

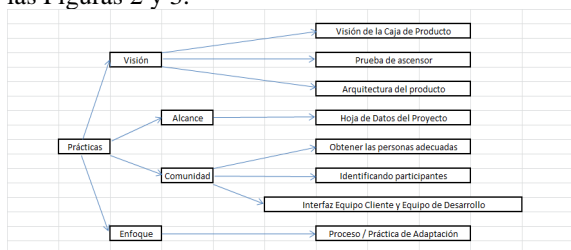


Figura 3. Prácticas de la Fase de Especulación.

Las metodologías analizadas se pueden combinar con otras que sean tradicionales o más actuales. Estas combinaciones pueden darse en una o en todas las fases para lograr cubrir todo el ciclo de vida. Las metodologías que cubren las etapas de requerimientos, suelen hacerlo en una o dos fases, en la práctica para obtener una visión general del sistema se trabajan con los requerimientos de más alto nivel, tanto funcionales como no funcionales, a través del consenso entre el equipo de desarrollo y los clientes; y en la especificación logran la especificación de requerimientos generales, refinándolos concurrentemente hasta llevarlos a especificaciones más detalladas [9].

Para los autores [12], [14] y [26] la documentación de los procesos ágiles es informal, prefiriéndose la comunicación cara a cara tratando de reducir los costos de mantenimiento.

En [14] y [25] se sugieren prácticas para aplicar a la gestión de requerimientos:

- Asegurar la participación del cliente. Negociar los requerimientos explícitos.
- Definir la trazabilidad de los requerimientos.

- Gestionar los cambios de requerimientos.
- Iterar-Inspeccionar-Adaptar.
- Construir prototipos y demostraciones.
- Realizar talleres de colaboración.
- Incorporar conceptos orientados a aspectos coincidiendo con [27].

### Líneas de investigación y desarrollo

Las líneas de investigación del grupo están centradas en las prácticas y artefactos para el análisis de requerimientos y el diseño de soluciones software, y en particular se están trabajando las siguientes:

- Modelado de requerimientos combinando BPMN y UML
- Modelado de requerimientos aplicando metodologías ágiles
- Integración de BPMN y SOA
- Arquitectura de Software y diseño de sistemas complejos

Por otro lado, no por ello menos importante, nos interesamos en la construcción del conocimiento en red, las siguientes líneas son las que involucra:

- La interacción que se produce en entornos colaborativos
- La construcción colaborativa del conocimiento

### Resultados y objetivos

Del análisis de literatura existente, encuestas realizadas por investigadores, experiencias exitosas, asociación a comunidades de desarrolladores ágiles locales y regionales, hemos obtenido información de las metodologías ágiles vigentes. En [9] hicimos una primera selección para analizar en detalle, teniendo en cuenta criterios como: (a) la metodología más presente en Internet, (b) la metodología más documentada, (c) la metodología con certificaciones y entrenamiento, (d) la metodología con comunidad propia, (e) la metodología más utilizada por organizaciones y (f) la metodología más utilizada en proyectos software; esto dio como resultado la selección de:

- Agile Project Management
- Crystal Methods
- Dynamic System Development Methods
- Extreme Programming
- Scrum
- Test Driven Development

De este listado surgieron aquellas que abarcaban las etapas iniciales de un proyecto, con prácticas y artefactos definidos y que, por supuesto, podríamos aplicar eran APM, XP y Scrum.

Existen trabajos que evalúan metodologías cualitativamente tales como [20], [21], [22], [23] y [24], otros lo hacen cuantitativamente [16] y en este caso evalúan si las metodologías ágiles cumplen o no con los principios del Manifiesto. Nosotros

reconocemos que seleccionar una metodología ágil a utilizar en un proyecto [3] requiere considerar cinco dimensiones de riesgo: tamaño, criticidad, dinamismo, equipo y cultura, y esto sin tener en cuenta el tipo de desarrollo.

Hemos generado una encuesta utilizando tecnología Google Docs que colocamos en línea durante cuatro meses y que difundimos en redes sociales, comunidades de desarrolladores ágiles y a través de correo electrónico pero no logramos obtener una muestra representativa (20 respuestas), es así que decidimos analizar otras encuestas, algunos de los resultados observados [4][5][6][7][8] muestran que:

- Los equipos ágiles están utilizando XP, Scrum, sus variantes y FDD.
- Las prácticas ágiles más utilizadas son las reuniones diarias.
- Las principales razones para implantar la agilidad son: incrementar el time-to-market, gestionar los cambios de prioridades y alinear el negocio e IT.
- Las prácticas menos utilizadas son: BDD<sup>5</sup>, la entrega continua y la programación por pares.
- Los procesos ágiles son apropiados para una gran proporción de proyectos.
- Las prácticas que más afectaron la calidad de las entregas fueron: guías de código comunes, refactorización, test de regresión y la integración continua.

Este estudio ha logrado que integrantes del proyecto incluyan esta temática en la planificación de los espacios curriculares a los que pertenecen y que por supuesto están relacionados con ingeniería de software y programación. Asimismo, los grupos de alumnos involucrados en la realización del trabajo final de carrera están utilizando algunas de sus prácticas.

El grupo espera durante este nuevo ciclo:

- Realizar eventos en el área de extensión de la Universidad relacionados con prácticas ágiles.
- Evaluar la aplicación de las prácticas ágiles durante el desarrollo del proyecto.
- Estudiar la integración entre las otras líneas de investigación y metodologías ágiles.
- Analizar las metodologías ágiles más adecuadas para trabajar en las organizaciones de la patagonia austral.

## Formación de Recursos Humanos

Entre los integrantes del proyecto existen dos alumnos de la Maestría de Informática y Sistemas – uno avanzado y otro iniciando -, dos

alumnos de la Maestría en Educación en Entornos Virtuales – uno de ellos en trabajo de tesis -, un alumno en proceso de trabajo final en la Especialización en Management Tecnológico y dos alumnos de pre-grado de Analista de Sistemas en proceso de realización de trabajo final de carrera, en estos casos uno de ellos está trabajando sobre esta temática en particular.

El grupo se complementa con un académico perteneciente a la Universidad de Magallanes doctorado en el área de computación y automatización industrial por la Universidad Estatal de Campiñas (Brasil) quien forma docentes y alumnos en la Unidad Académica Río Turbio (UART).

## Referencias

- [1] Beck, K., et al.: The Agile Manifesto. Manifesto for Agile Software Development. En: <http://www.agilemanifesto.org>
- [2] Agile Alliance, En: <http://www.agilealliance.org/>
- [3] Boehm, B. y Turner, R. “Balancing Agility and Discipline. A Guide for the Perplexed”. Addison-Wesley. 2004.
- [4] Ambler, S. Survey Says: Agile Works In Practice. En: <http://www.drdoobs.com/architecture-and-design/survey-says-agile-works-in-practice/191800169>. 2006.
- [5] Shine Technologies. Agile Technologies Survey Results. 2003.
- [6] VersionOne. State of agile survey: The State of Agile Development. 2011.
- [7] Alende, A. La utilización de los métodos ágiles en las empresas de desarrollo de software de Argentina. 2010.
- [8] Ali, M. Survey on the State of Agile Practices implementation in Pakistan. International Journal of Information and Communication Technology Research. Vol. 2. Nro. 5. 2012.
- [9] Rivadeneira, S. Metodologías ágiles enfocadas al modelado de requerimientos. ICT UNPA. 2013.
- [10] LNCS. Ingeniería del Software: Metodologías y Ciclos de Vida. Laboratorio Nacional de Calidad de Software. INTECO. España. 2009.
- [11] Meda, R. y Ierache, J. Una propuesta de conjunción de elementos metodológicos en común dentro de los enfoques ágiles para el desarrollo de software.
- [12] Reynoso, C. Métodos Heterodoxos en Desarrollo de Software. UBA. Argentina. 2004.
- [13] Larman, C. Agile and Iterative Development: A manager's guide. Addison-Wesley. 2004.

<sup>5</sup> Behavior-Driven Development

- [14]Cao, L. y Ramesh, B. Agile Requirements Engineering Practices: An Empirical Study. IEEE Software. 2008.
- [15]Mugridge, R. Managing Agile Project Requirements with Storytest-Driven Development. IEEE Software. 2008.
- [16]Mendes, K., Estevez, E., Fillotrani, P. Un framework para evaluación de metodologías ágiles.
- [17]Beck, K. Extreme Programming Explained. Embrace Change, Pearson Education, 1999.
- [18]Scrum Alliance. En: <http://www.scrumalliance.org/>
- [19]Highsmith, J. Agile Project Management: creating innovative products. Pearson Education. USA. 2004.
- [20]Abrahamsson, P., Salo, O., Ronkainen, J., Warsta, J., Agile Software Development Methods, Review and Analysis, VTT Publications 478. 2002.
- [21]Iacovelli, A., Souveyet, C., Framework for Agile Methods Classification, Workshop on Model Driven Informat. Systems Eng.: Enterprise, User and System Models. 2008.
- [22]Strode, D.E., The Agile Methods: An Analytical Comparison of Five Agile Methods and an Investigation of Their Target Environment, MSc Thesis in Information Systems, Massey University, Palmerstin North, Nueva Zelanda. 2005.
- [23]Visconti, M., Cook, C., An Ideal Process Model for Agile Methods, LNCS, ISBN 978-3-540-21421-2, Vol 3009, pp.439-441. 2004.
- [24]Oyeyipo E., Mueller C. Requirements Management in Agile-Scrum. Texas State University. San Marcos. 2011.
- [25]Gottesdiener E. "Requirements Practices on Agile Projects", Success with Requirements, Vol. 1, No. 8. 2007.
- [26]Qasaimah, M., Mehrfard, H., Hamou-Lhadj, A. Comparing Agile Software Process Based on the Software Development Project Requirements. IEEE Computer Society. CIMCA 2008
- [27]Araújo, J. y Ribeiro, J. Towards an Aspected-Oriented Agile Requirements Approach. WPSE'05.



## Evaluación de la calidad en sitios web bancarios

Pedro Alfonso<sup>1,2</sup>, Sonia I. Mariño<sup>1,2</sup>, María V. Godoy<sup>1,2</sup>,

<sup>1</sup>Área de Ingeniería Web. Facultad de Ciencias Exactas y Naturales y Agrimensura

<sup>2</sup>Departamento de Informática. Facultad de Ciencias Exactas y Naturales y Agrimensura.

9 de Julio 1449. CP: 3400. Corrientes. Argentina.

Universidad Nacional del Nordeste.

plalfonzo@hotmail.com simarinio@yahoo.com, mvgodoy@exa.unne.edu.ar

### CONTEXTO

En el trabajo se exponen los resultados iniciales obtenidos en la evaluación de la calidad en sitios bancarios, que operan en una provincia de la región NEA, República Argentina, en el marco del proyecto "Sistemas y TIC: técnicas y herramientas" (2012-2016), acreditado por la Secretaría General de Ciencia y Técnica (UNNE).

### RESUMEN

Se resume la línea de investigación vinculada con el estudio, análisis y evaluación de la calidad en sitios web bancarios.

En referencia a la formación de recursos humanos se concretó la defensa de una tesina de Especialización en Ingeniería de Software y se continúa con una tesis de Maestría de Ingeniería de Software, convenio Universidad de la Plata – Universidad Nacional del Nordeste.

**Palabras clave:** sistemas de información, calidad en portales bancarios, formación de recursos humanos.

### 1. INTRODUCCIÓN

En el marco del proyecto "Sistemas y TIC: técnicas y herramientas", una de las líneas temáticas que se aborda, se centra en los métodos y herramientas para evaluar la calidad de productos web. Desarrollar un sitio Web de calidad es una tarea imprescindible para todas aquellas empresas que desean brindar servicios a través de Internet. Por lo tanto, medir la calidad de los sitios que están siendo utilizados es de vital importancia para

todas aquellas organizaciones que pretenden posicionarse en Internet [1].

El desarrollo y evaluación de productos Web es una tarea difícil considerando todas las características y atributos deseables que se espera deba contener, por lo cual es recomendable contar con un modelo de calidad que sirva tanto para diseñar como para evaluar sitios. Los modelos de calidad surgen para describir dichas características, sus relaciones, como pueden ser medidas y como las mediciones pueden ser interpretadas [5].

Siguiendo lo expuesto por [5], existe una gran variedad de modelos de calidad, sin embargo, la mayoría están dirigidos a productos software genéricos y a la evaluación sobre el producto final y en menor medida orientado al contexto de la Web.

Por otra parte, para evaluar la calidad de sitios Web se han desarrollado diferentes métodos. Éstos se agrupan en categorías, como ser: testing, inspección, consulta, entre otros. Además, en relación al tipo de resultado que pueden obtenerse de su aplicación, pueden ser cuantitativos y cualitativos. En este sentido, existen diversas propuestas como las expuestas en [3], [10], [13] y [14], entre otros,

Según la encuesta realizada por D'Alessio IROL, para CertiSur S.A. [4], manifiesta que en Argentina, se incrementó el crecimiento de las operaciones de Home Banking en el período 2006-2012. El estudio expresa que el crecimiento de las operaciones de la banca online en 2006 versus las registradas en 2012

es muy significativo, pasando de 43% a 69%. Mientras que las operaciones de consultas y pagos se mantienen constantes, la banca transaccional se mantiene en crecimiento.

En este contexto, es imprescindible que los sitios bancarios sean de calidad. Además, de permitir que el sector financiero se vea beneficiado al suministrar información sobre los servicios ofrecidos y las personas puedan disponer de éstos sin restricciones espacio-temporales.

Los métodos expuestos, se enfocan al contexto en general y no distinguen el sector examinado, por lo tanto en este proyecto se consideran los trabajos específicos orientados al diseño y desarrollo de modelos de calidad aplicados a la generación y evaluación de sitios bancarios como los expuestos en [6], [7], [8], [11], entre otros. Teniendo en cuenta que los mencionados fueron diseñados y aplicados en diferentes países (Brasil y España), en [1] se contempló de relevancia estudiar las características propuestas por cada uno de ellos, a los efectos de la elaborar una propuesta integradora, aplicable a sitios Web bancarios en la Argentina.

Esta modalidad de comunicación ofrecida por los bancos, otorga beneficios para aquellas personas con algún tipo de discapacidad física. Además, evita a los usuarios desplazarse para realizar operaciones financieras, por lo que deben estar preparados para ser accesibles para la mayoría de las personas, presenten o no alguna discapacidad.

En este sentido, se considera que un sitio web, que evidencie problemas de acceso al contenido, está determinado por un diseño que no sigue las pautas y estándares web establecidos por la World Wide Web Consortium (W3C) [17]. Su objetivo es definir pautas que faciliten el acceso de las personas con discapacidad, a los contenidos web. A través de herramientas se evalúa y repara la accesibilidad y por consiguiente el diseño de dichos entornos.

Según la W3C, hablar de Accesibilidad Web es hablar de un acceso universal a la Web, independientemente del tipo de hardware, software, infraestructura de red, idioma, cultura, localización geográfica y capacidades de los usuarios.

Martín [9], expresa que en el contexto de la web, un sitio, portal y/o aplicación sea “accesible” significa que debe presentar la información a las personas, de manera tal, que independientemente de la tecnología que utilicen (computadora de escritorio, PC, PDA, teléfono, etc.) y de las capacidades diferentes que esta personas posean (físicas, psicológicas, sensoriales, etc.) todas ellas estén en igualdad de condiciones en lo que se refiere al acceso de contenidos y servicios ofrecidos por la Web.

Por lo tanto, uno de los temas abordados en el presente trabajo, es la inclusión de estándares técnicos en relación a las Pautas de Accesibilidad al Contenido en la Web 1.0 (WCAG 1.0).

## 2. LINEAS DE INVESTIGACION y DESARROLLO

El objetivo principal de esta investigación, es la elaboración de un método de evaluación, que permita la valoración cuantitativa de la calidad de sitios Web bancarios que operan en la República Argentina. El mismo se basa en características más relevantes consideradas para el dominio citado. Además, se tiene en cuenta la inclusión de estándares técnicos en concordancia a las WCAG 1.0, pensada esta como un atributo de calidad y aspectos relacionados con las normas regulatorias del sistema financiero de este país.

En la línea de I+D, con mira a la transferencia de conocimiento en sistemas web bancarios de calidad, se aborda:

- Relevamiento de modelos de calidad y métodos de evaluación orientados a productos Web.
- Estudio de modelos de calidad y métodos de evaluación.

- Relevamiento y estudio de modelos de calidad orientados al dominio bancario.
- Relevamiento, selección y estudio de metodologías ([12], [15], [16]) para el tratamiento de la accesibilidad web.
- Relevamiento y estudio de metodologías específicas para evaluar la accesibilidad de sitios web bancarios.
- Valoración cuantitativa de la accesibilidad de sitios bancarios en Argentina.

### 3. RESULTADOS OBTENIDOS/ESPERADOS

En esta sección se mencionan los logros alcanzados en la línea temática especificada y vinculados al mencionado proyecto que facilita la concreción de actividades de investigación aplicada, desarrollo y transferencia hacia la sociedad.

Los resultados se plasmaron en la tesina de Especialización en Ingeniería de Software, aprobada en el año 2012 [2] y se continúa una tesis de Maestría de Ingeniería de Software, convenio Universidad de la Plata – Universidad Nacional del Nordeste.

### 4. FORMACIÓN DE RECURSOS HUMANOS

En referencia a la formación de recursos humanos, se mencionan como logros del año 2012:

- En proyectos de TFA se incluyen como temas de calidad, la accesibilidad desde etapas tempranas del ciclo de vida. Es decir, se incorporó este concepto desde la etapa de análisis y diseño del sistema de información.
- Se seleccionan y estudian métodos herramientas, para la evaluación y medición de accesibilidad en sitios web bancarios de la región NEA y sus resultados serán publicados en el presente año.

### Referencias

[1] Alfonzo, P. L. y Mariño, S. I. (2011). “Revisión de modelos de calidad orientados a sitios Web bancarios. Estudio

preliminar”. Revista Técnica Administrativa. Vol. 10 Nro. 04 Buenos Aires, Argentina. 15-10-2011.

[2] Alfonzo, P. L. (2012). “Revisión de modelos para evaluar la calidad de productos Web. Experimentación en portales bancarios del NEA”. En: <http://sedici.unlp.edu.ar/handle/10915/19878>.

[3] Barnes, S. y Vidgen, R. (2001). “Assessing the Quality of Auction Web Sites”. Proceedings of the Hawaii International Conference on Systems Sciences, Maui, Hawaii, January 4-6.

[4] CertiSur S.A. Seguridad en Internet: la visión de los usuarios 2012. En: <https://www.certisur.com/>

[5] Coral, C., Moraga, A., Piattini, M. (2010). Calidad del producto y proceso del software. Ed. Rama.

[6] Córdoba, J., Cachero, C., Calero, C., Genero, M. y Marhuenda, Y. (2007). “Modelo de Calidad para Portales Bancarios”. XXXIII Conferencia Latinoamericana de Informática (CLEI'07). San José, Costa Rica. Octubre 2007. En: <http://www.dlsi.ua.es/~ccachero/papers/clei07.pdf>

[7] Diniz, E., Morena, R., Adachi, T. (2005). “Internet Banking in Brazil: Evaluation of Functionality, Reliability and Usability”. Electronic Journal of Information Systems Evaluations, 8(1):41-50.

[8] Hernández Ortega, B., Martínez Jiménez, J., DeHoyos, J. (2008). “Calidad de la información Web en la banca electrónica”. International Congress "Marketing Trends". Venice.

[9] Martín, A.; Gaetán, G.; Saldaño, V.; Miranda, G.; Molina, S.; Pastrana, S.

- (2012). "Diseño y Evaluación tempranos para priorizar la Accesibilidad en la WWW". Anales Workshop de Investigadores en Ciencias de la Computación. WICC 2012 (Argentina).
- [10] Mich, L.; Franch, M. y Gaio, L. "Evaluating and Designing Web Site Quality". IEEE Multimedia, vol. 10, no. 1, pp. 34-43, Jan. 2003, doi:10.1109/MMUL.2003.1167920
- [11] Miranda, F., Cortes, R. y Barriuso, C. (2006). "Quantitative Evaluation of e-banking websites: an empirical study of Spanish Banks". The Electronic Journal Information System Evaluation, 9(2), 73-82.
- [12] Montero, Y. H.; Fernández, F. J. (2004). "Propuesta de adaptación de la metodología de Diseño Centrado en el Usuario para el desarrollo de sitios web Accesibles". Revista española de documentación científica. Vol. 27, N° 3, 2004, págs. 330-344. En: [http://www.nosolousabilidad.com/hassan/DCU\\_accesible.pdf](http://www.nosolousabilidad.com/hassan/DCU_accesible.pdf)
- [13] Olsina, L. y Rossi, G. (2002). Measuring Web Application Quality with WebQEM, IEEE Multimedia Magazine, ol. 9, N° 4, pp. 20-29.
- [14] Olsina, L. y Rossi, G. (2002). "A Quantitative Method for Quality Evaluation of Web Sites and Applications". IEEE Multimedia 9(4): 20-29.
- [15] Segovia, C. (2008). "Accesibilidad e Internet...para que todas las personas, con distintas capacidades o recursos, puedan acceder a Internet". En: [http://www.archena.es/files/accesibilidad\\_e\\_internet.pdf](http://www.archena.es/files/accesibilidad_e_internet.pdf)
- [16] UWEM. Unified Web Evaluation Methodology. En: <http://www.wabcluster.org/> <http://www.wabcluster.org/uwem1/>
- [17] w3C Consortium. (1999) *Web Content Accessibility Guidelines 1.0*, W3C Working Draft, <http://www.w3.org/TR/WCAG10/>.



## Extendiendo BPMN2 para soportar Workflows Científicos de ESTECO

Corina Abdelahad, Daniel Riesco,  
Departamento de Informática  
Universidad Nacional de San Luis  
Ejército de los Andes 950 – San Luis – Capital – Argentina  
C.P.: 5700  
Tel.: 54-02652-424027 – Int. 251  
[cabdelah, driesco]@unsl.edu.ar

Alessandro Carrara, Carlo Comin, Carlos Kavka  
Research and Development Department  
ESTECO S.p.A.  
Area Science Park  
Padriciano 99  
34149, Trieste, Italia  
[carrara, comin, kavka]@esteco.com

### Resumen

En la última década, los procesos de negocio han adquirido gran popularidad, y derivado de ello han surgido diversos estudios al respecto. Un proceso de negocio es un conjunto de actividades coordinadas, diseñadas para producir una salida específica.

Por otra parte, la empresa ESTECO tiene una amplia experiencia en el ámbito de workflows de ingeniería aplicada en el contexto industrial [1]. ESTECO utiliza un modelo de workflow propietario que ha demostrado ser útil en el contexto de procesos de ingeniería. Si bien el modelo propietario cubre las necesidades actuales, el uso de un estándar permitiría claramente construir modelos de manera unificada.

Hay numerosos estándares para el modelado de procesos de negocios, pero por lo general no son directamente aplicables al ámbito de procesos de

ingeniería o computación científica. BPMN (Business Process Model and Notation) es un estándar clave en el modelado de procesos de negocio. Su notación gráfica muestra las distintas etapas de un proceso de negocio. Además su metamodelo está diseñado para ser extensible, esto permite extenderlo manteniendo la compatibilidad. La última versión del estándar ha sido desarrollada teniendo en cuenta algunas limitaciones y dándole mayor importancia a los objetos de datos persiguiendo el objetivo de poder interpretar y ejecutar directamente los modelos.

**Palabras claves:** BPMN, Procesos de Negocios, ESTECO, workflow, extensión, Metamodelo OMG.

### Contexto

El presente trabajo se enmarca en el Proyecto de Investigación: Ingeniería de

Software: Aspectos de alta sensibilidad en el ejercicio de la profesión de Ingeniero de Software – Facultad de Ciencias Físico-Matemáticas y Naturales, Universidad Nacional de San Luis, y la cooperación entre el LaCIS (Laboratorio de Calidad e Ingeniería de Software) de la UNSL y ESTECO S.p.A.

## Introducción

Es importante a la hora de modelar un proceso de negocio poder utilizar una herramienta independiente de la implementación, y por lo tanto, poder utilizar la especificación del proceso de negocio para diferentes plataformas.

La integración con plataformas del tipo Arquitectura Orientada a Servicios (Service Oriented Architecture - SOA) y sistemas de Computación en las Nubes (Cloud Computing) es esencial en el contexto de workflows industriales, esta característica no es manejada en modo adecuado con los típicos workflows de negocios [2][3].

BPMN es utilizado para el modelado de procesos de negocio y define una notación la cual es una plataforma independiente en relación con las definiciones específicas de los procesos de negocio. BPMN también permite la colaboración entre procesos de negocios de distintas organizaciones.

BPMN2 es la última definición del estándar BPMN para modelar procesos de negocios [7]. Este ha sido desarrollado solventando las limitaciones que poseía y dándole mayor importancia a los objetos de datos persiguiendo el objetivo de poder interpretar y ejecutar directamente los modelos. Además proporciona un conjunto de elementos de extensión, los cuales

permiten adoptar nuevos atributos y elementos a los elementos ya existentes al estándar BPMN. Esta nueva definición del estándar permite por primera vez la posibilidad de extender el uso de workflows del ámbito de procesos de negocios al ámbito de la ingeniería [4].

Los elementos de la notación están especificados en el metamodelo BPMN [3]. Este metamodelo se define en un nivel M2 de la OMG y se basa en MOF.

Un metamodelo es un modelo de un lenguaje de modelado [5] que describe el conjunto de modelos admisibles. Un metamodelo define el lenguaje con el cual se construyen modelos. Un metamodelo define formalmente los elementos de un lenguaje de modelado junto con sus relaciones y restricciones.

Por otro lado, ESTECO es una empresa la cual ha desarrollado un modelador de workflows propietario junto con un motor workflow asociado. Esta empresa es líder a nivel mundial como proveedor estratégico de tecnología de integración, simulación y optimización, así como servicios de consultoría a ingenieros e investigadores en distintos tipos de industrias. Si bien el modelador propietario cubre las necesidades actuales, el uso de un estándar permitirá construir modelos de manera unificada y estandarizada.

QVT (Query View Transformation) es un estándar establecido para crear consultas, vistas y transformaciones de modelos [6] y es el lenguaje estándar que propone la OMG para expresar las consultas y definir las transformaciones. Con QVT es posible definir transformaciones genéricas entre metamodelos. De esta manera, cualquier instancia de un metamodelo puede ser transformado en una instancia de otro metamodelo.

Una transformación de modelos es una transformación de uno o más modelos fuentes a uno o más modelos destinos, cada uno basados en sus metamodelos. Se puede decir entonces que instancias de un metamodelo pueden ser transformadas en instancias de otro metamodelo. Una definición de transformación es un conjunto de reglas de mapeo que todas juntas describen cómo un modelo origen puede ser transformado en un modelo destino. Cada regla de mapeo describe que uno o más elementos del modelo fuente deberán ser transformados en elementos del modelo destino. Cuando todas las reglas de mapeo han sido aplicadas la transformación se considera completa.

La especificación QVT tiene una naturaleza híbrida porque es declarativa e imperativa, con la parte declarativa dividida en una arquitectura de dos capas que formará el framework para la ejecución semántica de la parte imperativa.

## Líneas de investigación y desarrollo

Los procesos de negocio han adquirido gran popularidad en los últimos tiempos. De esto surge la necesidad de estudiar y profundizar más acerca de este tópico.

Existen muchos estándares en el ámbito de la construcción y ejecución de procesos de negocios, pero generalmente no son aplicables al ámbito de procesos ingenieriles y científicos.

El interés de nuestra investigación tiene como objetivo utilizar el mecanismo de extensión que ofrece BPMN2 para soportar workflows científicos y poder realizar las construcciones de sus modelos.

Con el fin de llevar adelante este objetivo, es necesario realizar una transformación, que soporte las extensiones desde el metamodelo de ESTECO al metamodelo de BPMN2. Esta transformación permitiría la conversión de la mayoría de los workflows industriales de ESTECO a BPMN2 consintiendo su ejecución en motores workflow BPMN2.

## Objetivos

El presente trabajo tiene como objetivo realizar una transformación en QVT entre metamodelos utilizando el mecanismo que provee BPMN2 para extender su metamodelo, con el fin de construir gráficamente un workflow científico y producir código BPMN2.

## Formación de Recursos Humanos

Este trabajo está siendo realizado para el desarrollo de una tesis de maestría en el marco del Proyecto de Investigación. Una tesista de maestría de la UNSL visitó ESTECO S.p.A. para llevar adelante el estudio del metamodelo de ESTECO junto con las posibles extensiones en BPMN2 como parte del desarrollo de tesis.

## Referencias

- [1] ESTECO S.p.A. Industrial applications. <http://www.esteco.com>.
- [2] Cui Lin, Shiyong Lu, Xubo Fei, Artem Chebotko, Darshan Pai, Zhaoqiang

- Lai, Farshad Fotouhi, and Jing Hua. "A Reference Architecture for Scientific Workflow Management Systems and the VIEW SOA Solution" IEEE Trans. Serv. Comput. 2, 1 (January 2009), 79-92.
- [3] Gideon Juve and Ewa Deelman. "Scientific workflows and clouds". ACM Crossroads 16, 3 (March 2010), pp. 14-18. Spring 2010
- [4] Yolanda Gil, Ewa Deelman, Mark Ellisman, Thomas Fahringer, Geoffrey Fox, Dennis Gannon, Carole Goble, Miron Livny, Luc Moreau, and Jim Myers. "Examining the Challenges of Scientific Workflows". Computer 40, 12 (December 2007), 24-32.
- [5] [11] J.M. Favre. "Towards a Basic Theory to Model Model Driven Engineering". 3Rd Workshop in Software Model Engineering, WiSME, 2004
- [6] OMG document number:  
Formal/2011-01-01  
<http://www.omg.org/spec/QVT/1.1/>
- [7] OMG document number:  
Formal/2011-01-03  
<http://www.omg.org/spec/BPMN/2.0/>



## Modelo para Aplicaciones Sensibles al Contexto: Validación y Evaluación

Gálvez, M.; Cáceres, N. R.; Velázquez, E. C.; Guzmán, A. N.

Facultad de Ingeniería

Universidad Nacional de Jujuy

mdpgalvezdiaz@fi.unju.edu.ar

### Resumen

El Grupo de Ingeniería de Aplicaciones Sensibles al Contexto (GRISECO) ha desarrollado un modelo para aplicaciones sensibles al contexto (MASCO), basado en capas y en mecanismos de dependencia en orientación a objetos, que contempla todas las categorías de la información y tipos de entidades que caracterizan el contexto. El modelo considera el tratamiento de más de una variable de contexto, la relación de una entidad con más de una variable de contexto, como así también la interacción entre variables de contexto y entidades. El propósito de este proyecto es continuar el refinamiento del modelo y realizar la validación y evaluación.

**Palabras clave:** Ingeniería de Software - Aplicaciones sensibles al contexto - Orientación a Objetos - Modelado - Validación - Evaluación.

### Contexto

La presente investigación se enmarca en el proyecto “Modelo para Aplicaciones Sensibles al Contexto: Validación y Evaluación”, cuyo objetivo principal es completar la validación y evaluación del modelo desarrollado por GRISECO.

Este proyecto, categoría A, está aprobado por la Secretaría de Ciencia y Técnica de la Universidad Nacional de Jujuy y se encuentra bajo incentivo.

La Ingeniería de Software constituye actualmente una línea prioritaria de investigación y desarrollo de la Facultad de Ingeniería de la Universidad Nacional de Jujuy.

### Introducción

MASCO tuvo su origen en una extensión del modelo presentado en Gordillo *et al.* (2006), que considera servicios sensibles a la variable de contexto ubicación y al perfil del usuario, tomando como referencia el framework Context Toolkit (Dey, 2001), basado en Widgets y el modelo de referencia para automatización de procesos industriales CIM (Computer Integrated Manufacturing, García Moreno, 1999) y fue evolucionando a través de los trabajos Quincoces *et al.* (2008, 2009, 2010) y Gálvez *et al.* (2009, 2010).

El modelo ha sido usado en el desarrollo de prototipos en distintos dominios, entre ellos: procesos industriales, domótica, juegos, robótica y aplicaciones con uso de telefonía celular. Esto ha permitido validar el modelo y comenzar con su evaluación.

Según Gómez la evaluación de una arquitectura se realiza para analizar e identificar riesgos potenciales en su estructura y sus propiedades que puedan afectar al sistema de software resultante, verificar que los requerimientos no funcionales estén presentes en la arquitectura, así como determinar en qué grado se satisfacen. (Gómez, 2007). También puede ayudar en la detección temprana de problemas, a tomar decisiones de diseño, lograr los atributos de calidad, validar requerimientos y lograr estimaciones más certeras, restricciones de implementación y mejorar su documentación. Cuanto antes se detecte un problema en un proyecto de software es mejor, y realizar una evaluación de la arquitectura es la manera más económica de evitar desastres (Brey *et al.*, 2005).

Existen un grupo de técnicas para evaluar que se clasifican en cualitativas y cuantitativas (Brey et al., 2005):

-*Técnicas de cuestionamiento o cualitativas.* Utilizan preguntas cualitativas en relación a la arquitectura: cuestionarios, checklists, escenarios.

-*Técnicas cuantitativas.* Sugieren tomar medidas cuantitativas utilizando métricas arquitectónicas, como acoplamiento, cohesión en los módulos, profundidad en herencias, modificabilidad. Se emplean simulaciones, prototipos y experimentos.

Por lo regular, las técnicas de evaluación cualitativas son utilizadas cuando la arquitectura está en construcción, mientras que las técnicas de evaluación cuantitativas, se usan cuando la arquitectura ya ha sido implantada (Gómez, 2007). Según Bass *et al.* (2003), una técnica efectiva es evaluar la arquitectura candidata antes que el proyecto sea aceptado debido a que reduce los riesgos y aumenta la calidad del producto a desarrollar.

Ante la posibilidad de efectuar evaluaciones en cualquier nivel del proceso de diseño, con distintos niveles de especificación, Kazman propone un esquema general en relación a la evaluación de una arquitectura con respecto a sus atributos de calidad. En este sentido, Kazman y sus colegas afirman que de la evaluación de una arquitectura no se obtienen respuestas del tipo “si - no”, bueno – malo” o “6.75 de 10”, sino que explica cuáles son los puntos de riesgo del diseño evaluado (Camacho *et al.*, 2004).

Su enfoque se orienta hacia la mitigación de riesgos, ubicando dónde un atributo de calidad de interés se ve afectado por decisiones arquitectónicas. Indica la importancia de la especificación exhaustiva de los atributos de calidad como base para efectos de la evaluación de una arquitectura de software. El punto es definir los atributos de calidad en función de sus metas y su contexto, y no como cantidades absolutas (Camacho *et al.*, 2004).

Según Kazman es posible realizarla en cualquier momento, pero propone dos

variantes que agrupan dos etapas distintas: temprana y tardía (Camacho *et al.*, 2004).

-*Temprana:* No es necesario que la arquitectura esté completamente especificada para efectuar la evaluación, y esto abarca desde las fases tempranas de diseño y a lo largo del desarrollo.

-*Tardía:* Cuando ésta se encuentra establecida y la implementación se ha completado. Este es el caso general que se presenta al momento de la adquisición de un sistema ya desarrollado

La evaluación la puede realizar gente Interna o Externa al proyecto aunque generalmente la realizan los miembros del equipo de desarrollo, arquitecto, diseñador, entre otros. (Gómez, 2007).

## Líneas de Investigación y Desarrollo

Se trabaja actualmente en el modelado y desarrollo de aplicaciones sensibles al contexto en distintos dominios, en el refinamiento y validación del modelo y en particular en su validación como modelo de arquitectura.

## Resultados y Objetivos

Este proyecto tiene estipulados dos años de duración (2012-2013) y los siguientes objetivos:

Objetivo general:

- Refinamiento, validación y evaluación del Modelo para Aplicaciones Sensibles al Contexto MASCO.

Objetivos particulares:

- Probar el modelo aplicándolo a distintos procesos industriales.
- Refinar el modelo obtenido por GRISECO.
- Validar el modelo
- Evaluar el modelo.
- Transferir los resultados obtenidos.

Durante el año 2012 se obtuvieron como resultado los siguientes trabajos:

- “Modelo para aplicaciones sensibles al contexto (MASCO): validación e inicio de

*evaluación*". María del P. Gálvez, Nélica R. Cáceres, Carolina V. Brouchy, Carola E. Velázquez, Omar M. González, Ariel N. Guzmán, Normando Romero y Viviana E. Quincoces. VIII Jornadas de Ciencia y Tecnología de las Facultades de Ingeniería del NOA. Facultad de Ciencias Exactas. ISSN 1853-7871, trabajo N° 357. Universidad Nacional de Tucumán. Argentina. 27 y 28 de septiembre de 2012.

- "Modelo para aplicaciones sensibles al contexto (MASCO): Descripción y comparación de implementaciones". María del P. Gálvez, Nélica R. Cáceres, Carola E. Velázquez y Ariel N. Guzmán. Primer Congreso Argentino de la Interacción- Persona Computador@, Telecomunicaciones, Informática e Información Científica – IPCTIIC 2012. UTN, Facultad Regional Córdoba. 30 de noviembre de 2012.

- "Integración de datos históricos y data stream en el modelo MASCO". Nélica R. Cáceres, Viviana E. Quincoces, María del P. Gálvez. VIII Jornadas de Ciencia y Tecnología de las Facultades de Ingeniería del NOA. Facultad de Ciencias Exactas. ISSN 1853-7871, trabajo N° 338. Universidad Nacional de Tucumán. Argentina. 27 y 28 de septiembre de 2012.

El desarrollo de estos trabajos permitió:

- Refinar la capa *Application* en el modelo para gestionar la interacción entre procesos cuando se trabaja con aplicaciones en donde intervienen varios procesos y la funcionalidad de coordinación de EntityAggregator.

- Trabajar en detalle con la capa *Hardware Abstraction*, en particular con el funcionamiento de los objetos Sensor y Actuator al trabajar con aplicaciones que usan varios sensores y actuadores.

- Refinar la capa *Context* con respecto al objeto variable de contexto ubicación (Location).

- Determinar los requisitos propios de una arquitectura en el modelo MASCO determinando que presenta: *Integridad*

*Conceptual*: es un modelo comprensible que tiene coherencia en su diseño esto queda demostrado al poder usar el modelo en diferentes aplicaciones rápidamente y sin complicaciones. *Corrección*: se pudo cumplir con los requisitos de todas las experimentaciones. *Compleción*: se cubrieron todos los requisitos planteados. *Capacidad de realización*: no se observaron mayores dificultades en la implementación de los ejemplos desarrollados.

- Se probó además que el modelo presenta los atributos de: *Modificabilidad*: el modelo permitió realizar cambios en sus componentes o relaciones para adaptarse a cada aplicación, esto deriva en la capacidad de extensibilidad al agregar nueva funcionalidad o mejorar la existente. *Simplificación de la funcionalidad*: el modelo permitió simplificar su funcionalidad al no usar aquellos componentes innecesarios según los requisitos del ejemplo sin que esto afectara a la funcionalidad general de MASCO. *Reusabilidad*: al permitir que sus componentes arquitectónicos pudieran ser reutilizados en el desarrollo de todas las experimentaciones.

## Formación de recursos humanos

Estructura del equipo de investigación:

Directora: Mg. Viviana Quincoces. Categoría de Investigación II.

Co- directora / Administradora del equipo: Mg. María del Pilar Gálvez. Categoría de Investigación III.

Investigadores:

- Ing. Nélica R. Cáceres. Categoría de Investigación V

Alumnos:

- Evelina C. Velázquez

- Ariel N. Guzmán

Los integrantes de este grupo de investigación participaron o participan en las siguientes actividades de formación de recursos humanos:

- Minicurso “*Context Aware y Data Mining*” dictado durante las VIII Jornadas de Ciencia y Tecnología de las Facultades de Ingeniería del NOA desarrollado en la Facultad de Ciencias Exactas y Tecnología de la Universidad Nacional de Tucumán, a cargo de: Mg. María del P. Gálvez, Ing. Nélide R. Cáceres, Carola E. Velázquez y Ariel N. Guzmán. 27 y 28 de septiembre de 2012.

- Curso de “*Modelado y Desarrollo de Software*”. Patrocinado por Programa Académico Microsoft Argentina & Uruguay y el grupo de Investigación GRISECO de la Facultad de Ingeniería de la Universidad Nacional de Jujuy, llevado a cabo los días 12,19 y 23 de Junio y 03 de julio del 2012 a cargo de las docentes: Mg. Viviana E. Quincoces, Mg. Pilar Gálvez e Ing. Nélide R. Cáceres, y de los alumnos Carola Velázquez y Ariel Guzmán. Resolución FI N° 165/12. del 2012.

- “Laboratorio de Desarrollo de Aplicaciones de *Windows 8*”, a cargo de Evelina C. González y Ariel Guzmán. Programa Académico Microsoft Argentina & Uruguay. Desarrollado en la Universidad Católica de Santiago del Estero, Departamento académico San Salvador, Jujuy. 21 de diciembre de 2012.

Dirección de Becas TICS

- Alumnas: Evelina C. Velázquez (DNI 29517497) y María C. Ledesma (DNI 31216450).

Dirección de Trabajo Final de Carrera:

- “Extensiones de UML para Sistemas de Tiempo *Real*”. Pablo Vilte (DNI 23.172.058) y Patricio Condorí (DNI 27.165.852). Carrera Ingeniería Informática. Res. FI 459/11.

## Referencias

Bass, L., Clements, P. & Kazman, R. *Software Architecture in Practice*. Segunda Edición. ISBN 0-321-15495-9. Editorial Addison- Wesley Professional. U.S. 2003.

Brey, G. A., Escobar, G., Passerini, N. & Arias, J. *Arquitectura de Proyectos de IT. Evaluación de Arquitecturas*. Buenos Aires: Universidad Tecnológica Nacional. Facultad Regional de Buenos Aires - Departamento de Sistemas, 2005.

Camacho, E., Cardeso, F., Nuñez, G., *Arquitecturas de Software*. 2004.

Dey, A.K., *Providing architectural support for building context aware applications, PHD Thesis*. Georgia Institute Technology, USA, 2001.

Gálvez, M.P., Quincoces, V.E. & Cáceres, N.R., *Refinamiento del Modelo para Automatización de Procesos Industriales MAPRIN*. Tercer Simposio Internacional de Investigación. La Investigación y su Transferencia a la Comunidad. Universidad Católica de Santiago del Estero. San Salvador de Jujuy, Jujuy. Publicado en Resúmenes pág. 105. 2009.

Gálvez, M.P.; Quincoces, V, Cáceres, N. & A. Vega, *Refinamiento de un Modelo en Capas que Provee Servicios de Ubicación para Aplicaciones Sensibles al Contexto*. III Congreso Internacional de Telecomunicaciones, Tecnologías de la Información y las Comunicaciones. Quito, Ecuador, 2010.

García, M., *Automatización de Procesos Industriales*. Editorial Universidad Politécnica de Valencia. España, 1999.

Gómez, O. S. *Evaluando Arquitecturas de Software. Parte I*. Panorama General. 01, México: Brainworx S.A, 2007. 1870-0888.

Gordillo, S., G. Rossi & A. Fortier, *Engineering Pervasive Services for Legacy Software*, Proceedings of the 1st International Workshop on Software Engineering for Pervasive Services, Lyon, France, 2006.

Quincoces, V.E, Gálvez, M.P, Méndez, S.A; Ramos, H.O, Zapana, J.V., Vega, A.A, & Cáceres, N.R., *Modelo orientado a objetos sensible al contexto basado en capas para aplicaciones de automatización de procesos industriales*. Investigaciones en Facultades de



Ingeniería del NOA, ISBN 978-987-1341-37-5. Editorial ECU UNCa, Catamarca, Argentina, 2008.

Quincoces, V. E., Gálvez, M. P., Cáceres, N. R., Vega, A. A., & Ramos, H. O. *Extensión de un modelo en capas que provee servicios para aplicaciones sensibles al contexto*. Investigaciones en Facultades de Ingeniería del NOA, ISBN 978-987-633-041-1. Editorial EUNSa, Salta, Argentina, 2009.

Quincoces, V. E., Gálvez, M. P., Cáceres, N. R., & Vega, A. A. *Modelo que provee servicios para aplicaciones sensibles al contexto: Validación en etapas tempranas*. Investigaciones en Facultades de Ingeniería del NOA, ISSN 3367-5072. EdiUNJu. Universidad Nacional de Jujuy. Jujuy, Argentina, 2010.

## Optimización de la Calidad de los Sistemas Móviles

Susana I. Herrera<sup>1</sup>, Pablo J. Najjar Ruiz<sup>1</sup>, Sergio H. Rocabado<sup>2</sup>, Marta C. Fennema<sup>3</sup>,  
María C. Cianferoni<sup>1</sup>

(1) *Instituto de Investigación en Informática y Sistemas de Información, Facultad de Ciencias Exactas y Tecnologías, Universidad Nacional de Santiago del Estero*  
*sherrera@unse.edu.ar, najarpablo@yahoo.com.ar, ceci.cianferoni@hotmail.com*

(2) *Departamento de Informática, Facultad de Ciencias Exactas, Universidad Nacional de Salta*  
*srocabad@cidia.unsa.edu.ar*

(3) *Departamento de Sistemas, Facultad de Tecnología y Ciencias Aplicadas y, Universidad Nacional de Catamarca*  
*crisfen@yahoo.com*

### Resumen

Los sistemas o aplicaciones móviles son aquellos que se ejecutan total o parcialmente en dispositivos portables y ubicuos como los teléfonos celulares. Dado que actualmente la mayoría de la gente dispone de estos equipos, se trata de sistemas ampliamente usados pero que poseen desventajas originadas principalmente en restricciones en el tamaño de pantalla, en la capacidad de procesamiento y en la disponibilidad de energía. Ante esto surgió la necesidad de estudiar aspectos claves de Ingeniería del Software que impactan sobre la calidad de los sistemas móviles, en cuanto a la comprensibilidad, usabilidad y eficiencia.

El presente proyecto investiga acerca de la optimización de los sistemas móviles tomando como referencia el estándar de calidad de la norma ISO/IEC 25000 y la norma de procesos de ciclo de vida del software ISO/IEC 12207. Los investigadores pertenecen a diferentes universidades nacionales del NOA, quienes abordan esta problemática mediante el desarrollo de estudios cualitativos y cuantitativos. Las principales tecnologías que se estudian son: arquitecturas móviles alternativas, nuevas representaciones visuales, realidad aumentada y nuevos métodos ágiles de desarrollo aplicables a sistemas móviles.

Los avances científico-tecnológicos que se logren se aplicarán en desarrollos en dominios: educación, aprendizaje organizacional, turismo y ciencias.

**Palabras clave:** Sistemas móviles, calidad del producto software, arquitecturas móviles, realidad aumentada, visualización, métodos ágiles de desarrollo de sistemas móviles.

### 1 Contexto

Esta investigación se lleva a cabo en el marco del proyecto denominado "Optimización de la calidad de los

Sistemas Móviles mediante la implementación de nuevas arquitecturas, realidad aumentada, técnicas de visualización y redes móviles Ad-Hoc. Aplicaciones en m-learning y en gestión del conocimiento". A su vez, el proyecto forma parte de un Programa de Investigación en el área Informática que tiene como objetivo realizar investigación aplicada sobre nuevas tecnologías que permitan mejorar la calidad tanto del proceso como del producto software.

El equipo de investigación del proyecto está conformado por docentes de diferentes universidades nacionales del NOA: del Instituto de Investigaciones en Informática y del Departamento de Matemática de la Facultad de Ciencias Exactas de la Universidad Nacional de Santiago del Estero (UNSE), del Departamento de Sistemas de la Facultad de Tecnologías y Ciencias Aplicadas de la Universidad Nacional de Catamarca (UNCA) y del Departamento de Computación de la Facultad de Ciencias Exactas de la Universidad Nacional de Salta (UNSA).

Además, el proyecto cuenta con el asesoramiento de investigadores de los Institutos de Investigación de la Facultad de Informática de la Universidad Nacional de La Plata.

Está financiado por el Consejo de Ciencia y Técnica de la Universidad Nacional de Santiago del Estero, durante el período Enero 2012-Diciembre 2015.

### 2 Introducción

Para lograr ubicuidad, los sistemas móviles se apoyan en el uso de dispositivos móviles que poseen recursos limitados en cuanto a capacidad de procesamiento, memoria, tamaño de pantalla, entre otros aspectos [19, 24, 26, 28]. Por ello, el desarrollo de sistemas móviles involucra cuestiones propias e importantes al momento de utilizarlos, como ser: el diseño de interacción, posicionamiento, representación del espacio, formas de

censado, tratamiento del contexto, arquitecturas y herramientas de implementación [8, 12, 18]. Todas estas cuestiones constituyen las variables que impactan directamente en la calidad de la aplicación que se desarrolla [3, 14, 20, 30].

Las características propias de este tipo de sistemas hacen que sea necesario seguir métodos apropiados para su desarrollo. En el ámbito académico y de la industria existen diversos estudios realizados sobre métodos alternativos de desarrollo [4, 21, 22, 23, 29], entre los que se destacan la metodología Mobile-D [1, 21]. Es una metodología desarrollada por investigadores del Centro de Investigación Técnica de Finlandia, que habría obtenido una certificación CMMI (Capacidad de Madurez del Modelo de Integración) de nivel 2.

Esta investigación se inició con el relevamiento de las mencionadas características, a propósito se presenta una breve descripción [12].

Los servicios que ofrece la computación móvil se desarrollan y ofrecen desde diferentes enfoques, según su finalidad. Por ello, se generan metáforas que definen el grado y tipo de interacción del usuario y la forma de recorrer el espacio. Estas metáforas se determinan de acuerdo a las historias interactivas y a las ubicaciones físicas de objetos y usuarios. Estas metáforas pueden ser: búsqueda del tesoro, rompecabezas, dominó, palabras cruzadas o scrabble, recolectando información [9, 10, 17].

En cuanto a la arquitectura, un sistema móvil generalmente se basa en una aplicación cliente que se conecta a un servidor de aplicaciones que se encuentra en Internet [5, 6, 13, 28]. Este servidor, a su vez, utiliza los servicios de un proveedor de ubicación, un Sistema de Información Geográfico (GIS) y de la información provista por diversos puntos de interés. Sin embargo existen aplicaciones que se ejecutan totalmente en el cliente [25]. Más precisamente, las arquitecturas que actualmente están vigentes son: Arquitectura Servidor, Arquitectura Cliente, Arquitectura Cliente-Servidor o Arquitectura híbrida. En este último tipo se desarrollan aplicaciones para el Sistema Operativo Android [16], detectado como el de mayor uso en la región.

Por otra parte, existen dos formas de posicionamiento para indicar el lugar en que se encuentra ubicado un elemento relevante para el sistema: Posicionamiento Geométrico, Posicionamiento Simbólico. Para posicionar a un usuario en una aplicación móvil, se necesita también una representación visual del espacio relevante para el usuario: un plano con las calles, un plano del edificio, etc. Esto se brinda mediante: imágenes (JPG, GIF) o por modelos de representación (cartografía: raster o vector) [11, 18].

Cuando se conoce la posición del usuario, se la obtiene utilizando una de las siguientes técnicas de censado de posiciones: GPS, sistema de antenas, tags [7]. Para el posicionamiento in-door es más apropiado el bluetooth o sistema de sensores.

Los teléfonos móviles o celulares utilizan, típicamente, tecnologías especialmente desarrolladas para ese tipo de dispositivos las cuales se han ido clasificando en diferentes generaciones [2, 28]. En Argentina, actualmente, está vigente la 3<sup>o</sup> Generación (3G), caracterizada por la convergencia de voz, datos y acceso inalámbrico a Internet a mayor velocidad; es apta para aplicaciones multimedia (envío y recepción de imágenes estáticas y video, servicios de ubicación geográfica, televisión en tiempo real, juegos, etc.). En países más avanzados como Japón, se utilizan dispositivos de 4<sup>a</sup> Generación y se experimenta con la 5<sup>a</sup>.

Se considera necesario el estudio de las diversas metodologías propuestas para sistemas móviles y diseñar una metodología propia que considere, en sus actividades y técnicas, las características de estos sistemas y que siga el estándar ISO/IEC 12207 [15]. De acuerdo a lo relevado hasta ahora, la propuesta debe basarse en el lenguaje de modelado UML, tomando buenas prácticas como lo proveniente del Modelo Vista Controlador (MVC) [5, 6, 17, 27].

Además, se investiga acerca de técnicas y herramientas de visualización y de realidad aumentada que podrían utilizarse para mejorar la calidad, principalmente las características de comprensibilidad, eficiencia y la usabilidad, de las aplicaciones móviles. En este sentido, se seguirá el estándar ISO/IEC 25000 [14, 20].

### 3 Líneas de investigación y desarrollo

La principal línea de investigación de esta propuesta consiste en la optimización de los sistemas móviles; lo cual significa aumentar su calidad, tomando las características de eficiencia, usabilidad y comprensibilidad [14, 20].

Respecto a la eficiencia (velocidad y administración de recursos), se investiga sobre la relación entre ésta y el diseño e implementación de arquitecturas alternativas para las aplicaciones móviles.

En relación a la comprensibilidad y usabilidad (capacidad de ser aprendido y operado, capacidad de atracción), se investiga sobre la construcción de interfaces de usuario usando técnicas de visualización y de realidad aumentada.

En relación al enfoque de calidad interna, se estudian métodos ágiles usados o factibles de ser usados en el desarrollo de aplicaciones móviles.

### 4 Objetivos y resultados

El objetivo general que guía esta investigación aplicada es:

*Contribuir al mejoramiento de la calidad de las aplicaciones móviles mediante nuevas arquitecturas de diseño, técnicas de visualización, recursos de realidad aumentada y métodos ágiles de desarrollo.*

Para ello, se fijaron los siguientes objetivos específicos:

- a) Definir un ecosistema móvil que identifique y caracterice los diversos componentes involucrados en la Computación Móvil (dispositivos, sistemas operativos, herramientas y librerías de programación y emulación, et.) de la región NOA, con un nivel de abstracción que permita aplicarlo alternativamente a otros contextos.
- b) Analizar las diversas arquitecturas de aplicaciones móviles (servidor, cliente, cliente-servidor, basados en web-services, etc.) y proponer una arquitectura que aumente su eficiencia en términos de velocidad, uso de memoria, procesador y energía.
- c) Determinar criterios y estrategias que permitan diseñar y construir interfaces de usuario que aumenten la usabilidad de las aplicaciones móviles, mediante el estudio de técnicas de visualización de información, realidad aumentada y metáforas de interacción.
- d) Diseñar una propuesta metodológica que guíe el desarrollo de aplicaciones móviles, considerando los aspectos de usabilidad, eficiencia y comprensibilidad.

En cuanto a los resultados, dado que el proyecto finaliza en el año 2015, solo se alcanzaron resultados parciales en relación a los objetivos a) y b).

Con respecto al objetivo a) se logró definir un ecosistema del NOA que se revisa permanentemente debido al avance constante de la tecnología móvil y a la ampliación de la muestra [11, 12].

Con respecto al objetivo b), se logró diseñar e implementar una arquitectura híbrida cliente-servidor sobre la cual se monta una aplicación de m-turismo. La arquitectura fue publicada en [13]. Asimismo, se diseñó e implementó una arquitectura cliente sobre la cual se monta una aplicación móvil de ayuda a personas con deficiencia visual [25].

## 5 Formación de recursos humanos

La Directora del proyecto pertenece al Departamento de Sistemas de la UNCA. La Codirectora, al Departamento de Informática de la UNSE. La asesora en Sistemas Móviles pertenece al Laboratorio de Investigación y Formación en Informática Avanzada (LIFIA) de la UNLP y posee vasta experiencia y conocimientos en el desarrollo de aplicaciones móviles [17, 18]. El resto de los integrantes son jóvenes que se inician en tareas de investigación, alumnos avanzados de carreras de Informática. Por lo tanto se requiere un proceso permanente de apoyo y formación de éstos.

Esta investigación contribuye al desarrollo de una tesis doctoral en Ciencias Informáticas (UNLP) referida a un marco para el análisis y evaluación de experiencias de m-learning [12] y desarrollo de aplicaciones de m-

learning. Además, se realizan cinco trabajos finales de alumnos de la carrera Licenciatura en Sistemas de Información (UNSE) e Ingeniería en Informática (UNCa): un trabajo final sobre arquitecturas y eficiencia de aplicaciones móviles, otro sobre aplicaciones con arquitecturas clientes para personas con deficiencia visual, otro sobre visualización, otro sobre realidad aumentada y otro sobre métodos de desarrollo de aplicaciones. Se tiene previsto incorporar más alumnos de grado en los próximos años, a efectos de brindarles un marco apropiado para sus trabajos finales.

## Referencias

1. Abrahamsson, P., Hanhineva, A., Hulkko, H., Ihme, T., Jääliñoja, J., Korkala, M., Koskela, J., Kyllönen, P., Salo, O. *Mobile-D: An Agile Approach for Mobile Application Development*, in Proc of the OOPSLA'04 Conference. 2004.
2. Aravind, A., Tahir, H. *Towards modeling realistic mobility for performance evaluations in MANET*; University of Northern British Columbia, Prince George, BC, Canada; 2010.
3. Ardagna, C.A., Jajodia, S. *Privacy Preservation over Untrusted Mobile Networks*. CSIS-George Mason University; 2007.
4. Blanco, P., Camarero, J., Fumero, A., Warterski, A., Rodríguez, P. *Metodología de desarrollo ágil para sistemas móviles. Introducción al desarrollo con Android y el iPhone*. Tesis de Doctorado en Ingeniería de Sistemas Telemáticos. Universidad Politécnica de Madrid. 2009.
5. Burbeck, S. *Applications Programming in Smalltalk-80(TM): How to use Model-View-Controller (MVC)*. (TM) Smalltalk-80. 1997. Disponible en <<http://st-www.cs.uiuc.edu/users/smarch/st-docs/mvc.html>>. Consultado el 01/03/2013.
6. Deacon, J. *Model-View-Controller (MVC) Architecture*. Ultima revision: 2009. Disponible en: <<http://www.jdl.co.uk/briefings/mvc.pdf>>. Consultado el 01/03/2013.
7. Denso Wave. *Qrcode.com*. Sitio oficial del Estandar QR Code. Disponible en: <http://www.qrcode.com/en/index.html>>. Consultado el 01/03/2013.
8. Fortier, A, Rossi G., Gordillo S., Challiol, C. *Dealing with Variability in Context-Aware Mobile Software*. Journal of Systems and Software 83(6): 915-936 (2010).
9. Hansen, F., Kortbek, K., Gronbaek, K., Spierling, U., Szilas, N. *Mobile Urban Drama: Setting the Stage with Location Based Technologies*. ICIDS 2008, LNCS 5334, pp. 20–31, Springer-Verlag, 2008.
10. Hansen, F., Kortbek, K., Gronbaek, K. *Mobile Urban Drama for Multimedia-Based Out-of-School Learning*. ACM. 978-1-4503-0424-5, 2010.



11. Herrera, S. I. & M. C. Fennema. *Tecnologías Móviles Aplicadas a la Educación Superior*. Congreso Argentino de Ciencias de la Computación. La Plata, 2011.
12. Herrera, S. I., J. L. Goñi & M. C. Fennema. *El m-learning en la educación universitaria de posgrado*. Jornadas de Ingeniería del NOA. Catamarca, 2011.
13. Herrera, S. I., Najjar Ruiz, P. J., Ledesma, E., Rocabado, S. *Sistema de Información Móvil para Turismo Receptivo*. Revista Gestao e Conhecimento, Edición Especial, Anales del 8° Congresso Brasileiro de Sistemas. ISSN 1808-6594. Poço de Caldas, Octubre 2012.
14. International Standar Organization. *ISO/IEC 25000. Estándares de Calidad del Software*. 2011.
15. International Standar Organization. *Norma ISO/IEC 12207. Procesos de Ciclo de Vida del Software*. 2011.
16. Komatineni, S., MacLean, D. *Pro Android 4*. Ed. Apres. ISBN 1430239301, 9781430239307. 2012.
17. Lliteras, A., Challiol, C., Gordillo S. *Juegos Educativos Móviles Basados en Posicionamiento: Una Guía para su Conceptualización*. 13th Argentine Symposium on Software Engineering, 41 JAIIO, La Plata, 2012.
18. Lliteras, A., Challiol, C., Mostaccio, C., Gordillo S. *Representaciones enriquecidas para la navegación indoor-outdoor en aplicaciones móviles*. XVII Congreso Argentino de Ciencias de la Computación. La Plata, 2011.
19. Pernici, B. *Mobile Information Systems. Infrastructure and Design for Adaptivity and Flexibility*. Ed. Springer-Verlag, Germany, 2006.
20. Piattini M., García F. y Caballero I. *Calidad de Sistemas Informáticos*. AlfaOmega, México, 2007.
21. Pikkarainen, M. *Mapping Agile Software Development on to ISO 12207*. Information Technology for European Advancement. Febrero, 2006. Disponible en: <[http://www.agile-itea.org/public/deliverables/ITEA-AGILE-D2.9\\_v1.0.pdf](http://www.agile-itea.org/public/deliverables/ITEA-AGILE-D2.9_v1.0.pdf)>. Consultado el 01/03/2013.
22. Rahimian, V. *Performance evaluation of mobile software systems: Challenges for a software engineer*. 5th International Conference on Electrical Engineering, Computing Science and Automatic Control CCE. Noviembre, 2008.
23. Rahimian, V., Ramsin, R. *Designing an Agile Methodology for Mobile Software Development: A Hybrid Method Engineering Approach*. The Second IEEE International Conference on Research Challenges in Information Science (RCIS). Junio, 2008.
24. Reza B'Far. *Mobile Computing Principles*. Cambridge University Press, 2005.
25. Richard, P. I., Aranda, M. D., Bivanco, W., Acosta Parra, C., Herrera, S. I. *Sistema móvil para asistencia a personas con deficiencia visual*. VIII Jornadas de Ciencia y Tecnología de Facultades de Ingeniería del NOA. San Miguel de Tucumán, Septiembre 2012.
26. Roy, N., Scheepers, H. and Kendall, E. *Mapping the Road for Mobile Systems Development*. In Proceedings of Pacific Asia Conference on Information Systems (PACIS 2003), paper 94, pp. 1358-1371, 2003.
27. Schiller, J. H., Voisard, A. *Location-Based Services*. Ed. Elsevier. ISBN 0080491723, 9780080491721. 2004.
28. Talukder, A.K., Ahmed, H., Yavagal, R. *Mobile Computing: Technology, Applications, and Service Creation*. 2° Edición. McGraw-Hill communications engineering series. ISBN 9780070144576. 2010.
29. Vainio, A. M., Tuunanen, T., Abrahamsson, P. *Developing Software Products for Mobile Markets: Need for Rethinking Development Models and Practice*. In Proc. of the 38th Hawaii International Conference on System Sciences (HICSS'05). Enero, 2005.
30. Zhang, L., B. Tiwana, Z. Qian. *Accurate Online Power Estimation and Automatic Battery Behavior Based Power Model Generation for Smartphones*. ACM, 2010.

## Metodologías Ágiles para el Desarrollo de Sistemas de Cómputo Distribuido

Oscar Martín Bianchi

[mbianchi@ejercito.mil.ar](mailto:mbianchi@ejercito.mil.ar)CIDESO<sup>1</sup>, DIGID<sup>2</sup> - Ejército ArgentinoEST<sup>3</sup>, IESE<sup>4</sup> - Ejército Argentino

Alejandro J. M. Repetto

[arepetto@ejercito.mil.ar](mailto:arepetto@ejercito.mil.ar)CIDESO<sup>1</sup>, DIGID<sup>2</sup> - Ejército ArgentinoEST<sup>3</sup>, IESE<sup>4</sup> - Ejército Argentino

José Ignacio Ariznabarreta Fossati

[ignacioariznabarreta@gmail.com](mailto:ignacioariznabarreta@gmail.com)EST<sup>3</sup>, IESE<sup>4</sup> - Ejército Argentino**RESUMEN**

El Manifiesto Ágil, escrito en el 2001, establece las pautas para la utilización de las metodologías ágiles, las metodologías más populares y las lecciones aprendidas, para el desarrollo de software (1). Éstas surgen como alternativa a los procesos de desarrollo con alto costo en documentación y procesos excesivamente prolongados.

De la misma forma, las arquitecturas de computación distribuida y computación de grilla surgen como respuesta a la necesidad de aportar una solución y alternativa a los procesos con alta demanda de recursos.

En la Escuela Superior Técnica del Ejército, durante el año 2011, se llevó adelante el proyecto CoDiSe (Computación Distribuida para Seguridad Informática) (2), que sirvió como prototipo experimental para el proyecto CODEC (Computación Distribuida para Experimentación Científica) (3), el cual pretende dotar a los laboratorios de las distintas especialidades de la EST con un sistema capaz de procesar grandes cantidades de información o realizar cómputos a gran escala, sirviéndose de los infraestructura informática ociosa de la universidad.

El presente trabajo tiene por objetivo compartir las experiencias obtenidas en el uso de las metodologías ágiles en el ámbito académico, en el marco del desarrollo de un sistemas de computación distribuida como parte del dictado de la asignatura Lenguajes de Programación I.

**Palabras Clave:** *Computación Distribuida, Metodologías Ágiles, Computación en Grilla de Escritorio, Ingeniería del Software.*

**CONTEXTO**

Elegir una metodología de desarrollo de software apropiada es una preocupación siempre presente en la mayoría de los proyectos de IT, así como en los ámbitos académicos asociados a la misma actividad.

Las nuevas metodologías de desarrollo de software, a diferencia de las clásicas, son consideradas mucho más ágiles y fáciles de adaptar a los cambios. Éstas no hacen foco en el largo ciclo de desarrollo sino en iteraciones más cortas y ligeras, con el sello de una proximidad más acentuada a los *Stakeholders* (4).

Al respecto, las metodologías ágiles proveen la oportunidad de evaluar la dirección del proyecto a medida que se desarrolla el ciclo de desarrollo natural del sistema. Esto se alcanza a través del ritmo regular establecido de trabajo, conocido como iteraciones o *sprint*, al final de los cuales se puede presentar algún estado de terminación del producto (sistema en este caso).

Enfocándose en la repetición de ciclos de trabajo abreviados (o iteraciones) los equipos de trabajo tienen la posibilidad de corregir tempranamente desvíos al inicio de cada iteración, reduciendo de esta forma el impacto de cualquier error o falla cometido en las etapas previas (4) (5).

Asimismo, esta metodología de desarrollo en paralelo propuesta está basada en el

<sup>1</sup> CIDESO: Centro de Investigación y Desarrollo de Software

<sup>2</sup> DIGID: Dirección General de Investigación y Desarrollo

<sup>3</sup> EST: Escuela Superior Técnica - Facultad de Ingeniería del Ejército Argentino

<sup>4</sup> IESE: Instituto de Enseñanza Superior del Ejército - Universidad del Ejército Argentino

enfoque de inspeccionar y adaptar (*inspect-and-adapt*), solucionando o evitando problemas como la parálisis por análisis (*analysis paralysis*), la sobrecarga de información y el problema de la racionalidad limitada (*bounded rationality*), entre otros.

En ese aspecto, en los últimos años se ha observado un aumento exponencial en la práctica de metodologías ágiles de desarrollo de software. Esto se debe en gran parte al hecho que estas prácticas facilitan el trabajo en equipo, especialmente si los grupos están fragmentados o distribuidos, haciendo hincapié en el uso de herramientas de gestión como los *Sprint* y los *Scrum*.

Ambas herramientas ofrecen la posibilidad de simplificar la aproximación a desarrollos complejos, facilitando tanto la comprensión del problema, como la implementación y seguimiento de la solución. Estas cualidades hacen del *Scrum* una metodología de desarrollo viable para su utilización en el ámbito académico (6).

En cuanto a la tecnología implementada, el uso de computación distribuida (CD) para solucionar problemas de rendimiento o disponibilidad de recursos (necesarios para grandes cómputos) es una tendencia en crecimiento. Se puede observar en áreas tan diversas como la experimentación científica (con proyectos como SETI<sup>a</sup> y BOINC<sup>b</sup>) y la seguridad cibernética (tanto en su uso como herramienta de testeo de nuevas tecnologías, como instrumento de *hacking*).

La incorporación de conceptos como “*Grid Desktop Computing*” (GDC) (7) y “*Cloud Computing*” a las tecnologías ya existentes de CD, la vuelven en una herramienta valiosa no sólo para el desarrollo de software de cómputo de alto rendimiento, sino también como herramienta pedagógica de enseñanza que aporta una variedad de conocimientos y conceptos amplia, fundamentales para la formación profesional de los estudiantes de las carreras de IT.

Teniendo presente esto, y como evolución de experiencias previas, en el año 2011 se

decide avanzar con un prototipo experimental de un sistema de CD, de la mano del proyecto CoDiSe (2), obteniendo en el 2012, como fruto de este trabajo el proyecto CODEC (3). Este proyecto tiene como propósito el desarrollo de un sistema de computación distribuida multipropósito construido sobre la base del sistema BOINC. El fin de CODEC es poner a disposición de la EST un sistema con gran capacidad de procesamiento basado en una arquitectura de Computación de Grilla de Escritorio (7), aprovechando a su vez la infraestructura informática de la universidad, bajo el paradigma Infraestructura como Servicio (IaaS – *Infrastructure as a Service*).

Esta variedad de conceptos y tecnologías hacen de la comprensión y ejecución de un proyecto como CODEC una tarea compleja para aquellos participantes del mismo. Por esto, y dadas las características del proyecto, específicamente su origen académico como trabajo práctico de la cátedra de Lenguajes de Programación I, el uso de metodologías ágiles surge como la alternativa natural.

## INTRODUCCIÓN

Una de las características determinantes de CODEC es el hecho que cada uno de sus *sprint* es semestral (y no mensual, o bimestral como recomiendan los especialistas en el tema) coincidiendo con el ciclo académico de las materias asociadas. Por la misma razón, el proyecto posee en sí mismo una alta variabilidad en los equipos de trabajo (los que rotan anualmente), perfiles de los miembros, conocimientos, dedicación horaria e incluso área de experticia. Esto constituye un ambiente altamente hostil para el desarrollo de software, lo que llevo a la evaluación de diferentes aproximaciones y metodologías de desarrollo, surgiendo el enfoque ágil como la opción más próxima a las necesidades y naturaleza del proyecto

Entre las ventajas que nos aporta el enfoque ágil podemos mencionar la posibilidad de contar con versiones de alcance limitado en donde se podrá tener un producto tangible, el cual le será entregado en sus diferentes ver-

<sup>a</sup><http://setiathome.berkeley.edu/>

<sup>b</sup><http://boinc.berkeley.edu/>

siones en los *sprints* propuestos, así como una versión final en el tiempo acordado.

Con respecto a lo estrictamente técnico, la implementación y las pruebas se simplifican sensiblemente al subdividirse cada una de las tareas programadas en el sprint en hitos más atómicos.

Como antecedente, la metodología utilizada durante el año 2012 puede ser considerado un “híbrido” debido a que, si bien se utilizó la definición base de la metodología *scrum*, el marco en el que fue desarrollado el proyecto forzó a un *sprint* semestral, debido a que la realización del mismo se llevó a cabo en la cátedra “Lenguaje de Programación 1”, dictada semanalmente, en donde se realizaba el *scrum* semanal. Esto colaboró en la resolución de los problemas identificados a lo largo del proceso de desarrollo, ya que el hecho de contar con equipos reducidos y metas atómicas facilitaba la posibilidad de cumplirlas de forma autodidacta.

Por último, no podemos dejar de mencionar un aspecto fundamental: los recursos humanos. Sobre este particular, es importante señalar que esta metodología nos abstrae de la problemática de encontrar alumnos con diferentes capacidades, así como diferentes conocimientos (adquiridos de manera extra curricular o de forma autodidacta), además de la dedicación que se le brindaba a dicho proyecto, trayendo a consideración nuevamente, la situación de estudiantes de Ingeniería en Informática, así como de Ingenieros en Electrónica.

## LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

La metodología utilizada en CODEC plantea una línea de investiga clara con respecto a la estrategia a utilizar, siendo esta una variación de la metodología ágil, adaptada a las necesidades del proyecto. Entre sus características más importantes podemos mencionar:

- *Scrum* reducido: en lugar de realizar *Scrum* diarios, se decidió llevar adelante una práctica de *Scrum* semanales. Esto se debió a las características de la asignatura, la cual tenía una carga horaria muy limitada, lo que

sumado a la disponibilidad de cada uno de los alumnos y docentes, dejaba un marco horario extremadamente estrecho. Por esta misma razón se decidió realizar los *scrum* una vez por semana durante el horario de dictado de clase.

- *Sprint* semestral: cada una de las iteraciones se basa en una fase del proyecto, cada una de las cuales tiene la duración de un ciclo lectivo.

- Generación de una métrica que permita evaluar los progresos del proyecto y la efectividad del trabajo realizado durante cada iteración con respecto a la anterior. Sobre este último punto, la métrica propuesta evaluara una serie de cuestiones entre las que cabe mencionar:

- Formación de equipos: los equipos han de ser formados de teniendo en cuenta las habilidades de cada miembro, así como los conocimientos transversales que deben poseer.

- Trabajo interdisciplinario: dándose la particularidad que alumnos de dos carreras diferentes forman parte de CODEC, esta oportunidad es aprovechada para evaluar el trabajo en equipo entre futuros profesionales de distintas especialidades, tomando como parámetro el rendimiento de cada grupo individualmente.

- Evaluación de conocimientos: siendo de interés tanto para la cátedra como para la institución, la evaluación de conocimientos de cada participante del proyecto, previa y posterior al *sprint*, ofrece un claro parámetro del grado de asimilación de los conocimientos y aptitudes que pretende ofrecer el proyecto.

- Éxito individual de los grupos: cada *sprint* planteara una serie de objetivos a cumplir (generalmente asociados a pequeños ciclos de desarrollo o componentes de alcance acotado), considerando el éxito de cada grupo en el trabajo en el grado de cumplimiento de cada tarea asignada.

- Integración del trabajo grupal: la tarea de integración del trabajo de cada grupo (su dificultad, nivel de éxito, resultados en general, etc.) ofrece una clara perspectiva para evaluar el apego a los estándares de integración fijados así como la facilidad de



interpretación del trabajo de terceros entre los distintos grupos.

### RESULTADOS Y OBJETIVOS

Los resultados obtenidos durante el 2012 en lo que hace al rendimiento del curso revelaron que el uso de las metodologías ágiles desarrollo lograron:

- Facilitar la comprensión de la temática, y el trabajo a realizar.
- Mejor aprovechamiento del tiempo dedicado al proyecto y flexibilidad a los cambios.
- Sencillez en el seguimiento del progreso, la supervisión y el soporte a los alumnos/desarrolladores.

En lo que hace a las mejoras que esperamos obtener respecto al rendimiento del curso, haremos referencia al incremento en la calidad del software, lo cual es consecuencia de la implementación de tareas, lo que trae una reducción en cuanto a la complejidad y, por ende, soluciones más rápidas ante cualquier eventualidad (como la metodología de trabajo ágil propone).

En la siguiente iteración se pretende cumplir los siguientes objetivos:

- Reducción de riesgos, al poder atomizarse el trabajo y la posibilidad de elegir comenzar por las tareas más críticas, se reducen los riesgos fracaso del proyecto.
- Disminución del tiempo de desarrollo, se pueden ir realizando entregas medida que vayan estando disponibles las diferentes partes/objetivos/sprint del proyecto.
- Análisis previo para identificar, de forma prematura, las tareas críticas en cuanto a la realización, para que las mismas sean abordadas desde un comienzo, reduciendo los riesgos de fracaso del proyecto.
- La liberación de productos finalizados (*releases*) por ciclo. Es por ello que al realizar diferentes betas/*releases* a medida que se van cumpliendo a los diferentes *sprints* del proyecto, esperamos lograr una disminución del tiempo de desarrollo del sistema.

### FORMACIÓN DE RECURSOS HUMANOS

Una marca distintiva de CODEC es el hecho de haber nacido de la colaboración de dos centros de investigación dentro del Ejército Argentino (3). Y la particularidad que uno de dichos centros es parte funcional de la EST. Esto pone al proyecto en un ámbito privilegiado para la formación de recursos humanos.

Por un lado, el CIDESO tiene amplia experiencia en la formación de recursos humanos en el terreno de la investigación aplicada en sistemas de información de diversa índole, incluyendo sistemas de simulación para el adiestramiento, sistemas de información geográfica, sistemas de visualización, sistemas inteligentes, sistemas móviles, sistemas de comunicación de alta complejidad y sistemas de cómputo de alto rendimiento. Por el otro, la EST tiene experiencia en investigación básica en el terreno de la ciencia de la informática, asociado estrechamente al hecho de poseer la carrera de Ingeniería Informática como parte de su oferta académica.

En los laboratorios de la institución se dispone del material y los recursos humanos específicos para la investigación en tanto en el área de la seguridad, comunicaciones así como en las asociadas a las tecnologías de CD. También cabe mencionar que cuenta con expertos informáticos, matemáticos y criptógrafos que dan cuerpo muchas veces a las investigaciones de los trabajos finales de carrera que se realizan en los posgrados que brinda la universidad.

Tanto el CIDESO como los laboratorios de la EST, a través del dictado de materias de grado en Ingeniería Informática, aportan recursos humanos a la misma universidad. Es así que investigadores de los laboratorios dan cátedras en la EST y, de manera análoga, alumnos de la escuela aportan sus análisis a los laboratorios a través de trabajos prácticos de laboratorio, prácticas profesionales supervisadas o tesis y tesinas de grado y posgrado.

En particular, el prototipo experimental del cual nace la idea de CODEC fue implementado por los alumnos de tercer año de Ingeniería Informática y Electrónica como

trabajo práctico de laboratorio de la materia Lenguajes de Programación I, y tutorado por investigadores del CIDESO y el CriptoLab, que a la vez son docentes de dicha cátedra.

Para el próximo paso, se pretende continuar con esta interacción fluida entre los centros de investigación y el alumnado, formado profesionales con conocimientos de campo en el terreno de la computación de alto rendimiento y un conocimiento acabado sobre temas criptográficos.

Además, al expandirse el sistema para ser aplicado en cualquier problema que requiera altos niveles de cómputo, se pretende incorporar alumnos y docentes de otras cátedras, de cualquiera de las ingenierías que se dictan en la Facultad.

Por otro lado, se continuará con la formación de profesionales en investigación utilizando estas nuevas tecnologías en los dos laboratorios participantes a través de la utilización de la GDC.

Así, pues, se formarán recursos humanos de todos los niveles, grado, posgrado o investigadores activos, incorporando más alumnos a los laboratorios y, potencialmente, becarios que se dediquen de modo formal (no sólo académico) a la profundización de los modelos propuestos.

## 5- BIBLIOGRAFÍA

1. **Cockbun, A.** *Agile Software Development*. s.l. : Addison-Wesley, 2001.

2. **Bianchi, Oscar Martín y Repetto, Alejandro.** *Sistema de Computacion Distribuida para Seguridad Informatica*. Posadas, Misiones : Workshop de Investigadores en Ciencias de la Computación 2012, 2012.

3. **Bianchi, Oscar Martín y Bönke, Matías Miguel.** *Computación Distribuida para Experimentación Científica*. Mar del Plata, Buenos Aires, Argentina : RIIPRO Joven - 3er CIIP, 2012.

4. **Abrahamsson, P., y otros.** *Agile software development methods Review and analysis*. s.l. : VTT Publications, 2002.

5. **Canós, José H., Letelier, Patricio y Penadés, M. Carmen.** *Métodologías Ágiles*

*en el Desarrollo de Software*. s.l. : Universidad Politécnica de Valencia, 2010.

6. **K., Schwaber, M., Beedle y R.C, Martin.** *Agile Software Development with SCRUM*. s.l. : Prentice Hall, 2001.

7. **Ahmar, Abbas.** *Grid Computing Technology: An Overview*. s.l. : Abbas, 2004.

8. *Grid Desktop Computing for Constructive Battlefield Simulation*. **Repetto, Alejandro J. M.** San Salvador de Jujuy : Red UNSI, 2009. XV Congreso Argentino de Ciencias de la Computación (CACIC 2009). 978-897-24068-4-1.

9. *Hybrid Architecture for Constructive Interactive Simulation: Evaluation and Outcomes*. **Repetto, Alejandro J. M.** Orlando, FL : s.n., 2010. /ITSEC'10, Interservice/Industry Training, Simulation and Education Conference.

10. **Berstis, Viktors.** *Fundamentals of Grid Computing*. s.l. : Redbooks Paper, IBM Corp, 2002.

11. **Constantinescu-Fulop, Nicolae-Zoran.** *A Desktop Grid Computing Approach for Scientific Computing*. s.l. : Department of Computer and Information Science Faculty of Information Technology, Norwegian University of Science and Technology, 2008.

12. **Computerized Bussiness Solutions.** *Centralized vs Distributed Computing: White Paper*. 2007.

13. **Opiyo, Elisha T. O., y otros.** *Computing Research Challenges and Opportunities with Grid Computing*.

14. **Neves, Rodrigo, y otros.** *Parallel and Distributed Computing BOINC Grid Implementation*.

## Reuso Orientado a Servicios

Andrés Flores, Alejandra Cechich, Martín Garriga, Marcelo Moyano, Israel Cors  
Grupo de Investigación en Ingeniería de Software del Comahue (GIISCO)  
<http://giisco.uncoma.edu.ar>  
Departamento de Ingeniería de Sistemas – Facultad de Informática  
Universidad Nacional del Comahue  
Buenos Aires 1400, (8300) Neuquén  
Contacto: [andres.flores, alejandra.cechich]@fai.uncoma.edu.ar

### Resumen

El reuso de artefactos software brinda oportunidades para proveedores y clientes, tanto para acelerar el proceso de desarrollo de software como para establecer oferta de productos reusables. El paradigma de Computación Orientada a Servicios (SOC), promueve el desarrollo de aplicaciones distribuidas en ambientes heterogéneos, que son construidas ensamblando o componiendo servicios reusables, que se publican a través de una red y se acceden mediante protocolos específicos. SOC ha sido ampliamente adoptado bajo su implementación con la tecnología de Servicios Web, que provee flexibilidad de ejecución remota que oculta las plataformas específicas de ejecución y permite descentralizar los procesos de negocios. SOC requiere la publicación de servicios en un registro (UDDI de acuerdo a Servicios Web), los cuales luego son identificados y evaluados para una aplicación en desarrollo. Sin embargo, aún este proceso necesita métodos exhaustivos y eficientes, tanto para identificación como para selección de servicios, en el cual se puede considerar la aplicación de técnicas de Pruebas de Software y el uso de dos conceptos actuales: Orquestación y Coreografía de servicios.

Palabras Clave: Ingeniería de Software basada en Reuso – Software Orientado a

Servicios – Servicios Web – Calidad de Software – Verificación y Validación.

### Contexto

La línea presentada se inserta en el contexto de los siguientes proyectos y acuerdos de cooperación:

- UNCo-FAI: “Reuso Orientado a Servicios”, sub-proyecto del Programa “Desarrollo Orientado a Reuso”. En proceso de acreditación por la Universidad Nacional del Comahue.
- PAE-PICT-2312: “Métodos y Herramientas para Sistemas masivamente Distribuidos”.
  - Investigaciones conjuntas con ISISTAN-UNICEN, Tandil.
- Acuerdo de Cooperación con el Grupo Alarcos, Escuela Superior de Informática, Universidad de Castilla-La Mancha, España.

### Introducción

Actualmente la industria de software observa cómo mediante el reuso de software se puede acelerar el proceso de desarrollo de software basándose en artefactos software que ya han sido probados en diferentes contextos de aplicación, con lo cual el proceso en si mismo resulta confiable. El concepto que se establece se denomina tercerización, por medio del cual se establecen contratos

para adquisición/provisión de artefactos software reusables y se entablan relaciones comerciales entre vendedores y clientes. Por lo tanto, desde el punto de vista de un cliente implica la posibilidad de acelerar un desarrollo de software para reducir el lanzamiento al mercado, y desde el punto de vista de un proveedor implica la posibilidad de observar sus productos con una perspectiva nueva que los coloque dentro del mercado de artefactos reusables.

Un paradigma que promueve altamente el reuso de software se denomina Computación Orientada a Servicios (SOC), donde la funcionalidad a ser reusada adopta la forma de servicios, o unidades lógicas que presentan entornos heterogéneos de ejecución y pueden ser ensambladas para formar otras unidades lógicas de mayor nivel de abstracción que resuelvan (directamente o en parte) los procesos de negocios para un contexto de aplicación [SH05,PTDL07]. El paradigma SOC encontró una plataforma potencial de aprovechamiento mediante la Web, desde donde se desarrolló la tecnología de Servicios Web [NSS03, Wetal05], cuya base es el esquema estándar XML y el Lenguaje de Descripción de Servicios Web (WSDL), que facilitan ampliamente el desarrollo y mantenimiento de especificaciones formales de servicios. Así el paradigma SOC bajo la implementación con Servicios Web ha logrado su amplia adopción en la industria, principalmente bajo la flexibilidad de ejecución remota que permite a las compañías descentralizar aún más sus procesos de negocios y la ventaja de que las plataformas específicas de ejecución se encuentran ocultas, por lo cual no se requiere de inversiones adicionales en tecnología (incluyendo costos y esfuerzo de aprendizaje) al adquirir funcionalidad de terceras partes. El beneficio que la tecnología de

Servicios Web provee al paradigma SOC se ha denominado “relación sin responsabilidad”, donde una aplicación cliente no requiere asumir cómo se ha implementado el servicio con el que se comunica. Sin embargo, los proveedores de servicios tienen la responsabilidad de evaluar la calidad de los productos ofrecidos como servicios y los consumidores de servicios a su vez deben ser capaces de identificar tal calidad que influirá sobre las aplicaciones en desarrollo. Para ello, las estrategias de Pruebas de Software deben ajustarse a este contexto específico donde los servicios están acordados como cajas negras que sólo permite evaluar el comportamiento y cualidades observables externamente [BDN10, Z08].

El funcionamiento concreto del paradigma SOC se basa en la Arquitectura orientada a Servicios (SOA) [SH05] que se encuentra compuesta por tres actores principales: un proveedor, un consumidor y un registro de servicios; donde el proveedor desarrolla y publica servicios en el registro, para que luego el consumidor busque servicios y establezca una comunicación con el proveedor. Sin embargo, la búsqueda de servicios publicados en un registro UDDI (según la tecnología de servicios Web) [OASIS04], en general requiere invertir un esfuerzo considerable para distinguir servicios candidatos que satisfagan los requerimientos de la aplicación cliente [NSS03, Wetal05]. En particular, cuando varios candidatos ofrecen funcionalidades similares se requieren métodos eficientes de selección de servicios que discriminen tanto aspectos funcionales como no-funcionales, considerando además las interacciones válidas para un servicio candidato en función de los procesos de negocio que implementará la aplicación cliente. En particular el ensamblaje de servicios considerando procesos de



negocio e interoperabilidad de servicios plantea el uso de dos conceptos de reciente investigación: Orquestación y Coreografía de servicios [P03, Wetal05]: el primero relacionado a una aplicación particular que describe un proceso de negocios específico, y el segundo relacionado a las interacciones válidas que pueden ocurrir entre distintos servicios predestinados a intervenir en una colaboración. En este contexto se cuenta actualmente con diversos lenguajes de descripción y frameworks de ejecución, tales como Microsoft XLANG (para BizTalk Server) [Micro10], IBM WSFL (Web Services Flow Language) [Snell10], de lo cual surgió el estándar BPEL4WS [OASIS07], y otro estándar WSCDL (Web Services Choreography Description Language) [W3C05] promovido por Sun, SAP, BEA e Intalio.

## Líneas de Investigación y Desarrollo

El perfil de esta línea puede definirse en base a las actividades de investigación y transferencia, a las que da soporte el grupo GIISCo. Los temas específicos consideran los desafíos diferentes asociados al crecimiento de la Tecnología de la Información y las Comunicaciones. Actualmente, abordamos los siguientes aspectos:

- Evaluación y selección de servicios.
- Composición y orquestación de servicios.
- Testing de servicios.
- Herramientas para evaluación, selección, composición y testing de servicios.
- Definición de aplicaciones en dominios específicos.

## Resultados y Objetivos

En [Cetal12] hemos enumerado una serie de contribuciones anteriores. Durante el año 2012, hemos profundizado la investigación en aspectos de *evaluación y selección de servicios*, generando métodos y herramientas enfocados en compatibilidad de interfaces y de comportamiento en base a testing. Este avance se ha efectuado en colaboración con investigadores del Grupo Alarcos [FP12] y de ISISTAN, UNICEN [GFCZa12, GFCZb12, GFCZc12, GFCZd12].

Las líneas de investigación convergen en el tratamiento del desarrollo de software basado en el reuso de servicios desde la perspectiva de las aplicaciones orientadas a servicios. Una aplicación orientada a servicios implica una solución de negocio que consume servicios de uno o más proveedores y los integra en un proceso de negocio. Además puede verse como una aplicación basada en componentes que integra dos tipos de componentes: internos localmente empotrados en la aplicación, y externos estática o dinámicamente enlazados a algún servicio. No solamente se enfocará en el reuso de servicios individuales, sino también en la composición de servicios como forma de tercerizar una funcionalidad. Se adoptará la visión de proceso de negocio para la definición de comportamiento, donde se aplicará testing de servicios para una evaluación dinámica. Se complementará el modelo de selección y composición de servicios mediante las últimas plataformas y avances tecnológicos incluyendo semántica y estandarización. Se prevee la aplicación de estos modelos y las herramientas de soporte a dominios específicos, con particular énfasis en aquellos que requieran rigurosidad como aporte de validación efectiva. La visión

de esta línea de investigación se resume en:

*“Definir técnicas y herramientas para la mejora del desarrollo de software, en función del reuso de servicios web. La definición de modelos de identificación y selección de servicios, y la posibilidad de composición de servicios”.*

## Formación de Recursos Humanos

Este proyecto se compone de 9 investigadores, entre los que se cuentan docentes y alumnos del Grupo GIISCo de UNComa y asesores externos. Algunos de los docentes-investigadores se encuentran realizando carreras de postgrado. Se cuenta actualmente con 2 doctores (1 investigador asistente CONICET), 1 doctorando (becario CONICET) y 1 maestrando entre los miembros del proyecto. Dirección de Tesis de Grado durante 2012: 3 tesis (2 finalizadas), y se incorporaron 3 alumnos becarios PNB-TICs.

## Referencias

- [BDN10] Baresi, L.; Di Nitto, E. (2010). *Test and Analysis of Web Services*. Springer.
- [Cetal12] Cechich, A.; Buccella, A.; Flores, A.; Aranda, G.; Martínez Carod, N.; Luzuriaga, J.; Martínez, R.; Moyano, M.; Mazalú, R.; Martín, A.; Garriga, M. (2012). *Desarrollo basado en Reuso*. WICC'12, XIV Workshop de Investigadores en Ciencias de la Computacion, Misiones, Argentina.
- [FPo12] Flores, A.; Polo, M. (2012). *Testing-based Process for Component Substitutability*. Software Testing, Verification and Reliability (STVR), 22(8):529–561, Wiley InterScience, John Wiley and Sons, Ltd.
- [GFCZa12] Garriga, M.; Flores, A.; Cechich, A.; Zunino, A. (2012). *Assessment Scheme-based Service Selection for SOC-based Applications*. In Book: “Computer Science & Technology Series – XVII Argentine Congress of Computer Science, Selected Papers”. Editorial EDULP, pp. 185-194.
- [GFCZb12] Garriga, M.; Flores, A.; Cechich, A.; Zunino, A. (2012). *Practical Assessment Scheme to Service Selection for SOC-based Applications*. EJS Special Issue. SADIO Electronic Journal of Informatics and Operations Research, 11(1): 16-30. Issue dedicated to ASSE'11 (Argentine Symposium on Software Engineering) durante JAIIO'11.
- [GFCZc12] Garriga, M.; Flores, A.; Cechich, A.; Zunino, A. (2012). *Behavior Assessment based Selection Method for Service Oriented Applications Integrability*. ASSE'12, 13th Argentine Symposium on Software Engineering, durante 41 JAIIO, SADIO. pp. 339-353. Córdoba, Argentina.
- [GFCZd12] Garriga, M.; Flores, A.; Cechich, A.; Zunino, A. (2012). *WirySOC: Service Selection Method based on a Practical Interface Assessment Scheme*. Software: Practice and Experience (SPE), pp. 31. Wiley InterScience, John Wiley and Sons, Ltd. [en revisión].
- [Micro10] Microsoft Corporation (2010). *XLANG/s Language*. MSDN Library. [http://msdn.microsoft.com/en-us/library/aa577463\(BTS.70\).aspx](http://msdn.microsoft.com/en-us/library/aa577463(BTS.70).aspx)
- [NSS03] Nagappan, R.; Skoczylas, R.; Sriganesh, R. (2003). *Developing Java™ Web Services: Architecting*

- and Developing Secure Web Services Using Java*. Wiley Publishing Inc.
- [OASIS04] OASIS Consortium (2004). *UDDI Version 3.0.2*. UDDI Spec Technical Committee Draft, October.
- [OASIS07] OASIS Standard (2007). *Web Services Business Process Execution Language Version 2.0*. <http://docs.oasis-open.org/wsbpel/2.0/OS/wsbpel-v2.0-OS.html>
- [P03] Peltz, C. (2003). *Web Services Orchestration and Choreography*. IEEE Computer, 36(10):46–52.
- [PTDL07] Papazoglou, M.; Traverso, P.; Dustdar, S.; Leymann, F. (2007). *Service-Oriented Computing: State of the Art and Research Challenges*. IEEE Computer 40(11): 38-45.
- [SH05] Singh M.; Huhns, M. (2005). *Service-oriented computing: Key concepts and principles*. IEEE Internet Computing, 9(1):75–81.
- [Snell10] Snell, J. (2010) *The Web services insider, Part 4: Introducing the Web Services Flow Language*. IBM Emerging Technologies. <http://www.ibm.com/developerworks/library/ws-ref4/index.html>
- [W3C05] W3C Candidate Recommendation (2005) *Web Services Choreography Description Language Version 1.0*. <http://www.w3.org/TR/ws-cdl-10/>
- [Wetal05] Weerawarana, S.; Curbera, F.; Leymann, F.; Storey, T.; Ferguson, D. (2005). *Web Services Platform Architecture: SOAP, WSDL, WS-Policy, WS-Addressing, WS-BPEL, WS-Reliable Messaging, and More*. Prentice Hall PTR.
- [Z08] Zhou, X. (2008). *Testing and Verifying Web Services. From the Researcher's Perspective*. VDM Verlag.

## Interoperabilidad de Modelos de Simulación

Mónica del Carmen Gil, Germán Montejano, Mario Berón

Facultad de Ciencias Físico-Matemáticas y Naturales - Universidad Nacional de San Luis

Ejército de los Andes 950 - San Luis - Argentina

sistemas97@hotmail.com, [gmonte@unsl.edu.ar](mailto:gmonte@unsl.edu.ar), [mberon@unsl.edu.ar](mailto:mberon@unsl.edu.ar)

### Resumen

En la actualidad los diseñadores de sistemas y sus usuarios tienen dificultad a la hora de querer lograr la interconexión entre diferentes sistemas, a menudo por la falta de documentación o las deficiencias en la misma.

Para salvaguardar este inconveniente se necesita de un equipo de investigadores e instituciones comprometidas con el desarrollo de programas y estándares de datos, los cuales faciliten un entendimiento funcional de cómo operan los sistemas.

El uso de técnicas de simulación permite el diseño, la evaluación de alternativas, el pronóstico, la experimentación de estrategias, el contraste de teorías y el apoyo a la toma de decisiones, facilitando abordar la complejidad que presentan los sistemas.

La construcción de simuladores que sean **interoperables** entre sí, implica el manejo de diversos aspectos complejos tales como el progreso del tiempo, los modelos matemáticos y sistemas bien definidos y documentados a **nivel conceptual**, como así también, el compromiso de las distintas organizaciones responsables de brindar la información requerida. El objetivo final es proveer una infraestructura que permita desarrollar sistemas con los atributos de flexibilidad, extensibilidad, mantenibilidad y reusabilidad.

### Contexto

El presente trabajo se enmarca en el Proyecto de Investigación: Ingeniería de Software: Aspectos de alta sensibilidad en el ejercicio de la profesión del Ingeniero de Software” de la Facultad de Ciencias

Físico-Matemáticas y Naturales, Universidad Nacional de San Luis.

### Introducción

La simulación de sistemas, y en específico la simulación basada en modelos computacionales, supone el estudio de conceptos sobre las ciencias sociales, matemáticas y de la computación, y busca la creación de modelos simplificados de la realidad estudiada. La simulación de sistemas implica la construcción de modelos. El objetivo es averiguar que pasaría en el sistema si acontecieran determinadas hipótesis. Para ello se utilizan los modelos de simulación que son un importante medio para aumentar la eficiencia de la investigación, ya que estos auxilian a los investigadores en la asimilación del conocimiento adquirido mediante la experimentación.

La ingeniería de software por su parte se fundamenta en una aproximación denominada ciclo de vida del software el cual es un paradigma universalmente aceptado y que se compone de las fases genéricas de: conceptualización o construcción, transición, producción implementación y mantenimiento.

Un modelo es un subconjunto de la realidad. Cuando se construye el modelo, parte del mundo real y sus relaciones quedan excluidas.

A pesar de que los modelos de simulación han sido usados con éxito para entender diversos procesos sociales así como para asistir la formulación y evaluación de políticas en diferentes sistemas, existen pocos esfuerzos para formalizar, sistematizar y comunicar métodos para su descripción en la etapa conceptual.

Las investigaciones realizadas hasta el momento



[1,2,3,4], permiten comprobar que no existen metodologías para unificar criterios de especificaciones en las primeras etapas de diseño de cualquier sistema. Esta peculiaridad conduce a problemas de interoperabilidad al no modelar las relaciones que son necesarias para asegurarla. La mayoría de los modelos no están bien especificados y documentados desde un nivel inicial.

Durante el Panel de discusión SCS-SISO (*Simulation Interoperability Standards Organization*) [5] / SCS (*Society for Modeling and Simulation International*) sobre las prioridades para estándares de modelos y simulación, Seigler explicitó en su presentación que para asegurar la interoperabilidad entre sistemas la estandarización debe ser alcanzada en el nivel de modelado, por ej: el nivel de estandarización debe ser más alto que los estándares del nivel de programación aplicados actualmente. Para una “significativa interoperabilidad”, compartir los datos estandarizados vía protocolo estandarizado, tales como el DIS protocolo (*Distributed Interactive Simulation*) o el HLA (*High-Level Architecture*), son necesarios pero no suficientes [6].

La coordinación de los modelos conceptuales subyacentes, la armonización de las ideas de operación simuladas es el **quid de la cuestión** para crear soluciones interoperables. En lugar de estandarizar solamente los requerimientos de intercambio de información se debe coordinar la cadena causa efecto en el modelo subyacente.

La única instancia del proceso de desarrollo del modelo y simulación que lleva a asegurar esto es en el modelo conceptual. El modelo conceptual describe qué parte del mundo real es modelado y bajo qué restricciones, y algunas veces cuál es más importante y qué partes no serán modeladas.

Si bien, la interpretación no ambigua del significado de los datos intercambiados entre dos sistemas es necesaria; como en los modelos NMI (*Reference Model for Interoperability*) y LISI (*Levels of Information Systems Interoperability*) [7,8]; esto no es suficiente.

El hecho de establecer estándares de metadatos permite un mayor uso abierto de los datos dentro de los sistemas, aunque los datos no hayan sido estandarizados en si mismos, permiten la interpretación de los datos en un contexto dado.

Para lograr una interoperabilidad significativa entre los sistemas de simulación y dar soporte a las distintas etapas de una forma integrada, consistente y flexible; se aplicarán métodos de ingeniería basándose en el estándar LCIM (“*Levels of Conceptual Interoperability Model*”) [9], hasta poder cumplir principalmente con lo requerido en el Nivel Conceptual para aplicaciones de modelos y simulación.

### **Línea de Investigación**

Los esfuerzos científicos, para cubrir la demanda de los usuarios sin el continuo deterioro de los datos para aplicaciones específicas, requiere de un método que reconozca la complejidad del mundo real de naturaleza físico-químico y biológico como así también de consideraciones de naturaleza socio-económica, culturales y políticas.

Como consecuencia de los avances en tecnología de computación y de la ciencia de la informática, en la actualidad se dispone de las herramientas de apoyo para la integración del conocimiento adquirido a nivel disciplinario. Esta característica permite que este entendimiento pueda ser usado para analizar la complejidad del comportamiento de los sistemas. Es relevante destacar que en estas herramientas se incluyen los modelos de simulación, la definición y adopción de estándares y las prácticas adecuadas para tal fin.

Además de los estándares, existe el mismo problema en cuanto a arquitecturas: el uso de diferentes metodologías de diseño, sistemas marco, componentes, etc. dificulta la interconexión entre diferentes sistemas, a menudo por la falta de documentación o las deficiencias en la misma. La esencia del problema es que no se trabaja a conciencia cuando se definen los sistemas. A la hora de diseñarlo se pasa por alto las

especificaciones a un nivel conceptual, muchas veces por premuras de tiempo.

El desafío no es el intercambio de datos entre sistemas, ya que la arista técnica está suficientemente tratada por los estándares de interoperabilidad [10]. El problema es que los conceptos de los modelos (subyacentes) necesitan también estar alineados.

Con esto se quiere decir que hacer las cosas bien desde un principio, como definir los datos y dejarlos documentados, ahorra tiempo y esfuerzos en futuras modificaciones o usos de los mismos.

### ***Soluciones posibles basadas en LCIM***

Los modelos de simulación son un importante medio para aumentar la eficiencia de la investigación y toma de decisiones. Esto se debe a que: i) Pueden auxiliar a los investigadores en la asimilación del conocimiento adquirido mediante la experimentación, ii) Proporcionan un marco de referencia para aportes de carácter multidisciplinario, iii) Promueven el método de sistemas para la solución de problemas y iv) Facilitan una organización sistemática del conocimiento existente.

Teniendo en cuenta lo mencionado en el párrafo precedente, se propone usar una estructura estandarizada de elementos que envuelvan el contenido de los diferentes modelos, de manera que sea integrable en una arquitectura orientada a servicios, o en un servicio en particular dentro de esa estructura. Esta característica propenderá a mejorar la interoperabilidad y la reusabilidad. Las ventajas que se derivan de tal enfoque son que los componentes del sistema y el contenido puedan redistribuirse, ser ofrecidos y recibidos por diferentes actores y partes.

Así, cuando se vayan sumando modelos y soluciones interoperables, se podrán reutilizar las especificaciones semánticas anteriores ampliándolas sólo cuando por particularidades específicas se requiera. Se evitará de esta manera la

duplicación de esfuerzos y se ampliará el alcance de la interoperabilidad.

Para lograr la integración y el trabajo coordinado de los sistemas de información se deben solucionar problemas: i) Semánticos, como las incongruencias de estructuras de información incomprensibles de una organización a otra; ii) Organizacionales, causados por procesos administrativos descoordinados que se repiten innecesariamente; iii) Técnicos, causados por programas informáticos incompatibles entre sí y iv) De gobernanza, debidos a la falta de normas y de la institucionalidad necesarias para lograr acuerdos de intercambio de información entre las organizaciones que terminen siendo prácticas homogéneas y estándares aceptados por todas ellas.

El mejoramiento de las condiciones de intercambio de información puede buscarse desde ahora, en cada organización, para evitar que se sigan desarrollando soluciones informáticas aisladas ignorando las lecciones del pasado.

Para el logro de los objetivos mencionados previamente es necesario contar con sistemas que sean interoperables, es decir, que compartan estándares técnicos, semánticos y organizacionales acordados y acatados por todos.

La solución propuesta consiste en la definición de una ontología que condense el vocabulario común y las relaciones existentes entre todos los modelos de simulación. Para llevar a cabo esta tarea se necesita: Adoptar un Modelo de Interoperabilidad a Nivel Conceptual el que da rigor ingenieril a las diferentes etapas por las que pasará la definición de la Ontología, evitando la elaboración de soluciones ad hoc. Luego de realizada la tarea mencionada previamente, se llevará a cabo una investigación profunda de los modelos de simulación. Esta tarea tiene como finalidad poder extraer un vocabulario común y las relaciones existentes entre los términos para poder, cumpliendo con las etapas del modelo de Interoperabilidad, dar soporte a las distintas etapas de una forma integrada, consistente y flexible.

### **Caso de estudio para la aplicación de las soluciones propuestas**

Los sistemas de información geográfica (SIG) se están convirtiendo en importantes sistemas de gestión de información geo-referenciada en un amplio dominio de aplicación, especialmente en las ciencias de la tierra, la atmósfera y el océano. Esta gestión comprende procesos de análisis, predicción, estudios socioeconómicos y ecológicos, entre otros, para lo que resulta importante la posibilidad de inter-operar entre ellos [11].

Hoy en día, el empleo de modelos hidrológicos es indispensable, ya que pueden ser utilizados para analizar diferentes aspectos del agua como por ejemplo: los escurrimientos, las inundaciones, la operación de embalses, así como los sistemas de abastecimiento, los usos del agua, su aprovechamiento y protección del agua subterránea [12].

Con la llegada de los Sistemas de Información Geográfica y su aplicación en la modelación hidrológica, el hidrólogo encontró una forma de automatizar y hacer más eficiente los análisis que llevaba a cabo.

Para ello surgieron distintos simuladores de flujos de agua como ser: el Hydrus (simula el flujo unidimensional, transporte de un sólo soluto y movimiento de calor, en un medio variablemente saturado), Feflow (modelación avanzada tridimensional de elemento finito para flujo de aguas subterráneas, calor y transporte de contaminantes), Visual Help (estándar internacional para la modelación de hidrología de rellenos sanitarios y para la estimación de las tasas de recarga de aguas subterráneas), Visual Modflow (estándar de la industria en modelación tridimensional de flujo de aguas subterráneas, transporte de contaminantes y de calor) pero ninguno de estos simuladores pueden integrarse entre sí por no existir un Modelo Conceptual Compartido en cada uno de esos programas, es decir los resultados que brinda un simulador no pueden ser utilizados por otro [13].

Teniendo en cuenta lo importante que es el uso de estos simuladores para el dominio hidrológico se definirá e instanciará la ontología para la interoperabilidad de modelos y simulación con los conceptos del dominio hidrológico.

Con el propósito de mostrar la utilidad de las soluciones interoperables propuestas por esta línea de investigación, las mismas se aplicarán al: “estudio hidrogeológico de la llanura norte de la provincia de San Luis” para consolidar la hidrogeología local y regional y los datos hidrogeológicos disponibles, en un conjunto de hipótesis y conceptos que puedan ser evaluados cuantitativamente.

Los pasos para llevar a cabo la aplicación mencionada en el párrafo precedente son: i) Estudiar el dominio, ii) Definir el modelo conceptual para especificar luego su ontología.

Los conceptos se describen explícitamente para entender su significado, mediante acuerdos ontológicos. Con ello, un usuario que desee reutilizar una ontología desarrollada por otros, puede conseguir la información de todos los conceptos que soporta, su taxonomía y los axiomas.

La ontología se construirá a partir de los conceptos que se quieren relacionar, particularmente se enfocará al análisis de las relaciones entre los ríos y las curvas de nivel.

Una vez definida la ontología se la editará y se le realizarán chequeos de consistencia utilizando el PROTÉGÉ [14].

PROTÉGÉ es un software libre, un editor de ontología y un marco de trabajo de base de conocimiento. Utiliza el OWL (Ontology Web Language) que tiene como objetivo facilitar un modelo de marcado construido sobre RDF y codificado en XML.

### **Formación de Recursos Humanos**

En cuanto a la formación de recursos humanos en esta línea de investigación se espera:

- I) La finalización de tesis de maestría de dos maestrandos que actualmente están trabajando en el tema, y una tesis doctoral de un investigador del proyecto;
- II) Incrementar la interrelación con la Universidad Federal de Minas Gerais, institución con quien se mantiene un convenio de bi-titulación y
- III) Posibilitar la generación de nuevas tesis de doctorado, maestría y tesinas de licenciatura.

## Bibliografía

- [1] F. Javier Sanchez San Román, “*Introducción a Visual Modflow*”, Dpto. Geología Universidad de Salamanca(España).
- [2] Enric Peig Olivé; “*Interoperabilidad de Metadatos en Sistemas Distribuidos*”; Universitat Pompeu Fabra, Octubre 2003.
- [3] Mark Kasunic, William Anderson; “*Measuring Systems Interoperability: Challenges and Opportunities*” Software Engineering Measurement and Analysis Initiative, CMU/SEI-2004-TN-003, Abril 2004
- [4] José Aurelio Martínez García; “*Simulación de link 16 sobre HLA*”; Universidad Europea de Madrid, Escuela superior politécnica -Área de telecomunicación, Curso 2006-2007
- [5] SISO/SCS Panel Discussion: “*Priorities for M&S Standards*,” Spring Simulation Interoperability Workshop, Orlando, Florida, March 2003.
- [6] IEEE Standard Group 1516: “*High Level Architecture*” (Revision 2000), IEEE CS Press.
- [7] C4ISR Architecture Working Group, “*Levels of Information Systems Interoperability (LISI)*”. U.S. Department of Defense, March 1998.
- [8] Lisa L. Brownsword, David J. Carney, David Fisher, Grace Lewis, Craig Meyers, Edwin J. Morris, Patrick R. H. Place, James Smith, Lutz Wrage; “*Current Perspectives on Interoperability*”, TECHNICAL REPORT, CMU/SEI-2004-TR-009, ESC-TR-2004-009, March 2004.
- [9] Andreas Tolk, Ph.D., James A. Muguirra, “*The Levels of Conceptual Interoperability Model*”, Virginia Modeling Analysis and Simulation Center Old Dominion University, Tolk/Muguirra 2003.
- [10] Moreno, H., S. Silveira-Netto y H. Sin (2007), “*Conceptualización de arquitectura de gobierno electrónico y plataforma de interoperabilidad para América Latina y el Caribe*”, serie Sociedad de la información (LC/W.140), Comisión Económica para América Latina y el Caribe (CEPAL), Santiago de Chile, julio.
- [11] Carlos Mario Zapata Jaramillo, Guillermo González Calderón, María Isabel Marín Morales, “*Un caso de estudio en interoperabilidad entre sistemas de información geográfica*”, Universidad de Medellín, 2009.
- [12] M. en C. Víctor Hugo Guerra Cobián, “*Análisis del efecto de discretización espacial en el modelado de cuencas hidrológicas utilizando el modelo distribuido CEQUEAU-ONU*”, Toluca México, Septiembre de 2007.
- [13] F. Javier Sanchez San Román, “*Introducción a Visual Modflow*”, Dpto. Geología Universidad de Salamanca (España).
- [14] Esmeralda Ramos y Haydemar Nuñez “*ONTOLOGÍAS: componentes, metodologías, lenguajes, herramientas y aplicaciones*”, Centro de Ingeniería de Software y Sistemas -ISYS- Laboratorio de Inteligencia Artificial -LIA- Caracas, Julio, 2007.



# Desarrollo de Modelos de Evaluación Usando Operadores de una Lógica Continua

Aristides Dasso\*, Ana Funes\*

\*Universidad Nacional de San Luis

Ejército de los Andes 950

San Luis

Argentina

{arisdas, afunes}@unsl.edu.ar

## Resumen

En la línea de investigación aquí presentada, nos ocupamos del desarrollo de modelos para la evaluación de sistemas usando el método LSP (Logic Score of Preference) [Duj08]. Actualmente, nos encontramos avocados al desarrollo de modelos para la evaluación de bienes inmobiliarios, particularmente –aunque no exclusivamente– modelos que puedan ser aplicados al cálculo del avalúo fiscal, aunque también pueda ser aplicado a otras áreas.

El empleo del método LSP permite expresar aspectos en la evaluación que otras técnicas meramente aditivas no permiten, ofreciéndonos la posibilidad de construir modelos que se ajusten con una mayor precisión a las necesidades del usuario, sea este un ente recaudador de impuestos fiscales, un agente inmobiliario o cualquier otro interesado en obtener una tasación de un bien inmueble.

**Palabras clave:** *Lógica Continua. Avalúo de Inmuebles. Métodos de Tasación. Logic Score of Preference. LSP.*

## Contexto

Este trabajo de investigación se encuentra enmarcado dentro de una de las líneas de investigación del Proyecto de Incentivos código 22/F222 “Aspectos de alta sensibilidad en el ejercicio de la profesión de Ingeniero de Software”, de la Universidad Nacional de San Luis. Dentro del contexto de desarrollo de

métodos y herramientas, esta investigación tiene como objetivo el concretar el desarrollo de herramientas de software para la construcción parametrizada de modelos de avalúo de inmuebles, ya sea para necesidades fiscales o de marketing, haciendo uso del método LSP. Este método es altamente preciso ya que se basa en el uso de operadores de una Lógica Continua.

## Introducción

El avalúo de inmuebles es una tarea que realizan distintos “stakeholders”, tanto privados como entes estatales. Estos stakeholders pueden querer valorar un inmueble por razones diversas; por ejemplo, un banco para garantía de un préstamo, un ente gubernamental para propósitos impositivos, o un agente inmobiliario por razones de compra o de venta.

En general, en todos los casos, los modelos de avalúo fiscal, cuando existen en forma de sistemas informáticos, suelen estar basados en simples sumatorias pesadas de distintas características del bien a ser valuado.

Por otro lado, el método LSP, un método que se basa en el empleo de una lógica continua, que permite la creación de funciones complejas de evaluación y su aplicación en la evaluación y comparación de sistemas de índole general, permite la creación de modelos precisos y fácilmente adaptables a las necesidades del usuario.

En el presente trabajo apuntamos a generar, en un solo valor, haciendo uso del método LSP,

un indicador que se obtiene al agregar las distintas características o atributos de un bien bajo evaluación.

Esta agregación tiene como resultado final un único valor, que es un índice entre 0 y 100 que sirve para determinar el valor final, ya sea del bien mismo o del impuesto que le corresponda de acuerdo a cómo dicho índice sea empleado.

En la sección siguiente ilustramos, por medio de un ejemplo sencillo, cómo el índice es obtenido y cómo el mismo puede ser aplicado en un entorno impositivo.

## Resultados y Objetivos

Como resultado de los objetivos planteados, se espera obtener un modelo LSP para el cálculo del avalúo fiscal.

Para lograr nuestro objetivo, en primer término es necesario identificar claramente los parámetros o características principales a tener en cuenta en el cálculo del avalúo para, luego, explotar las características de primer nivel en sub características y armar lo que en el método se da en llamar *árbol de preferencias*. Las hojas de dicho árbol reciben el nombre *variables de performance*.

Cabe hacer notar que cada una de las características propias de un bien (p.e. superficie, servicios, etc.) provee una perspectiva diferente en el espacio de valuación de dicho bien, brindando resultados diferentes de acuerdo con el modelo de evaluación (o avalúo, en nuestro caso). Es por eso que el modelo que proponemos nos permite contar con un resultado único (un índice) proveniente de un modelo comprensible que combine las múltiples características de un bien y permita al usuario calibrar dicho modelo de acuerdo a sus necesidades.

Una vez identificadas las características que intervienen en la valuación de un bien inmueble, nuestro objetivo será, entonces, integrar dichas características en un único índice, el cual puede ser empleado para el cálculo del correspondiente valor del bien, ya sea para calcular un impuesto o considerar el

valor de mercado.

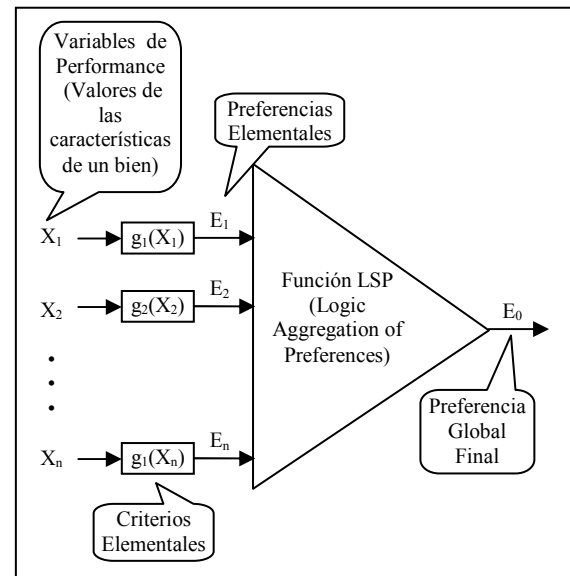


Figura 1. El proceso de evaluación de LSP.

Para construir el modelo de evaluación, que agrega las características, empleamos los operadores que provee la Lógica Continua, propuesta en el método Logic Score of Preferences (LSP), cuyo proceso se encuentra resumido en la Figura 1. En ella mostramos una visión global del método LSP con sus correspondientes partes.

Este método nos permite construir modelos de evaluación que consisten en *estructuras de agregación*, las que se basan en el uso combinado de distintas funciones GCD (Generalized Conjunction Disjunction) del método LSP. Los datos obtenidos de las distintas características del inmueble, las cuales, de acuerdo a LSP, serían las variables de performance del sistema, son mapeados por medio de funciones, llamadas *criterios elementales*, en *preferencias elementales*. Son estas preferencias elementales las que van siendo agregadas por medio de funciones GCD en estructuras de agregación y que nos permiten obtener un único valor final (*preferencia global final*  $E_0$  en la Figura 1).

Los criterios elementales son funciones que transforman un valor real proveniente de una variable de performance en un valor perteneciente al intervalo  $[0,100]$ . Estos valores resultantes son llamados preferencias elementales y representan el grado de cumplimiento con un requisito del sistema

bajo evaluación. Así, por ejemplo, si se trata de la superficie del inmueble esa superficie se transformará, con el correspondiente criterio elemental, en un valor del intervalo [0,100]. La relación entre la superficie y el intervalo será justamente propia de la elección del criterio elemental por los encargados de construir el modelo.

Estos criterios elementales nos permiten establecer, por ejemplo si así se desea, valores mínimos para una variable de performance.

Ilustraremos esto por medio de un ejemplo simple. En la Figura 2 puede verse un posible ejemplo de un árbol parcial de preferencias, de donde se obtendrán las variables de performance. Así obtendremos valores para la variable ‘Superficie’ del inmueble, y para la característica ‘Valor de Mercado’, el cual se conforma de dos variables: ‘Ventas anteriores’ de inmueble similares en la zona y ‘Costo’ (estimado) del mismo; la ‘Ubicación’, que puede tener en cuenta distintas zonas de una provincia; la ‘Accesibilidad’ a la propiedad que en este caso hemos considerado compuesta de dos variables: ‘Distancia a centros urbanos’ y ‘Tipos de Caminos’. Por supuesto que el árbol de preferencias es una elección de quién o quienes construyan el modelo.

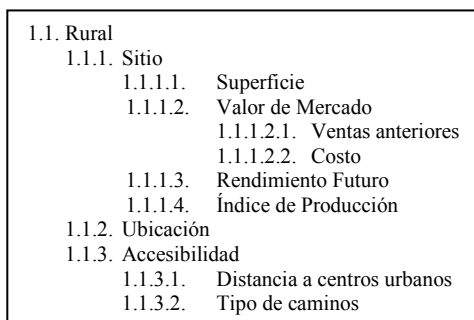


Figura 2. Algunas variables de performance para el caso de inmueble rural.

A partir de las variables de performance mostradas en el ejemplo de la Figura 2, como ya se dijo, será necesario definir los correspondientes criterios elementales, y luego construir la estructura de agregación.

Un ejemplo de la posible estructura de agregación es mostrada en la Figura 3.

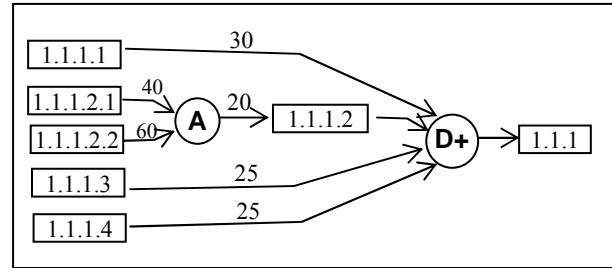


Figura 3. Estructura de agregación de una parte (1.1.1) del árbol de preferencias de la Figura 2.

Nótese que en la estructura de la Figura 3, el ítem 1.1.1.2.2 tiene mayor peso que el 1.1.1.2.1, esto es porque, para el ejemplo, se consideró que el costo estimado actual del inmueble es más importante que los precios de ventas anteriores del mismo bien o similares. El operador A (media aritmética) de la lógica continua devuelve la media ponderada (pesada) de los valores de 1.1.1.2.1 y 1.1.1.2.2. La función D+, que agrupa a los ítems que constituyen la característica 1.1.1, es una función fuertemente disyuntiva, que permite que cualquiera de los valores de los ítems que integran a 1.1.1 pueda faltar, es decir que sea valuados a cero, sin afectar el valor resultante de la agregación de los valores de los otros ítems.

La Tabla 1 ilustra un posible uso del modelo construido para el cálculo del valor del impuesto inmobiliario. En ella se ve que el avalúo, tanto para los bienes rurales como urbanos, tienen un máximo y un mínimo (a los que hemos denominado  $Ar_{mx}$ ,  $Ar_{min}$ ,  $Au_{mx}$ ,  $Au_{min}$ , respectivamente). Estos valores deben ser elegidos por las autoridades correspondientes. Entonces si el índice que se obtiene para un bien rural es, p.e., 100 le corresponderá un avalúo  $Ar_{mx}$ . Si p.e., está entre 0 y un cierto límite inferior  $l_l$ , le corresponderá un avalúo  $Ar_{min}$ , si tiene un índice entre  $l_l$  y 100 le corresponderá un avalúo proporcional entre  $Ar_{min}$  y  $Ar_{mx}$ .

Tabla 1. Matriz de uso impositivo Avalúo

	Máximo	Mínimo
Rural	$Ar_{mx}$	$Ar_{min}$
Urbano	$Au_{mx}$	$Au_{min}$

Otra alternativa de empleo del índice obtenido del modelo construido es asignar un valor monetario a la unidad o fracción de cada valor

entre 0 y 100; de esa manera el avalúo estará en directa relación con el ‘precio’ del índice que le corresponda al bien a evaluar.

## Líneas de Investigación y Desarrollo

Este trabajo es llevado a cabo dentro de la línea de “Métodos Formales y Prototipos Evolutivos” del proyecto de incentivos de la Universidad Nacional de San Luis, código 22/F222: “Aspectos de alta sensibilidad en el ejercicio de la profesión de Ingeniero de Software”. Cabe destacar que, en esta línea de trabajo de desarrollo de modelos de evaluación, en la cual venimos trabajando desde hace varios años, hemos desarrollado diversos modelos para evaluación de distintos tipos de sistemas, tales como servicios de gobierno electrónico, herramientas de modelado en UML, browsers, entre otros (ver [DF11], [CDF09], [DDF00], [DDF07], [DFP04], [DFPS01], [DPS03], [FDD00], [FDPS05], [MDU00]). Dichos modelos han sido creados aplicando el método LSP (Logic Score of Preferences) el cual hace uso operadores de una Lógica Continua ([Duj08], [Duj07], [Duj96], [Duj97], [DB97]).

El desarrollo y calibración de modelos para la evaluación de clasificadores es uno de los proyectos inmediatos en esta línea, al mismo tiempo que se continúa trabajando en la construcción de modelos de evaluación de calidad de software como otro de los objetivos en esta línea.

A partir de estos desarrollos se analiza la posibilidad de nuevas tesis de grado y posgrado.

## Formación de Recursos Humanos

La evaluación de sistemas, métodos y herramientas es una de las áreas en la cual venimos trabajando desde hace varios años y que ha producido varias publicaciones como ya se ha mencionado más arriba. Este trabajo continuo nos ha conducido, entre otros, a la

evaluación de sitios de gobierno electrónico, lo que ha dado como resultado tesis de maestría y de licenciatura, a la vez que otras se encuentran en preparación.

Los aspectos propios del trabajo aquí presentado son ambiciosos y se espera que las distintas tareas a desarrollar sirvan para la realización de tesis de posgrado así como de grado.

## Referencias

- [CDF09] M. Castro, A. Dasso, A. Funes. “Modelo de Evaluación para Sitios de Gobierno Electrónico”. 38 JAIIO/SIE 2009, Simposio de Informática en el Estado 2009, Mar del Plata, Argentina, August 26-28, 2009.
- [DB97] J. J. Dujmovic and A. Bayucan, “Evaluation and Comparison of Windowed environments”, Proceedings of the IASTED Interna Conference Software Engineering (SE'97), pp 102-105, 1997.
- [DDF00] N. Debnath, A. Dasso, A. Funes, G. Montejano, D. Riesco, R. Uzal, “The LSP Method Applied to Human Resources Evaluation and Selection”, Journal of Computer Science and Information Management, Publication of the Association of Management/International Association of Management, Volume 3, Number 2, 2000, ISBN 1525-4372, pp.1-12.
- [DDF07] Narayan Debnath, Aristides Dasso, Ana Funes, Roberto Uzal, José Paganini. “E-government Services Offerings Evaluation Using Continuous Logic”. 2007 ACS/IEEE International Conference on Computer Systems and Applications, AICCSA '2007, Amman, Jordan. Sponsored by IEEE Computer Society, Arab Computer Society, and Philadelphia University, Jordan. May 13-16, 2007
- [DF11] Aristides Dasso, Ana Funes. “A Model for E-voting Systems Evaluation”. 40 JAIIO/SIE 2011, August 29 to September 2, 2011. Córdoba, Argentina.



- [DFP04] A. Dasso, A. Funes, M. Peralta, C. Salgado, "User Oriented Evaluation Models for DBMSs", 33 Jaiio (ASIS 04), Córdoba, Argentina, 20-24 de Septiembre, 2004.
- [DFPS01] A. Dasso, A. Funes, M. Peralta, C. Salgado, "Una Herramienta para la Evaluación de Sistemas", Workshop de Investigadores en Ciencias de la Computación, WICC 2001, Universidad Nacional de San Luis, San Luis, Argentina, May 2001.
- [DPS03] N. Debnath, M. Peralta, C. Salgado, A. Funes, A. Dasso, D. Riesco, G. Montejano, R. Uzal, "Web Programming Language Evaluation using LSP", CAINE03 Proceedings, Las Vegas, USA, 11-13 de Noviembre, 2003. ISBN: 1-880843-49-8, pp 302-305.
- [Duj08] Dujmovic, J.J., "Characteristic forms of generalized conjunction/disjunction"; En Fuzzy Systems, 2008 (FUZZ-IEEE 2008). (IEEE World Congress on Computational Intelligence). 1-6 June 2008, pp. 1075 – 1080, ISSN: 1098-7584, E-ISBN: 978-1-4244-1819-0, Print ISBN: 978-1-4244-1818-3.
- [Duj07] Jozo J. Dujmovic, "Continuous Preference Logic for System Evaluation", IEEE Transactions on Fuzzy Systems, Vol. 15, N° 6, December 2007.
- [Duj96] J. J. Dujmovic, "A Method for Evaluation and Selection of Complex Hardware and Software Systems", The 22nd International Conference for the Resource Management and Performance Evaluation of Enterprise Computing Systems. CMG96 Proceedings, vol. 1, pp.368-378, 1996.
- [Duj97] J. J. Dujmovic, "Quantitative Evaluation of Software", Proceedings of the IASTED International Conference on Software Engineering, edited by M.H. Hamza, pp. 3-7, IASTED/Acta Press, 1997.
- [FDD00] A. Funes, A. Dasso, J. Dujmovic, G. Montejano, D. Riesco, R. Uzal, "Web Browsers Performance Analysis using LSP Method", Proceedings de la International Conference on Software Engineering Applied to Networking & Parallel/Distributed Computing (SNPD'00), Mayo, 2000, Reims, Francia. ISBN: 0-9700776-0-2, pp 551-558.
- [FDPS05] Ana Funes, Aristides Dasso, Carlos Salgado, Mario Peralta, "UML Tool Evaluation Requirements". Argentine Symposium on Information Systems ASIS 2005. Rosario, Argentina. September 29-30, 2005.
- [LU04] Z. Lu et al., Predicting Subcellular Localization of Proteins using Machine-Learned Classifiers, Bioinformatics, Volume 20, Issue 4, March 2004, pp. 547 - 556.
- [MDU00] G. Montejano, J.J. Dujmovic, R. Uzal, D. Riesco, A. Dasso, A. Funes, "A Prototype Tool for Decision Support based in the LSP Method", Proceedings de IASTED, Las Vegas, Nevada, USA, 6-9 de Noviembre, 2000. ISBN: 0-88986-306-7, pp 1-4.

# Los Sitios Web Gubernamentales como facilitadores del Gobierno Electrónico - Análisis de la evolución de los sitios web municipales del Conurbano Bonaerense

Daniel Giulianelli, Rocío Rodríguez, Pablo Vera,  
Artemisa Trigueros, Isabel Marko

GIDFSI (Grupo de Investigación, Desarrollo y Formación en Innovación de Software)  
Departamento de Ingeniería e Investigaciones Tecnológicas

Universidad Nacional de La Matanza

Florencio Varela 103, San Justo, Buenos Aires, Argentina

4480-8900 (interno 8751)

dgiulian@ing.unlam.edu.ar; rrodriguez@ing.unlam.edu.ar; pvera@ing.unlam.edu.ar;  
atrigueros@ing.unlam.edu.ar; imarko@ing.unlam.edu.ar

## Resumen

Este artículo presenta una línea de Investigación y Desarrollo en la cual se ha realizado un exhaustivo análisis de los sitios web municipales del conurbano bonaerense. Con la finalidad de analizar en qué grado estos sitios implementan aspectos que permitan conseguir los objetivos perseguidos por el gobierno electrónico planteados en el presente artículo por medio de 5 pilares fundamentales.

El grupo de trabajo ha analizado los 30 sitios web municipales del conurbano bonaerense en dos periodos de tiempo (2007-2008 y 2011-2012) con la finalidad de analizar el grado de evolución de los mismos.

En el presente artículo se mostrarán las generalidades de la línea de investigación y desarrollo, así como los resultados de dicha evolución.

**Palabras Clave:** Gobierno Electrónico, Web, e-Servicios, e-Transparencia, e-Democracia

## 1. Contexto

En la Universidad Nacional de La Matanza se llevan a cabo diversas líneas de Investigación y Desarrollo (I+D). Entre ellas la presente, dirigida a la calidad de los sitios web gubernamentales. Esta línea no sólo permite analizar a los sitios, sino también establecer lazos entre la Universidad y los

Municipios, permitiendo incrementar la calidad de los sitios gubernamentales en pos de brindar al usuario una experiencia satisfactoria al utilizar los mismos. Esta línea de investigación es financiada por la universidad a través del programa CyTMA 2.

## 2. Introducción

### 2.1. Sitios Web Gubernamentales

Es posible plantear la Gobernabilidad Electrónica como la intersección de dos fenómenos sociales contemporáneos: la comunicación global, generada a partir del desarrollo de las TIC (Tecnologías de la Información y la Comunicación) y una nueva visión de gobernabilidad, donde el ciudadano asume un rol activo, en vez del tradicional rol de receptor pasivo de las acciones de gobierno. Esta concepción de gobernabilidad ha ido perfeccionándose con el tiempo y la experiencia basada en la implementación en distintos países y comunidades. Uno de los conceptos que ha evolucionado es el de comunicación gobierno/ciudadano (G2C). Dicha comunicación puede ser pasiva, donde el papel del gobierno municipal es el de proveedor/emisor de noticias, informaciones y servicios; mientras que el papel del ciudadano es simplemente de receptor pasivo de lo que el gobierno decide entregarle, siendo esta concepción primitiva y antigua.

La ONTI (Oficina Nacional de Tecnologías de Información), define al Gobierno Electrónico como: “el uso de las Tecnologías de Información y Comunicación para redefinir la relación del gobierno con los ciudadanos, mejorar la gestión y los servicios, garantizar la transparencia y la participación y facilitar el acceso a la información pública, apoyando la integración y el desarrollo de los distintos sectores”. [ONT05]

La presente línea de I+D se centra en analizar como los sitios web gubernamentales facilitan los objetivos principales del Gobierno Electrónico.

## 2.2. Pilares del Gobierno Electrónico

Uno de los propósitos de los sitios web municipales es favorecer las comunicaciones y el nivel de servicios ofrecidos a los ciudadanos de la comunidad a la que pertenecen, esta investigación analiza la calidad de los mismos en base a la implementación de 5 pilares básicos (los tres primeros se encuentran basados en los campos de aplicación planteados por la UNESCO [UNE01]):

1. Administración electrónica: Se refiere a la mejora de los procesos gubernamentales internos, la gestión de los funcionarios del sector público y los procesos de ejecución e información.
2. Servicios electrónicos: Tienen como objetivo mejorar la provisión de información y ofrecer a los ciudadanos el acceso a los servicios públicos. También están incluidas las informaciones sobre eventos, espectáculos, transporte público, bolsa de trabajo, políticas de empleo, licitaciones, mapas, etc. Como ejemplos de servicios interactivos se pueden mencionar: solicitudes de documentos públicos, solicitudes de documentos legales y certificados, expedición de permisos y licencias, otorgamiento de turnos, pagos on line de impuestos, tasas y servicios.
3. Democracia electrónica: Implica una mayor y más activa participación ciudadana en el proceso de toma de decisiones gracias a las TIC. Como ejemplos se pueden mencionar:

encuestas, foros, chat, blogs, paneles, referendums, listas de correo, boletín por mail, contacto directo con autoridades, responsables de áreas y representantes, opiniones y sugerencias de los ciudadanos, libro de quejas, preguntas frecuentes y sus respuestas.

4. Transparencia electrónica: Recorre los conceptos de e-Servicios y e-Democracia en forma transversal. Puede verse como e-Servicios al proveer información de actos y decisiones de gobierno, ya que a partir de esa información los ciudadanos conocen las acciones de gobierno y sus motivaciones. Por otra parte, también puede entenderse como e-Democracia ya que el ciudadano, al informarse y tener los medios de comunicación adecuados, puede participar activamente en las decisiones y control de su gobierno.
5. Comunicación Activa y Pasiva: La Comunicación Pasiva puede pensarse como comunicación simplex, donde un actor, en este caso el gobierno municipal, siempre es el emisor y el otro actor, en este caso el ciudadano, siempre es el receptor. Mientras que la comunicación Activa puede ser pensada como comunicación full duplex, donde ambos actores pueden emitir y recibir comunicación en forma simultánea.

La aplicación de la Gobernabilidad Electrónica, provee una nueva forma de gobernar, al fomentar la participación ciudadana, propende necesariamente hacia la transparencia de la gestión pública y conlleva un nuevo estilo de dirección y de liderazgo, más amplio, fluido y cotidiano. La comunicación con los ciudadanos cambia radicalmente, ya que se borran las barreras de tiempo y distancia, permitiendo a cualquier habitante de la zona conocer, informarse, opinar, peticionar y, por qué no, criticar y controlar a sus representantes. Es a la vez un juego de conocer y recibir servicios e informaciones, antes tediosos y desgastantes, pero al mismo tiempo, se convierte en una responsabilidad para el ciudadano, ya que la Gobernabilidad Electrónica a través de las TIC le provee el medio para ejercer activamente su condición de ciudadano

responsable a través de un control efectivo de las acciones de gobierno.

También para el gobierno existen ventajas cuando la Gobernabilidad Electrónica se implementa en forma eficiente ya que éste puede proporcionar un mejor servicio en cuanto al tiempo, tramitaciones, consultas y pagos más eficaces. Los costos administrativos bajan y los ciudadanos pierden menos tiempo y sufren menos inconvenientes para realizar sus gestiones ante el gobierno.

A través del e-Governance se hace efectiva la transparencia en la gestión gubernamental, ya que es el medio ideal para comunicar, exponer, mostrar y controlar dicha gestión.

### 2.3. Muestra Considerada

Los gobiernos locales están cerca de los ciudadanos y constituyen para muchos la principal representación de gobierno. La relación de los ciudadanos y las autoridades locales tiende a ser una relación basada en la proximidad ya que los intereses en juego de ambas partes están claramente entrelazados con respecto a temas como:

- los servicios públicos
- el desarrollo urbano
- la planificación escolar
- los problemas del medio ambiente
- la política local

Es a nivel local que el impacto de las TICs en las relaciones entre gobiernos y ciudadanos puede ser más eficaz.

Se consideran a los 30 Municipios del Conurbano Bonaerense, son habitados por aproximadamente 10.000.000 de personas, constituyendo el 25 % de la población de la Argentina (INDEC Censo 2010 [IND10]) y donde, por sus características, incluyen toda la gama sociocultural y económica posible en nuestro país, constituyendo una magnífica muestra para el estudio en cuestión. Paralelamente, la tecnología se ha hecho más accesible para los ciudadanos, ej: mayor ancho de banda, netbooks, celulares, etc. [CNC10], [ITU10]. La Universidad Nacional de La Matanza está ubicada en el Conurbano

Bonaerense, es así como un amplio porcentaje de sus alumnos son ciudadanos de sus municipios, constituyendo este estudio, un aporte a la comunidad educativa en la cual está inmersa esta casa de altos estudios.

### 2.4. Metodología

Se realiza un primer estudio en los años 2007-2008, el cual se replica en el 2011-2012 para poder advertir la evolución de los sitios web municipales del conurbano bonaerense. Por cada uno de los sitios se analiza un total de 80 aspectos que influyen en 1 ó más de los pilares de Gobernabilidad Electrónica. El 75% de esos aspectos fueron considerados en base a recomendaciones de la ONTI [ONT05] y del W3C [W3C07]. El resto de publicaciones de colegas y elaboración propia del grupo de investigación.

A fin de resumir lo realizado no se consigan los 80 aspectos, pero pueden ser consultados online en la publicación [GIU08] efectuada en la primer etapa del estudio (2007-2008). Dichos aspectos fueron ponderados según el grado de significancia con una puntuación del 1 al 5 (ver tabla 1).

**Tabla 1.** Nivel de Relevancia y Puntaje Asociado

NIVEL DE RELEVANCIA	PUNTAJE
Indispensable	5
Muy Importante	4
Importante	3
Deseable	2
Optativo	1

La tabla 2, muestra la cantidad de aspectos por pilar y por relevancia, lo que permite visualizar los puntajes involucrados que permitirán obtener un valor total por pilar (resaltado en negrita sobre la última fila de la tabla). Ese valor total por pilar, es el valor ideal que obtendrá un sitio que cumpla con todos los aspectos analizados.

Cabe aclarar que tanto en la tabla 2 como en la 3, solo se ha indicado una letra de cada pilar siendo S: e-Servicios, D: e-Democracia, T: e-Transparencia y finalmente para la Comunicación Activa y Pasiva se ha asignado A y P respectivamente.



**Tabla 2.** Puntajes Ideales por Pilares

NIVEL		S		D		T		A		P	
		Cant.	Puntos	Cant.	Puntos	Cant.	Puntos	Cant.	Puntos	Cant.	Puntos
5	Indispensable	10	50	17	85	12	60	15	75	13	65
4	Muy Importante	1	4	1	4	1	4	1	4	0	0
3	Importante	21	63	11	33	8	24	23	69	10	30
2	Deseable	11	22	0	0	2	4	10	20	5	10
1	Optativo	9	9	0	0	1	1	8	8	2	2
<b>TOTALES</b>		52	<b>148</b>	29	<b>122</b>	24	<b>93</b>	57	<b>176</b>	30	<b>107</b>

En base a la metodología formulada antes del primer periodo de relevamiento, se analizan nuevamente cada uno de los sitios con una planilla de relevamiento en donde se indica por cada aspecto si es cumplido o no. En caso de ser cumplido se suma el valor asignado a dicho aspecto. El resultado será, para cada pilar, la sumatoria de todos los valores de los aspectos cumplidos. En este artículo se presentan la evolución analizada de los estudios realizados en ambos períodos.

### 3. Líneas de Investigación y Desarrollo

A partir de la presente línea de investigación se involucran los ejes temáticos:

- Modelos de Calidad para Gobierno Electrónico
- Calidad en Sitios Web
- Establecimiento de Métricas para evaluar aspectos web
- Establecimiento de Métodos para la Revisión y Evaluación de Métricas

### 4. Resultados

En la tabla 3 se muestran los pilares y por cada uno de ellos dos columnas la primera corresponde al primer período de relevamiento (2007-2008) y la segunda al último período (2011-2012). Cabe destacar que en el primer período había municipios que no tenían sitio web oficial, en algunos casos si existían sitios con comentarios o algo de información mínima construida por individuos particulares.

Se han considerado los sitios web oficiales de los municipios. Es por ello que hay municipios con celdas en blanco en la tabla 3, en la primera columna de cada uno de

los pilares, esto indica que no había sitio a analizar en esa época.

Analizando los valores obtenidos en ambos períodos (presentados en la tabla 3) pueden alcanzarse los siguientes resultados:

- El 60% de los sitios de la muestra considerada han evolucionado en la entrega de e-Servicios, mientras que el 40% ha empeorado su puntaje con respecto al primer período de análisis.
- En el período inicial (2007-2008) el 20% de los sitios no cumplía con ninguno de los aspectos asociados al pilar de e-Democracia, en cambio en el periodo 2011-2012 puede observarse que todos los sitios tienen una puntuación resultante, la más alta es de 103 sobre un total esperado de 122 puntos, mientras que la más baja es de 64 representando al 53% del puntaje ideal. Esto muestra una importante evolución en los sitios municipales.
- En el período inicial el 17% de los sitios no tenían implementación de comunicación activa, siendo sitios netamente informativos, sin posibilidad de que el ciudadano interactuó en ellos. Actualmente, todos los sitios han implementado comunicación activa, oscilando los puntajes logrados desde 39% al 74% del puntaje esperado. En cuanto a comunicación pasiva se puede observar una notable mejoría en los sitios de la muestra.
- El 20% de los sitios en el período inicial no tenían implementación de e-Transparencia. Actualmente todos los sitios han agregado contenidos o recursos que mejoran el grado de transparencia, las puntuaciones obtenidas oscilan desde el 46% al 89% del valor máximo posible.

TABLA 3. Resultados por Pilares en Ambos Períodos de Análisis

MUNICIPIO	S.		D.		C				T.	
					A.		P.			
3 DE FEBRERO	82	88	24	74	15	49	91	113	31	66
ALTE. BROWN	62	105	59	87	25	58	96	134	45	88
AVELLANEDA	52	78	49	94	18	64	83	108	35	87
BERAZATEGUI	4	106	0	95	0	66	4	135	0	94
ESTEBAN ECHEVERRIA	52	68	41	75	27	46	66	97	20	65
EZEIZA		76		92		54		114		81
FLORENCIO VARELA	81	76	53	64	29	36	105	104	45	62
GRAL RODRIGUEZ	10	85	0	90	0	57	10	118	0	84
HURLINGHAM	54	98	31	79	13	57	72	120	24	83
ITUZAINGÓ	75	108	78	87	38	69	115	126	63	79
JOSE C. PAZ	31	43	21	67	8	43	44	67	20	49
LA MATANZA	78	77	48	90	28	57	98	110	33	84
LANÚS	71	75	59	89	25	47	105	117	55	83
L.DE ZAMORA	88	100	53	87	37	64	104	123	46	82
LUJÁN	103	81	59	74	49	47	113	108	39	75
MALVINAS ARG.	96	84	70	77	54	44	112	117	56	67
MARCOS PAZ	75	89	53	69	42	56	86	102	49	73
MERLO	2	81	0	82	0	45	2	118	0	77
MORENO	74	79	74	87	38	53	110	113	67	74
MORÓN	108	85	78	79	42	59	144	105	73	71
PILAR	118	102	82	92	56	64	144	130	70	94
PTE. PERÓN	65	41	36	77	24	50	77	68	24	59
QUILMES	85	94	38	79	20	52	103	121	35	89
SAN FERNANDO	110	93	88	95	57	58	141	130	75	95
SAN ISIDRO	101	75	67	103	46	59	122	119	65	80
SAN MARTÍN	95	87	80	97	48	55	127	129	61	91
SAN MIGUEL		84		90		49		125		80
SAN VICENTE	95	59	0	79	0	46	5	92	0	74
TIGRE	74	90	60	85	34	57	100	118	45	82
VICENTE LOPEZ	98	63	41	92	34	53	105	102	40	78

## 5. Formación de Recursos Humanos

El grupo de investigación se encuentra conformado por 9 Docentes, 3 Graduados y 5 Alumnos de la Universidad que se encuentran formándose en actividades de I+ D.

En el área de Gobierno Electrónico se han aprobado 2 tesis doctorales y 1 de maestría. Actualmente integrantes del grupo de investigación se encuentran dirigiendo 1 tesis doctoral y 1 tesina de grado en este área.

## 6. Referencias

[CNC10] CNC (Comisión Nacional de Telecomunicaciones), Buenos Aires, 2010. <http://www.cnc.gov.ar>

[ITU10] ITU (International Telecommunication Union), 2010. [http://www.itu.int/ITU-D/ict/publications/idi/2010/Material/MIS\\_2010\\_without\\_annex\\_4-e.pdf](http://www.itu.int/ITU-D/ict/publications/idi/2010/Material/MIS_2010_without_annex_4-e.pdf)

[GIU08]Giulianelli, Marko, Rodríguez, Trigueros, Vera, Larrosa. Análisis de Gobernabilidad Electrónica: Relevamiento de e-Servicios, e-Democracia, e-Transparencia y Comunicación en Sitios Web Municipales.

[http://sedici.unlp.edu.ar/bitstream/handle/10915/21447/Documento\\_completo.pdf?sequence=1](http://sedici.unlp.edu.ar/bitstream/handle/10915/21447/Documento_completo.pdf?sequence=1)

[IND10] INDEC (Instituto Nacional de Estadísticas y Censos). Censo Poblacional, 2010. <http://www.indec.mecon.ar/>

[ONT05] ONTI. “Plan de Gobierno Electrónico” Decreto 378/2005 – Subsecretaría de la Gestión Pública, 2005. [http://www.sgp.gov.ar/contenidos/onti/productos/pnge/docs/pnge\\_decreto\\_378\\_2005.pdf](http://www.sgp.gov.ar/contenidos/onti/productos/pnge/docs/pnge_decreto_378_2005.pdf)

[UNE01] UNESCO. Definiciones de e-Government. <http://portal.unesco.org/ci/en/files/14896/11412266495e-governance.pdf/e-governance.pdf>

[W3C07] W3C. “Guía Breve de Accesibilidad Web”, 2007. <http://www.w3c.es/divulgacion/guiasbreves/Accessibilidad>

## QUCO2: Una herramienta para Medir la Calidad de las Aplicaciones Web.

Noelia Pinto<sup>1</sup>, Nicolás Tortosa<sup>1</sup>, Liliana Cuenca Plestch<sup>1</sup>, César Acuña<sup>1</sup>, Marcelo Estayno<sup>2</sup>

<sup>1</sup>Departamento de Ingeniería en Sistemas de Información, UTN-FRRe, French 414, (3500), Resistencia, Chaco, Argentina. Tel. 362-4432683

<sup>2</sup>Departamento de Informática - Facultad de Ingeniería – Universidad Nacional de Lomas de Zamora, (1832), Lomas de Zamora, Buenos Aires, Argentina.

### Resumen

Lograr un alto nivel de calidad de un producto o servicio es el objetivo de la mayoría de las organizaciones que desarrollan software. La administración de la calidad del software utiliza procedimientos y estándares durante el desarrollo del software, además del correspondiente proceso que verifica que todo el personal siga estos estándares. En un esfuerzo por definir el concepto de calidad, algunos autores argumentan que un atributo de calidad puede contribuir a la obtención de mejoras en el funcionamiento y operación del software. El presente artículo sintetiza los resultados y proyecciones de la implementación de una herramienta computacional para evaluar la calidad de aplicaciones Web, teniendo en cuenta las características propias de este tipo de software. La herramienta gestiona los elementos de un modelo de calidad jerárquico, desarrollado a tal fin, y permite recolectar la información necesaria para evaluar productos diversos y con características contrastantes.

**Palabras Clave:** Modelo de Calidad de Software, Métricas de Calidad, Calidad en Aplicaciones Web.

### Contexto

La presente línea de investigación se enmarca en el proyecto de investigación interinstitucional, titulado

“Modelos y Métricas para la evaluación de la calidad del software”, que llevan adelante docentes investigadores de las Facultades Regional Resistencia de la UTN y de Ciencias Exactas y Naturales de la UNNE, bajo la dirección de un investigador de la Universidad Nacional de Lomas de Zamora.

### Introducción

Las cuestiones de calidad en los productos o servicios de cualquier índole se ha hecho algo cotidiano y cada vez más importante puesto que se convirtió en un factor diferenciador cuando se está optando por adquirir un servicio o un bien. El concepto de calidad tiene diferentes definiciones, pero una ampliamente aceptada es la establecida por la ISO 9000[1] la cual define la calidad como el “grado en el que un conjunto de características inherentes cumple con los requisitos”. Estos requisitos son establecidos por los usuarios y por tanto se puede decir que la calidad es un tema subjetivo dependiente del nivel de satisfacción del usuario sienta frente al producto utilizado. El software, como cualquier otro producto, también está sujeto a la evaluación de su calidad, de tal manera que los usuarios puedan establecer el grado con el cual éste satisface sus necesidades.

Incluso, la principal intención de la Ingeniería de Software es mejorar la

“calidad de sus productos” para lograr que éstos sean competitivos, y así se ajusten a los requerimientos establecidos por los usuarios finales.

Un tipo especial de productos software son las aplicaciones web, que exhiben factores en su proceso de desarrollo y en los resultados obtenidos que dificultan evaluar su calidad y por tanto obtener productos que satisfagan los requerimientos de origen.

Actualmente ha aumentado la necesidad que las aplicaciones web aseguren más confiabilidad al usuario y mejoren aspectos en su calidad de uso. Este último aspecto es uno de los más nuevos incorporado al estándar ISO/IEC 9126 2001 [2] [3] [4] [5]. En éste se define al modelo de calidad como “la opinión que tiene un usuario al utilizar una aplicación de software”, la cual se deriva de los resultados obtenidos de una evaluación.

En el presente artículo se describe cómo a partir del modelo de calidad en uso propuesto por la ISO/IEC 9126 2001, pero teniendo en cuenta las características propias de software web, se ha logrado la implementación exitosa de una herramienta tecnológica destinada a la evaluación de aplicaciones web. Asimismo se presenta la comparación del software desarrollado con otros, existentes en el mercado, y orientados, también, a la evaluación de aspectos de calidad en aplicaciones web.

Debido a la importancia que la calidad de software en Internet ha despertado en los últimos años, la Conferencia Internacional de la Ingeniería de Software del año 2002 (ICSE 2002) se centró en los aspectos de Calidad para los Sistemas en Internet [6] [7]. En esta conferencia se concluyó que las métricas más importantes son los siguientes: Fiabilidad, Usabilidad, Mantenibilidad,

Seguridad, Disponibilidad y Escalabilidad [8].

Gracias al trabajo de investigación llevado adelante, se ha concluido que el estudio de los atributos de calidad para sistemas de software puede ser muy extenso. Por lo tanto, en este trabajo se orienta al estudio e implementación de las tres primeras métricas nombradas, que se han considerado como las más relevantes para la evaluación de aplicaciones web.

## **Líneas de Investigación y Desarrollo**

En este proyecto se ha desarrollado un Modelo de Calidad para la Evaluación de aplicaciones web. El objetivo fundamental del mismo es contribuir a la mejora en la calidad de los productos de software mediante modelos y métricas aplicados al producto y al proceso de creación, diseño, desarrollo y mantenimiento, como medio para aumentar la competitividad de las pymes de la región NEA en el contexto de la industria del software.

De este modo, uno de los ejes que se presentan tiene que ver con el desarrollo de un Framework para integrar la información de la gestión de calidad de un producto software. El objetivo del mismo es evidenciar y monitorear los esfuerzos del equipo de desarrollo en pos de la calidad, permitiendo que tanto el cliente como el evaluador de calidad puedan visualizar el proceso de gestión de una forma clara y uniforme [9].

Se pretende además estudiar los resultados de la implementación de la herramienta desarrollada para lo cual se requirió de la elaboración de un caso de estudio de ejemplo. Este caso de estudio tiene como finalidad mostrar el uso de la herramienta desarrollada en una evaluación real de calidad en diferentes



sitios web. El caso de estudio contempló las siguientes actividades: a) definir un contexto de uso, b) seleccionar la población, c) definir los parámetros de evaluación y d) comparar resultados arrojados por QUCO con otras herramientas.

## Resultado y Objetivos

Mediante el análisis de los resultados obtenidos en el caso de estudio presentado, se demostró el correcto funcionamiento de la herramienta en una evaluación de calidad sobre un entorno real. El uso del plugin no presentó dificultades en su curva de aprendizaje de acuerdo al feedback obtenido de parte de los evaluadores participantes.

Una vez finalizada la etapa de evaluación con QUCO2, se procedió a obtener un promedio de los valores obtenidos para cada sitio propuesto. En base a una ponderación de los pesos de las métricas consideradas, se obtuvo que el valor ideal o de máxima satisfacción ocupa el rango entre 20 y 22, el valor de calidad media el rango entre 14 y 20, el valor de calidad regular entre 8 y 14 y el mínimo nivel con un valor por debajo de 7.

La herramienta presentó fiabilidad al momento de aplicar la evaluación del caso de estudio, y su utilización proporcionó datos confiables que permiten a los Clientes obtener una visión aproximada de la calidad en uso del sitio web evaluado. Luego, teniendo en cuenta la comparación de los resultados que ha arrojado frente a los valores obtenidos con el uso de las otras dos herramientas, ya disponibles en el mercado, se observa similitud en los niveles de calidad para los sitios web del caso de estudio; contribuyendo a validar el modelo de calidad utilizado en el desarrollo del framework.

Además se pudo comprobar que QUCO2 representa una herramienta que engloba diferentes aspectos obteniendo un valor final de evaluación considerando un conjunto de métricas, y no centrándose en un único aspecto de calidad. Destacando también que al ser una aplicación opensource facilita su implementación y uso en cualquier ámbito.

Los objetivos que se han planteado para llegar a estos resultados fueron:

- Realizar estudios de investigación de los diversos modelos de calidad actuales.
- Elaborar un modelo de calidad que se ajuste a la evaluación de calidad de productos de software web.
- Desarrollar una herramienta tecnológica basada en el modelo de calidad elaborado.
- Comparar los resultados obtenidos de la aplicación creada con la evaluación generada a partir de otras herramientas existentes.
- Contribuir al proceso de mejora continua en empresas PyMes dedicadas al Desarrollo de Software en la región NEA.

## Formación de Recursos Humanos

### Equipo de Trabajo

El equipo de trabajo involucrado en el desarrollo del trabajo aquí presentado está formado por:

- Mg. Liliana Cuenca Pletsch
- Dr. César Acuña
- Mg. Marcelo Estayno
- Ing. Noelia Pinto
- Sr. Nicolás Tortosa

### Formación de Becarios

La Ing. Noelia Pinto, egresada de Ing. en Sistemas de Información, ha sido beneficiada con la beca BINID de Investigación otorgada por Rectorado de la UTN con el objetivo de contribuir a su formación profesional y promover la investigación en la docencia.

El Sr. Nicolás Tortosa, alumno avanzado de la carrera Ing. en Sistemas de Información, también ha sido beneficiado con la beca BINID otorgado por la misma Institución nombrada anteriormente. Él forma parte del equipo de trabajo y colabora con las actividades del proyecto de investigación.

### Referencias

- [1] ISO. “Sistemas de gestión de la calidad-Conceptos y vocabulario. Norma Internacional ISO 9000”, 2001.
- [2] ISO, “ISO/IEC 9126-1 – Software engineering–Product quality – Part 1: Quality Model”, 2001.
- [3] ISO, “ISO/IEC 9126-2 – Software engineering– Product quality – Part 2: External Metrics”, 2003.
- [4] ISO, “ISO/IEC 9126-3 – Software engineering– Product quality – Part 3: Internal Metrics”, 2003.
- [5] ISO, “ISO/IEC 9126-4 – Software engineering– Product quality – Part 4: Quality in Use Metrics”, 2004.
- [6] J. OFFUTT. Quality Attributes of Web Software Applications. IEEE Software, 0740-7459/02. March/April 2002.
- [7] BECKER, F.E. MOTTAY. A Global Perspective on Web Site Usability. IEEE Software, 0740-7459/00, January/February 2001.
- [8] DÁVILA NICANOR, MEJÍA ALVÁREZ. “Evaluación de la Calidad de Software en Sistemas de Información en Internet”. CINVESTAV-IPN. Sección de Computación.
- [9] ACUÑA, C., GARCÍA, L., FERRARO, M., CASIVA, A., CUENCA PLETSCH, L. “Calidad del Software aplicada a los servicios web y a los servicios web semánticos”. WICC-2011. Publicado en anales. ISBN 978-950-673-892-1.

# Generación Automática de código RMI a partir del Modelo de Análisis utilizando Reglas Relations/QVT

Ariel Arsaute, Marcela Daniele, Mariana Frutos, Ariel Gonzalez, Marcelo Uva, Fabio Zorzan.

Departamento de Computación, Facultad de Ciencias Exactas, Físico-Químicas y Naturales,  
Universidad Nacional de Río Cuarto.

Ruta 36 Km. 601 –CP 5800 - Río Cuarto – Córdoba - Argentina Tel. (0358) 4676235  
{aarsaute, marcela, mfrutos, agonzalez, uva, fzorzan}@dc.exa.unrc.edu.ar

## Resumen

Este trabajo realiza una contribución tendiente a la mejora del proceso de desarrollo de software, siguiendo la filosofía de Model Driven Architecture (MDA). Dentro de esta línea de investigación en trabajos previos [1,2,10] se definió la transformación del modelo de CU de sistema al modelo de análisis y del modelo de análisis al modelo de diseño considerando la tecnología de implementación Remote Method Invocation (RMI). Para lograr estas transformaciones se especificaron una serie reglas de transformación Query/View/Transformation (QVT) que permitieron establecer la correspondencia entre los modelos fuentes y destinos correspondientes. Los modelos de análisis y diseño obtenidos como resultado de aplicar las distintas transformaciones QVT (modelos objetivos), atañen a modelos UML correspondiente a los diagrama de clases de análisis y diseño respectivamente.

En este trabajo se pretende la generación automática de código, tomando como entrada el modelo de diseño resultado y haciendo uso de la herramienta Acceleo que es un generador de código que transforma los modelos en código.

El objetivo final es la realización de aportes tendientes a lograr un producto de software siguiendo los principios de abstracción, automatización y estandarización establecidos por MDA, de manera totalmente automática, o bien automatizando el proceso de transformación lo mayormente posible.

**Palabras clave:** MDA, QVT, Proceso Unificado, UML.

## Contexto

La línea de investigación presentada en este trabajo se desarrolla en el marco del proyecto “Ingeniería de Software: Automatización de Procesos de Desarrollo de Software”, perteneciente a los Proyectos y Programas de Investigación (PPI) de la secretaría de Ciencia y Técnica de la Universidad Nacional de Río Cuarto.

## Introducción

La ingeniería de Software implementa un enfoque sistemático, disciplinado y cuantificable para el desarrollo, operación y mantenimiento del software. Las metodologías de desarrollo junto con las herramientas de software se han adoptado con éxito en un amplio espectro de aplicaciones industriales. Dentro de las metodologías orientadas a objetos, el Proceso Unificado [3] define actividades y responsabilidades estableciendo quién está haciendo qué, cuándo, y cómo para construir o mejorar un producto de software. Esta metodología, divide el desarrollo del producto de software en fases utilizando UML (Unified Modeling Language) como lenguaje de modelado [4].

### ARQUITECTURA DIRIGIDA POR MODELOS

La Arquitectura Dirigida por Modelos (MDA) [5] se ha concebido para dar soporte a la ingeniería dirigida por modelos de los sistemas software, cuyo objetivo central es resolver el problema producido por el cambio de tecnologías junto con la integración del sistema de software y que todo esto no impliquen un alto costo.

La idea subyacente en MDA es usar modelos, de modo que las propiedades y características de los sistemas queden contenidas en un modelo abstracto independiente de los cambios producidos en la tecnología. MDA proporciona una serie de guías o patrones expresadas como modelos. MDA propone cuatro niveles de abstracción que componen la jerarquía o arquitectura de modelos. Estos son: Modelo independiente de cómputo CIM (Computation Independent Model), Modelo independiente de la plataforma PIM (Platform Independent Model), Modelos específicos de la plataforma PSM (Platform Specific Model), y la aplicación final.

Los modelos CIM describen el entorno en el que se utilizará el sistema, sin referencia directa a su implementación.

Los PIM modelan funcionalidad y estructura de un sistema de información sin considerar detalles tecnológicos de la plataforma en la cual se implementará el sistema.

Los PSM describen los modelos específicos de plataforma, concretamente de la plataforma tecnológica donde se ejecutará el sistema.

MDA plantea el siguiente proceso de desarrollo: de los requisitos se obtiene un modelo independiente de la plataforma (PIM), luego este modelo es transformado con la ayuda de herramientas en uno o más modelos específicos de la plataforma (PSM), y finalmente cada PSM es transformado en código. Por lo tanto, MDA incorpora la idea de transformación de modelos (PIM a PSM, PSM a código) y se necesitan herramientas para automatizar estas tareas. Estas herramientas constituyen uno de los elementos básicos de MDA.

### QVT (Query / View / Transformation)

El planteamiento QVT se basa principalmente en la definición de un lenguaje para las consultas (Queries) sobre los modelos Meta Object Facility (MOF) [6], la búsqueda de un estándar para generar vistas (Views) que revelen aspectos específicos de los sistemas modelados, y finalmente, la definición de un lenguaje para la descripción de transformaciones de modelos MOF.

En este trabajo se plantea la utilización de QVT para definir transformaciones entre modelos. Estas transformaciones describen relaciones entre un metamodelo fuente F y un metamodelo objetivo O, ambos metamodelos deben estar especificados en MOF. Luego esta transformación se utiliza para obtener un modelo objetivo, el cual es una instancia del metamodelo O, a partir de un modelo fuente que es una instancia del metamodelo F. Una característica muy importante de estas transformaciones es que pueden ser bidireccionales y multidimensionales.

La herramienta CASE utilizada para la definición de las reglas QVT es MediniQVT [7]. Esta herramienta implementa la especificación QVT/Relations de la OMG en un poderoso motor QVT. MediniQVT está diseñada para transformaciones de modelos permitiendo un rápido desarrollo, mantenimiento y particularización de reglas de transformación de procesos específicos. La herramienta está integrada a Eclipse y utiliza Eclipse Modeling Framework (EMF) [8] para la representación de modelos.

### Líneas de investigación y desarrollo

Como se mencionó anteriormente, MDA propone basar el desarrollo de software en modelos, separando el diseño de la arquitectura y de las tecnologías de construcción, posibilitando que el diseño y la arquitectura puedan ser alterados independientemente. Una vez definidos los modelos, se definen una serie de transformaciones que generan nuevos modelos, permitiendo ir de modelos más abstractos a otros más concretos. MDA define un framework para procesar y relacionar modelos.



La transformaciones entre modelos se realizan haciendo uso de herramientas automáticas, como herramientas de transformación de modelos, las cuales permiten el refinamiento de los mismos.

Nuestra propuesta consiste en tomar el modelo de diseño UML obtenido como resultado del trabajo anterior, el cual corresponde a una instancia del metamodelo UML. A partir de este modelo pretendemos generar código utilizando Acceleo [9]. Esta herramienta permite la generación de código a través de cualquier tipo de metamodelo compatible con EMF como UML 1, UML 2 e incluso metamodelos adaptados a nuestro dominio específico.

De esta manera se pretende lograr un acercamiento a la automatización de la construcción del software aplicada a aquellos procesos de desarrollo dirigidos por casos de uso.

### Resultados y Objetivos

En esta etapa se trabajará en la generación automática de código a partir del modelo de diseño obtenido como resultado de aplicar las sucesivas transformaciones QVT definidas entre el modelo de casos de uso de sistema y el modelo de diseño, como se observa en la Fig.1.

La aplicación de las transformaciones QVT produce un modelo de diseño que consiste en un diagrama UML de clases de diseño.

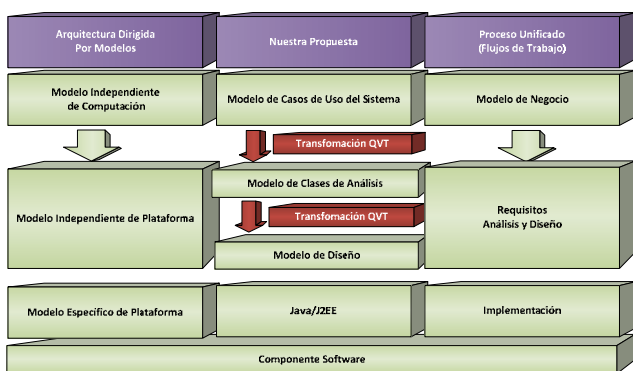


Fig. 1. Transformación modelo CUs a Modelo de Clases de Diseño

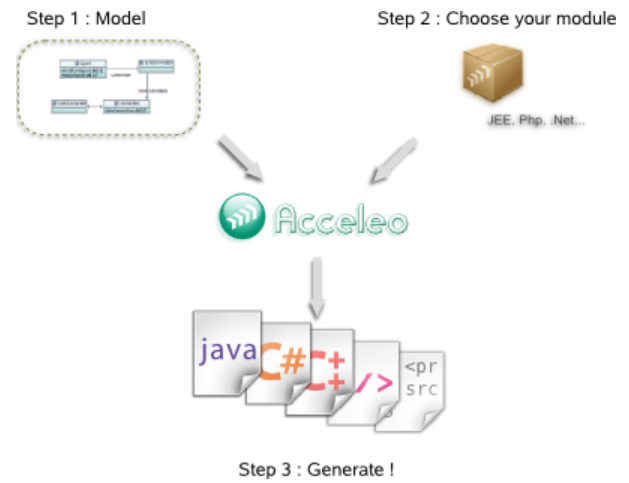


Fig. 2. Generación de código utilizando Acceleo

Este modelo nos resulta conveniente para ser utilizarlo como modelo entrada para la herramienta Acceleo como podemos ver en la Fig.2, obteniendo como resultado el código generado del lenguaje seleccionado.

Acceleo es una herramienta de transformación de modelo a texto, con un lenguaje basado en plantillas o templates. Teniendo en cuenta los estándares definidos por EMF, Acceleo brinda un enfoque MDA simple permitiendo la generación de archivos a partir de modelos UML, MOF o EMF

Acceleo brinda un asistente para la generación automática de las mismas. Primeramente se debe indicar el metamodelo a ser utilizado. Luego se debe seleccionar el modelo en sí. Acceleo está basado en los principales estándares MDA, lo que garantiza su compatibilidad e interoperabilidad con otras aplicaciones, como GMF. Es compatible con XMI, fue diseñado para trabajar con cualquier metamodelo y permite extender la funcionalidad ofrecida mediante la importación de librerías de Java, que pueden utilizarse para agregar funcionalidad a las plantillas y la generación de código. Una vez seleccionado el metamodelo y el modelo de entrada, se debe indicar las plantillas utilizadas en el proceso de generación. Luego se guarda la cadena indicando su nombre y la ubicación donde se alojará.

El proceso de transformación comenzará con la ejecución de una de las cadenas de generación creadas. El archivo producto de la transformación será generado en el formato que se haya indicado.

El resultado buscado está estrechamente vinculado al logro de un acercamiento hacia la automatización de las actividades que componen el proceso de desarrollo de software, transformado modelos ya validados como lo son las Plantillas Genéricas de descripción, análisis y diseño de casos de uso, con el fin de obtener un producto de software de calidad, cumpliendo con los requerimientos iniciales.

### Formación de Recursos Humanos

Durante el desarrollo de esta línea de investigación han logrado obtener el título Magister en Ingeniería de Software tres integrantes del grupo de trabajo. Otro integrante está actualmente trabajando es su tesis de Magister. Dos grupo de alumnos están realizando su trabajo final de fin de carrera de Analista en Computación, otro grupo ha logrado obtener el título de Licenciado en Ciencias de la Computación.

También, se han formado ayudantes de segunda en las asignaturas de Análisis y Diseño de Sistemas, Ingeniería de Software, Base de Datos y Proyecto.

Los temas abordados en esta línea de investigación brindan un fuerte aporte al proceso de perfeccionamiento continuo de los autores de carreras de computación en Universidades Nacionales como del exterior.

### Referencias

- [1] Ariel Arsaute, et al. "Hacia una integración de MDA y el Proceso Unificado a través de reglas de transformación QVT". 13th Argentine Symposium on Software Engineering, ASSE 2012.
- [2] Ariel Arsaute, et al. "Transformación del modelo de análisis al modelo de diseño utilizando QVT". XIV Workshop de Investigadores en Ciencias de la Computación 2012.
- [3] Jacobson, I. El Proceso Unificado de Desarrollo de software. Addison -Wesley, EE.UU. 2000.
- [4] UML, Unified Modeling Language (UML) Resource Page. <http://www.omg.org/#UML2.0>
- [5] Miller, J., Mukerji, J., MDA Guide Version 1.0.1 Document number omg/2003-06-01, Disponible en: <http://www.omg.com/mda>, 2003
- [6] Object Management Group, "Meta Object Facility (MOF)2.0Query/View/Transformation Specification" <http://www.omg.org/docs/ptc/05-11-01.pdf>.
- [7] ikv++: medini QVT. <http://www.ikv.de/>, ultimo acceso Agosto 2011
- [8] Eclipse Modeling Framework, URL: <http://www.eclipse.org/modeling/emf/>, ultimo acceso Abril 2009.
- [9] Acceleo. Home: <http://www.acceleo.org/>. Última visita Marzo 2012.
- [10] N. Debnath, F. A. Zorzan, G. Montejano and D. Riesco, "Transformation of BPMN Subprocesses Based in SPEM Using QVT", 2007 IEEE INTERNATIONAL CONFERENCE on ELECTRO/INFORMATION TECHNOLOGY, May 17-20, 2007, Marriott O Hare, Chicago, IL, USA. <http://www.eit-conference.org/eit2007/>

## Modelado y Desarrollo de Sistemas Robóticos Educativos

Claudia Pons, Gonzalo Zabala, Gabriela Pérez  
 CONICET, Universidad Abierta Interamericana (UAI) y Facultad de Informática de la UNLP  
 Buenos Aires, Argentina

### RESUMEN

La inclusión del sistema educativo argentino al modelo de “una computadora portátil por niño” a través del programa “Conectar Igualdad”, entre otros, junto con el sólido propósito tanto del Ministerio de Educación como del Ministerio de Ciencia y Tecnología de equipar a las escuelas con artículos tecnológicos concretos, tales como robots y adquirentes de datos, nos da una oportunidad única de desarrollar plataformas robóticas / automatizadas dirigidas a la educación. Mientras que los sistemas robóticos crecen hasta ser más y más complejos, la necesidad de formalizar su proceso de desarrollo de software crece también. Los enfoques tradicionales que se utilizan en el proceso de desarrollo de estos sistemas de software están alcanzando sus límites; las metodologías y el conjunto de herramientas no alcanzan para atender las necesidades de tal complejo proceso de desarrollo de software. El objetivo general de este proyecto es la definición de un marco metodológico sustentado por un conjunto de herramientas tecnológicas capaces de lidiar con los requerimientos del proceso de desarrollo del software robótico. Un desafío muy importante es dar el paso desde “code-driven” a “model-driven” en el desarrollo de sistemas de software de robótica. La separación del conocimiento de robótica de la aplicación de tecnologías de ciclo corto es esencial para fomentar el reuso y el mantenimiento. El contexto en el cual serán aplicados los resultados comprende el campo de desarrollo de los sistemas robóticos en general, pero los remitiremos específicamente a los sistemas robóticos educativos que podrían ser potencialmente usados en las aulas argentinas.

**Palabras clave:** robótica, ingeniería de software, educación.

### 1. CONTEXTO

Este proyecto cuenta con la participación de profesores / investigadores del CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas), de la Universidad Abierta Interamericana (UAI), de la Universidad Nacional de La Plata (UNLP) y de la Universidad de la República (Uruguay). Los investigadores que participan en este proyecto se complementan muy

bien de acuerdo a sus áreas de especialización; contribuyen al conocimiento en las áreas de modelado de software, ingeniería de software y robótica.

### 2. INTRODUCCIÓN Y PLANTEO DEL PROBLEMA

Los sistemas robóticos (SR) juegan un papel cada vez mayor en la vida cotidiana. También la necesidad de sistemas automatizados en entornos industriales y educativos aumenta y se vuelve más exigente. Mientras que los sistemas robóticos crecen y cada vez son más complicados, la necesidad de formalizar su proceso de desarrollo de software crece también. Los enfoques tradicionales que se utilizan en el proceso de desarrollo de estos sistemas de software están alcanzando sus límites; las metodologías utilizadas actualmente y los conjuntos de herramientas no alcanzan para atender las necesidades de dicho proceso de desarrollo de software complejo.

La inclusión del sistema educativo argentino al modelo de “una computadora portátil por niño” a través del programa “Conectar Igualdad” [1], entre otros, junto con el sólido propósito tanto del Ministerio de Educación como del Ministerio de Ciencia y Tecnología de equipar las escuelas con artículos tecnológicos concretos, tales como robots y adquirentes de datos, nos da una oportunidad única de desarrollar plataformas robóticas dirigidas a la educación.

Está ampliamente aceptado que nuevos enfoques deben ser establecidos para satisfacer las necesidades del complejo proceso de desarrollo SR de hoy. En esta dirección, el desarrollo basado en componentes (CBD, por sus siglas en inglés Component-based development) [5], la Arquitectura Orientada a Servicios (SOA, por sus siglas en inglés Service Oriented Architecture) [12] [13], así como la Ingeniería de software dirigida por modelos (MDE, por sus siglas en inglés Model Driven software Engineering) [15] [16] y el Modelado Específico de Dominio (DSM, por sus siglas en inglés Domain-Specific Modeling) [14] se encuentran entre las principales y prometedoras tecnologías en el ámbito de SR. En estos días, estamos realizando una investigación sobre el uso actual de estas técnicas modernas de ingeniería de software para el desarrollo de sistemas de software

robótico y su nivel de automatización real. Un desafío muy importante es dar el paso desde “code-driven” a “model-driven” en el desarrollo de sistemas de software de robótica

### 3. OBJETIVOS, HIPÓTESIS Y RESULTADOS ESPERADOS

El objetivo general de este proyecto tiene dos partes: desde el punto de vista técnico queremos definir un marco metodológico con el apoyo del proceso MDD para hacer frente a las exigencias del proceso e desarrollo de software robótico, y desde el punto de vista educativo, queremos desarrollar el software y hardware para complementar las netbooks existentes distribuidas en las escuelas primarias de Argentina. El contexto en el cual los resultados serán aplicados comprende el campo de desarrollo de sistemas robóticos en general, pero los vamos a dirigir específicamente a los sistemas educativos robóticos que pueden ser potencialmente utilizados en las aulas argentinas. Esas plataformas robóticas educativas deben poseer una capacidad de adaptación muy dinámica, acompañando a la tasa de desarrollo de esas tecnologías y a las diferencias locales de cada una de las plataformas de hardware. Debido a este hecho, es esencial contar con potentes herramientas de ingeniería de software que facilitan la construcción de frameworks para hacer frente a tal complejidad y volatilidad tecnológica.

Basado en el contexto y los objetivos explicados anteriormente, vamos a trabajar bajo las siguientes hipótesis, en las que combinamos nuestro conocimiento de los aspectos técnicos de los sistemas robóticos y el impacto social que generan el uso de nuestros resultados en las netbooks distribuidas en el programa "Conectar-Igualdad".

- Es indispensable propiciar la aplicación de los principios de ingeniería para hacer frente a la complejidad de los sistemas de software robóticos, porque no podemos esperar un crecimiento significativo con sistemas hechos artesanalmente.
  - Las Interfaces y el comportamiento de los sistemas robóticos se deben definir en un nivel superior de abstracción para que puedan ser reutilizados con diferentes plataformas tecnológicas. Lograr una separación de los conocimientos de robótica de las tecnologías de ciclo corto es esencial para fomentar la reutilización y el mantenimiento.
  - La aplicación de tecnologías existentes de ingeniería de software, tales como SOA, MDE, y CBD, para construir sistemas de software robóticos podría ahorrar una gran cantidad de tiempo y esfuerzo, además de favorecer la reutilización y el mantenimiento de dichos sistemas.
- Dichas técnicas de ingeniería de software

podrán ser aplicadas a la construcción de software educativo robótico dirigido a la enseñanza de la tecnología y la ciencia en las escuelas argentinas, dando lugar a mejoras en el rendimiento educativo.

En este contexto, los objetivos específicos del proyecto son los siguientes:

- Resumir la evidencia existente sobre la aplicación de las tecnologías de ingeniería de software, tales como SOA, MDE y el CBD en el campo de desarrollo de sistemas robóticos;
- Identificar las carencias en la investigación actual con el fin de sugerir áreas para futuras investigaciones;
- Proporcionar un fundamento teórico con el fin de posicionar adecuadamente las nuevas actividades de investigación;
- Mejorar las técnicas actuales para ofrecer un avance en el estado del arte;
- Definir las líneas de base para una metodología abierta para el proceso de desarrollo de software robótico.
- Crear herramientas de apoyo al proceso de desarrollo de software robótico. Ejemplos de estas herramientas son: un lenguaje de modelado específico de dominio equipado con editores gráficos, facilidades de generación de código, integración con servicios web, editores de definición de componentes, etc.
- Usar los resultados para la construcción de sistemas robóticos educativos que se utilizarán en las aulas argentinas a bajo costo.
- Realizar una serie de experimentos para evaluar la eficacia y la viabilidad de la utilización de sistemas robóticos para mejorar el proceso educativo de los niños argentinos

### 4. LINEAS DE INVESTIGACIÓN Y RELEVANCIA DEL PROBLEMA

Los primeros proyectos que se centraban en la aplicación de las computadoras en el ámbito educativo tenía el objetivo de vincular los elementos tecnológicos concretos con las herramientas digitales ofrecidas por esa tecnología. Un ejemplo es el desarrollo de Lego Logo, donde se controla un robot físico por medio de comandos que se introducen en la computadora. Tales órdenes causar un efecto tanto en el robot concreto, así como en su representación virtual. Por varias razones, esta primera propuesta de mantener la presencia de material no-digital en las áreas de educación tecnológica fue desapareciendo con los años, en proporción inversa a la introducción de computadoras en las escuelas. Así, la educación tecnológica se refiere únicamente al mundo digital, perdiendo la conexión con el material de tecnología



concreto. Debido al hecho de que ciertas estructuras cognitivas no se desarrollan en ausencia de material concreto, esta desconexión muy probablemente podría originar una grave laguna en la educación de los niños. La situación se hace aun más difícil por el hecho de que en la actualidad los juegos habituales de los niños también se asocian sobre todo con el mundo virtual.

En los últimos años, la aparición de kits de robótica orientados a usuarios no expertos dio lugar al desarrollo de un número significativo de proyectos educativos utilizando robots. Estos proyectos se aplican los robots en diferentes niveles educativos, desde pre-escolar hasta educación superior, especialmente en las áreas de la física y tecnología. Los dispositivos que son construidos por los alumnos que utilizan este material ofrecen la posibilidad de nuevos experimentos científicos relacionados con los fenómenos cotidianos. La fuerte motivación que el uso de estos materiales genera en los estudiantes ha alentado a los docentes de otras disciplinas para desarrollar propuestas de utilización de robots en el campo de las artes, las ciencias sociales y otros. Estas nuevas propuestas pueden revertir los déficits cognitivos causados por la no utilización de material concreto en juegos diarios y la educación formal.

En este contexto, uno de los problemas que nos encontramos es que el hardware de los kits de robótica está en constante cambio, además su uso no es uniforme en las diferentes regiones e incluso en los niveles de educación ni desde el punto de vista técnico. Por lo tanto, las interfaces técnicas de estos robots deben ocultar estas diferencias para que los educadores no se vean forzados a cambiar su material educativo una y otra vez. Un ejemplo de estas interfaces es "*Physical Etoys*" [2], un proyecto en el que participamos y que propone una plataforma de enseñanza estándar para la programación de robots, independientemente de si están basados en Arduino, Lego, u otras tecnologías. *Physical Etoys* ha logrado muy buena aceptación en la comunidad académica en todo el mundo y proporciona un punto de partida prometedor para el desarrollo de una poderosa plataforma de ingeniería de software para el desarrollo de sistemas robóticos aplicables en las aulas argentinas.

Aunque la complejidad del software robótico es alta, en la mayoría de los casos su reutilización todavía está restringida al nivel de bibliotecas. En el nivel inferior, se han creado una multitud de bibliotecas para realizar tareas como cálculos matemáticos para cinemática, dinámica y visión de la máquina, como por ejemplo [3]. En vez de construir los sistemas a partir de bloques que provean servicios ya verificados, el proceso de construcción de software de un nuevo sistema robótico a menudo sigue siendo la re-

implementación de la lógica para unir las diferentes bibliotecas. A menudo, la integración global está orientada a un middleware determinado y sus capacidades. Los middlewares se utilizan para ocultar la complejidad de la comunicación entre componentes, por ejemplo OpenRTM-AIST [4] es un middleware basado en CORBA para las plataformas de robots. Obviamente, esto no sólo es caro y consume enormes recursos sino que desaprovecha la madurez alcanzada por los procesos de desarrollo de software existentes.

En concreto, los sistemas de robótica tienen necesidades especiales, a menudo relacionadas con su naturaleza de tiempo real y las propiedades del ambiente. Este tipo especial de sistemas necesita más calidad que un sistema de propósito general y tiene que ser capaz de hacer frente al entorno físico incierto y dinámico en el que están inmersos. Atributos como la fiabilidad y la seguridad son una necesidad en este dominio.

En este contexto, es ampliamente aceptado que nuevos enfoques deben ser aplicados para satisfacer las necesidades de este proceso de desarrollo tan complejo. El Desarrollo Basado en Componentes (CBD), la Arquitectura Orientada a Servicios (SOA), así como la Ingeniería de software Dirigida por Modelos (MDE) y el Modelado de Dominio Específico (DSM) se encuentran entre las tecnologías más prometedoras en el ámbito de SRs.

En primer lugar, el paradigma CBD [5] establece que el desarrollo de aplicaciones debe lograrse mediante la vinculación de partes independientes, los componentes. Las interfaces de los componentes, basadas en estrictos patrones de interacción, separan la esfera de influencia y por lo tanto particionan la complejidad global. Esto se traduce en componentes débilmente acoplados que interactúan a través de servicios con contratos. Los componentes como unidades arquitectónicas permiten especificar con mucha precisión, utilizando el concepto de puerto, tanto los servicios prestados como los servicios requeridos por un determinado componente. Además definen una estrategia de composición basada en la noción de conectores. La tecnología de componentes ofrece altas tasas de reutilización y facilidad de uso, pero poca flexibilidad con respecto a la plataforma de implementación. Los componentes actuales están relacionados con C / C++ y Linux (por ejemplo, Microsoft Robotics Developer Studio [6], EASYLAB [7], Player/Stage Project[10]), aunque algunos alcanzan más independencia, gracias a la utilización de alguno middleware (por ejemplo, Smart Software Component Model [11], Orocos [3] Orca [8] CLARATy [9]).

En segundo lugar, necesitamos una manera de definir las interfaces y comportamiento a un

nivel más alto de abstracción para que puedan ser utilizados en sistemas con diferentes plataformas. Esto originó la idea de componentes abstractos, es decir, independientes de la plataforma de implementación. De este modo, para superar los problemas actuales resulta imperioso migrar de diseños basados en el código hacia diseños basados en los modelos. Una descripción basada en el modelo es un medio adecuado para expresar los contratos de las interfaces de los componentes, verificar el comportamiento general de los sistemas integrados y finalmente obtener automáticamente el software ejecutable. En lugar de construir herramientas de soporte para cada framework a partir de cero, ahora la estrategia consiste en expresar los modelos necesarios en lenguajes estándar de modelado tales como UML, logrando la separación de los componentes del hardware subyacente. En el contexto de la ingeniería de software, el desarrollo dirigido por modelos (MDE) y el enfoque de modelado de dominio específico (DSM) han dado lugar a un cambio de paradigma de “code-centric” a “model-centric”. Estos enfoques promueven la sistematización y la automatización de la construcción de artefactos de software. Los modelos son considerados construcciones de primera clase en el desarrollo de software, y el conocimiento de los desarrolladores es encapsulado por medio de transformaciones de modelos. La característica esencial del MDE y DSM es que el principal objetivo del desarrollo de software son los modelos. Su principal ventaja es que los modelos se pueden expresar en diferentes niveles de abstracción y por lo tanto están menos ligados a una tecnología específica. Esto es especialmente relevante para los sistemas de software en el dominio de la computación ubicua, que consisten en aplicaciones dinámicas y distribuidas y corren sobre plataformas heterogéneas de hardware, tales como los sistemas robóticos.

Por último, la arquitectura orientada a servicios (SOA) es un conjunto flexible de principios de diseño utilizados durante las fases de desarrollo e integración de sistemas de computación. Un sistema basado en una arquitectura SOA empaqueta su funcionalidad como un conjunto de servicios interoperables que pueden ser utilizados dentro de múltiples sistemas, pertenecientes a dominios de negocios varios. SOA define cómo integrar aplicaciones muy dispares para un entorno basado en la Web y utiliza múltiples plataformas de ejecución. En lugar de definir una API, SOA define la interfaz en términos de protocolos y funcionalidad. SOA requiere el acoplamiento flexible de servicios con los sistemas operativos, y otras tecnologías que subyacen en las aplicaciones. SOA separa las funciones en unidades distintas, o servicios [12], que los desarrolladores hacen accesibles a través de

una red a fin de permitir a los usuarios combinarlos y reutilizarlos en la producción de sus aplicaciones. Estos servicios y sus consumidores se comunican entre sí pasando los datos en un formato común, bien definido, o mediante la coordinación de una actividad entre dos o más servicios [13].

Por lo tanto, los paradigmas CBD y SOA proporcionan un punto de partida para lograr un enfoque MDE en robótica, donde las diferencias entre las distintas plataformas de software y sistemas middleware pueden quedar completamente ocultos al usuario debido a la definición de niveles de abstracción intermedios. Por el momento, no hay ninguna propuesta que tome ventaja de la aplicación combinada de CBD, SOA y MDE para el desarrollo de sistemas de software robóticos en general, ni tampoco para el desarrollo de sistemas robóticos educativos en particular.

## 5. CONTRIBUCION ORIGINAL

La contribución original de este proyecto consiste en el desarrollo de un marco metodológico equipado con herramientas para la construcción de sistemas de software robóticos, en particular nos hemos centrado en la propuesta educativa de Physical Etoys y Butia [23], pero consideramos que la metodología es de carácter lo suficientemente general para ser aplicada a otros sistemas robóticos también. Hoy en día existen sólo propuestas preliminares sobre la aplicación de MDE en la robótica (ver por ejemplo los trabajos descritos en [18], [19], [20] y [21]). Ninguna de estas propuestas provecha la combinación del paradigma MDE con enfoques orientados a servicios y basados en componentes, tal como proponemos en este proyecto.

Resultados preliminares de nuestro proyecto han sido publicados en [25,26 y 27].

## 6. RECURSOS HUMANOS

El equipo está integrado por cinco investigadores senior, el Prof. Paul Puleston es actualmente Profesor Titular en la UNLP e investigador del Consejo Nacional de Investigación Científica y Técnica, en el Laboratorio de Electrónica Industrial, Control e Instrumentación (LEICI). Su área de experticia son los sistemas de control automático, teoría y aplicaciones. La Dra. Claudia Pons conduce, desde hace varios años, un grupo de investigación dedicado al estudio de metodologías de desarrollo de software basado en modelos y métodos formales. Actualmente dirige un centro de investigación en la UAI, especializado en robótica (ver <http://www.caeti.uai.edu.ar>). MSc. Gonzalo Zabala es el director del Laboratorio de Robótica de la Universidad Abierta Interamericana (UAI), ver <http://tecnodacta.com.ar/gira/>. Es el creador de Physical Etoys y participa activamente

en el programa "Conectar Igualdad" (<http://www.conectarigualdad.gob.ar>), integrando su Consejo Asesor. Gonzalo Zabala y su grupo tienen amplia experiencia en desarrollo de proyectos educativos utilizando robots. La Dra. Gabriela Arévalo es experta en el campo de la ingeniería de software y específicamente en las metodologías de refactorización/reingeniería. Ella trabaja en el CONICET y contribuye a los proyectos de investigación en la UAI.

El resto del equipo está integrado por estudiantes de pregrado y postgrado que trabajan en la robótica y en MDD. Tres tesis de licenciatura en Informática y dos tesis doctorales se están desarrollando en el contexto de este proyecto.

## REFERENCIAS

- [1] Programa Conectar Igualdad. <http://www.conectarigualdad.gob.ar/> (accedido en May 2011)
- [2] Physical Etoys. GIRA Grupo de Investigación en Robótica Autónoma del CAETI. <http://tecnodacta.com.ar/gira/> (accedido en May 2011)
- [3] Bruyninckx, H.: Open robot control software: The OROCOS project. In: Proceedings of 2001 IEEE International Conference on Robotics and Automation (ICRA'01), vol. 3, pp. 2523–2528. Seoul, Korea (2001)
- [4] Ando, N., Suehiro, T., Kitagaki, K., Kotoku, T., Yoon, W.K.: RT-middleware: Distributed component middleware for RT (robot technology). In: International Conference on Intelligent Robots and Systems 2005 (IROS 2005), pp. 3933–3938 (2005)
- [5] Clemens Szyperski Component Software: Beyond Object-Oriented Programming. 2nd ed. Addison-Wesley Professional, Boston ISBN 0-201-74572-0 (2002).
- [6] Microsoft, "Microsoft robotics developer studio," 2009, <http://msdn.microsoft.com/en-us/robotics/default.aspx>, visited on March 11th 2009.
- [7] Barner, S., Geisinger, M., Buckl, C., Knoll, A.: EasyLab: Model-based development of software for mechatronic systems. In: IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications. Beijing, China (2008)
- [8] Brooks, A., Kaupp, T., Makarenko, A., Oreback, A., Williams, S.: Towards component-based robotics. In: Proc. of 2005 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS'05), pp. 163–168. Alberta, Canada (2005)
- [9] Nesnas, I., Wright, A., Bajracharya, M., Simmons, R., Estlin, T.: CLARAty and challenges of developing interoperable robotic software. In: Proceedings of 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003), vol. 3, pp. 2428–2435 (2003).
- [10] Gerkey, B.P., Vaughan, R.T., Howard, A.: Most valuable player: a robot device server for distributed control. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 1226–1231. Wailea, Hawaii (2001) Player Stage
- [11] C. Schlegel, "Communication patterns as key towards component interoperability," in Software Engineering for Experimental Robotics (Series STAR, vol. 30), D. Brugali, Ed. Berlin, Heidelberg: Springer-Verlag, pp. 183–210. Smartsoftware (2007)
- [12] Bell, Michael. "Introduction to Service-Oriented Modeling". Service-Oriented Modeling: Service Analysis, Design, and Architecture. Wiley & Sons. pp. 3. ISBN 978-0-470-14111-3. (2008).
- [13] Bell, Michael. SOA Modeling Patterns for Service-Oriented Discovery and Analysis. Wiley & Sons. pp. 390. ISBN 978-0470481974. (2010).
- [14] Steven Kelly, Juha-Pekka Tolvanen. Domain-Specific Modeling. John Wiley & Sons, Inc. 2008.
- [15] Stahl, M Voelter. Model Driven Software Development. John Wiley. (2006).
- [16] Claudia Pons, Roxana Giandini, Gabriela Pérez. "Desarrollo de Software Dirigido por Modelos. Teorías, Metodologías y Herramientas", Editorial: McGraw-Hill Education. (2010).
- [17] Jacobson, Ivar; Grady Booch; James Rumbaugh. The Unified Software Development Process. Addison Wesley Longman. ISBN 0-201-57169-2. (1998).
- [18] Christian Schlegel, Thomas Hafler, Alex Lotz and Andreas Steck. Robotic Software Systems: From Code-Driven to Model-Driven Designs. In procs. Of ICAR 2009. International Conference on Advanced Robotics. IEEE Press (2009)
- [19] Iborra, A., Alonso, D., Cáceres, F. Ortiz, J., Sanchez Palma, P. and Alvarez, B.. Design of Service Robots. Experiences Using Software Engineering. IEEE Robotics & Automation Magazine 1070-9932/09/ IEEE MARCH 2009
- [20] Barner, S., Geisinger, M., Buckl, C., Knoll, A.: EasyLab: Model-based development of software for mechatronic systems. In: IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications. Beijing, China (2008)
- [21] A. Brooks, T. Kaupp, A. Makarenko, S. Williams, and A. Oreback, "Towards component-based robotics" in Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems, 2005, vol. 1, pp. 163–168. (2005)
- [22] Brooks, R., "A robust layered control system for a mobile robot". Robotics and Automation, IEEE Journal, 1986.
- [23] "Butiá Project", <http://www.fing.edu.uy/inco/proyectos/butia/>, visited on August 2011
- [24] Murphy R. R., An Introduction to AI Robotics (Intelligent Robotics and Autonomous Agents), Chapters 3-6, MIT Press, 2000.
- [25] Claudia Pons, R. Giandini, G. Arévalo (2012). "A systematic review of applying modern software engineering techniques to developing robotic systems". Vol. 32 No. 1. Revista Ingeniería e Investigación. Categoría A1 en el IBN-Publiindex. Index ISI ScienceCitationIndexExpanded.
- [26] C. Pons, R. Giandini, G. Arévalo, Dimitris Karagiannis: "Exploring Software Engineering Techniques for Developing Robotic Systems". Computer Science & Technology Series. Editorial de la Universidad Nacional de La Plata (EDULP). Octubre 2012.
- [27] Applying Software Engineering Techniques to the Development of Robotic Systems. C. Pons, G. Arevalo, G. Zabala, R. Morán. Congreso Argentino de Ciencias de la Computación CACIC 2011. La Plata. Octubre 2011.



## INGENIERÍA DE SOFTWARE EN CRECIMIENTO: ENFOQUE CONCEPTUAL Y METODOLÓGICO, Y HERRAMIENTAS

Vilallonga, Gabriel<sup>1,2</sup>; Beron, Mario<sup>2</sup>; Sanchez, Alejandro<sup>2</sup>; Moreno, Juan Pablo<sup>1</sup>; Acosta Parra, Carlos<sup>1</sup>.

<sup>1</sup>Departamento Sistemas/Facultad de Tecnología y Ciencias Aplicadas/Universidad Nacional de Catamarca  
Maximio Victoria N° 55 - C.P: 4700 - San Fernando del Valle de Catamarca  
Telefono: 03834- 435112 – int 168

<sup>2</sup>Departamento de Informática de la Facultad de Ciencias Físico, Matemáticas y Naturales de la Universidad Nacional de San Luis. Ejercito de los Andes 950. San Luis  
Tel: +54 (0266) 4520300. Int. 2127

[gvilallo@tecno.unca.edu.ar](mailto:gvilallo@tecno.unca.edu.ar), [unsl.edu.ar](mailto:unsl.edu.ar), [mberon@unsl.edu.ar](mailto:mberon@unsl.edu.ar), [asanchez@unsl.edu.ar](mailto:asanchez@unsl.edu.ar), [jpmoreno@tecno.unca.edu.ar](mailto:jpmoreno@tecno.unca.edu.ar), [carlosacostap@tecno.unca.edu.ar](mailto:carlosacostap@tecno.unca.edu.ar)

### Resumen

Las temáticas abordadas por la Ingeniería de Software (IS) abarca problemas que van desde la gestión del proyecto de desarrollo de software, hasta el mantenimiento del producto resultante. Los desafíos de la IS se encuentran en los entornos de ejecución heterogéneos y de diferentes plataformas, los tiempos de entrega de los sistemas, y la continua evolución hacia sistemas de computación extremadamente grandes, altamente dinámicos y heterogéneos que desafían los fundamentos, modelos y herramientas de la ingeniería de software. El objetivo del proyecto es el estudio de los retos de la IS y su ámbito de aplicación. Se plantea como hipótesis de trabajo la necesidad de retroalimentación entre la actividad de investigación y el ámbito real de aplicación de desarrollo de software, empresas y organizaciones gubernamentales. Desde esta perspectiva, el marco de trabajo propuesto se centra en las experiencias que, de ser pertinentes, propongan modificaciones a las prácticas y, por efecto sinérgico, a los conceptos que la sustentan. Esto resultará en la revisión de técnicas, metodologías, y el desarrollo de herramientas que asistan al desarrollo de sistemas de software.

**Palabras clave:** Ingeniería de Software  
Arquitectura de Software, Patrones de Diseño de Software, Ingeniería Reversa, Gestión de Proyectos de Software.

### Contexto

El proyecto de investigación: “Ingeniería de software en crecimiento: enfoque conceptual y metodológico, y herramientas”, es evaluado por la Secretaría de Ciencia y Tecnología de la Universidad Nacional de Catamarca y se desarrolla y ejecuta en ámbitos del Departamento de Informática de la Facultad de Tecnología y Ciencias Aplicadas de la UNCa y en el Departamento de Informática de la Facultad de Ciencias Físico, Matemáticas y Naturales de la Universidad Nacional de San Luis y, donde los integrantes se desempeñan como docentes de la carrera de Ingeniería en Informática (UNCa), e Ingeniería en Informática y Licenciatura en Ciencias de la Computación (UNSL) respectivamente.

El proyecto establece como prioridad la formación científico-técnica de los integrantes con la premisa de proceder a la transferencia de resultados a los distintos actores sociales con especial énfasis en los alumnos de las carreras a fines.

### Introducción

Como es de conocimiento general, la Ingeniería de Software (IS) abarca problemas que van desde la gestión del proyecto hasta el mantenimiento del producto resultante.

Entre los ítems relevantes de la IS analizados por los integrantes del proyecto se destacan la movilidad de infraestructuras, arquitecturas y el dinámica de cambio en los procesos de negocios, tamaños de los sistemas, por



enumerar algunos de los desafíos a los que se enfrenta la IS.

Otro ítem importante es el desarrollo de aplicaciones de software complejas y de grandes dimensiones. Esta, no solo cubre aspectos técnicos de la construcción de sistemas de software, sino también que abarca el estudio de problemas de administración, tales como la dirección de equipos de programación, tareas de planificación y de presupuestado, en términos generales la gestión integral del proyecto de desarrollo de sistemas de software.

Otro punto importante es el desarrollo de software en la actualidad plantea cambios continuos y crecientes, tanto tecnológicos como en cambios de paradigmas, que repercuten en el seno de los equipos de desarrollo de software. Estos, a menudo, están físicamente distribuidos, quizás móviles, tienen habilidades muy variadas, y lenguajes distintos. Es aquí donde la ingeniería de software debe dirigir esos cambios a través del desarrollo y refinamiento de nuevas técnicas, prácticas y herramientas que tienden a tener fuerte base formal. Una de las tareas a llevar a cabo es acercar los desarrollos formales a actores de las tareas involucradas en la construcción de software sin necesidad que estas personas estén altamente capacitadas en la temáticas formales. El desarrollo de herramientas amigables, con interfaces adecuadas y sistemas formales subyacentes, es uno de los objetivos a cumplir por los investigadores en IS.

Otro de los ítems que despierta real interés en la IS es el referido a la actualización de procesos, encrucijadas en la que se encuentran distintas empresas, E-Business, para mantener su competitividad en el contexto global que determinan las reglas de mercado actuales. En la misma línea de la automatización de procesos se encuentran los organismos gubernamentales, E-Government, para realizar un manejo eficiente del volumen de trámites que se procesan en su esfera. En estos dos ámbitos, es posible reconocer procesos con características similares que permiten encuadrarlo en un determinado patrón de negocio. El estudio de este ítem de la

ingeniería aporta la herramienta necesaria para el desarrollo de esos procesos a partir de casos ya estudiados. En esta área se ha trabajado en forma conjunta entre el Laboratorio de Tecnologías de la Información y las Comunicaciones (LaTICs) de la FTyCA de la UNCa y el Ministerio de Salud de la Provincia de Catamarca en la aplicación de reingeniería y desarrollo de sistemas en el seno de esta repartición pública.

En este punto es donde se ha hecho visible una de los numerosos casos de desarrollo de sistemas donde ya existen sistemas con lo que se debe interactuar. Son pocos los casos un ingeniero de software se encuentra con un sistema a desarrollar de cero. Este escenario implica aplicación de reingeniería en donde deberán coexistir diferentes aplicaciones que han sido desarrolladas con diferentes tecnologías, en algunos casos, obsoletas, donde la adaptación de los sistemas existentes es una de las tareas. Ámbito en donde juega un papel fundamental la ingeniería reversa, comprensión de código y arquitecturas de software.

Otro de los tópicos de gran importancia en el desarrollo de software es el análisis y verificación a que deben ser sometidos el software desarrollado para cumplir con normas básicas de calidad. En sistemas altamente es imperioso la utilización de métodos formales permiten el estudio de los sistemas para garantizar el correcto desempeño de estos. Se prevé el estudio de técnicas de model checking y la factibilidad de incorporación en herramientas de modelado de sistemas.

Los resultados en estas áreas pueden mejorar el proceso de formación académica, de docentes y alumnos en el desarrollo de aptitudes y de valores, contribuir en la toma de decisiones tácticas y estratégicas de la Facultad, proporcionando un generador de conocimiento y elevando de esta manera la calidad de la educación que favorezca la formación de ingenieros en la Universidad.

### **Líneas de investigación y desarrollo**

Basados en la necesidad de retroalimentación entre la actividad de investigación y el ámbito real de aplicación de desarrollo de software,

se propone el estudio de casos reales de desarrollo de software, técnicas, métodos y herramientas utilizadas y a partir de estos proceder a realizar un riguroso estudio centrado en las prácticas de IS. De estos se prevé obtener información valiosa para realizar aportes concretos a la práctica realizada desde la visión de la IS moderna.

La metodología de trabajo planteada anteriormente se llevará a cabo con mayor énfasis, sin descartar otras áreas específicas, en las siguientes temáticas:

- Gestión de Proyectos de Desarrollo de Software,
- Patrones de Negocios,
- Reingeniería de Procesos,
- Ingeniería de Sistemas de Tiempo Real,
- Arquitecturas de Software.

Estas temáticas son las áreas de mayor interés sobre las cuales se trabaja.

### **Resultados y Objetivos**

El grupo de trabajo, en diversas reuniones, ha realizado diferentes análisis que han dado como resultado el establecimiento de objetivos centrales y particulares que en un posterior análisis permitirán sacar conclusiones acerca de resultados del presente proyecto.

El objetivo principal del proyecto consensuado es el de contribuir de manera efectiva y mensurable en la concreción de aportes concretos a la IS aplicada, lo que implica revisión, o nuevas propuestas, de técnicas, metodologías, y el desarrollo de herramientas que asistan al desarrollo de software; e incidir significativamente en las actividades académicas de grado y post grado, la formación de recursos humanos, la transferencia al medio social y productivo del "know how" y habilidades, revisión conceptos, metodologías, técnicas y herramientas de desarrollo de software.

El marco de trabajo propuesto se centra en las experiencias basadas en la interacción entre la actividad científica y proyectos concretos de desarrollo de software, que, de ser pertinentes, propongan modificaciones a las prácticas como así también la revisión de conceptos que

la sustentan.

Estos objetivos están siendo alcanzados gracias a la firma de distintas actas acuerdo realizadas bajo el convenio establecido entre la FTyCA de la UNCa y el Ministerio de Salud de la Provincia de Catamarca. En el Acta Acuerdo N° 1 se ha acordado la realización de actividades encuadradas en "Relevamiento, Reingeniería de Procesos y Captura de requisitos para la Gestión de Turnos en las dependencias del Ministerio de Salud de Catamarca", la cual ha finalizado con éxito. Esta última actividad ha derivado en la firma de una nueva Acta Acuerdo que se encuentra en desarrollo, Acta Acuerdo N°2 "Desarrollo y Aseguramiento de Calidad de la solución informática para la Gestión"

Estas actividades de extensión y transferencia se instrumentan por medio del LaTICs que coordina las actividades de investigación de la carrera de Ingeniería en Informática de la FTyCA de la UNCa.

Estas actividades han exigido realización de cursos acerca de la temática, como así también el estudio de material bibliográfico y de publicaciones científicas.

Es de hacer notar que estos grupos de trabajos están coordinados por docentes investigadores de la carrera y donde las actividades de campo son llevadas a cabo en conjunto entre docentes y alumnos de los últimos años, próximos a recibir el grado de ingenieros.

### **Formación de Recursos Humanos**

El proyecto cuenta con integrantes docentes que se encuentran en la etapa de desarrollo de sus tesis de posgrado en el marco de la carrera de maestría en Ingeniería del Software en temas específicamente relacionados al área del proyecto. Además los participantes pertenecen a distintas cátedras de las carreras de ingeniería y licenciatura de las universidades que participan.

Se ha procedido a la incorporación de alumnos de los últimos años con el finalidad de incorporarlos a actividades de investigación y desarrollo en las áreas específicas del proyecto

Específicamente el proyecto de investigación prevé el programa de capacitación y formación

de recursos humanos, que contempla las siguientes actividades:

- Incorporación de alumnos de los últimos años de la carrera de Ingeniería en Informática de la Facultad de Tecnología y Ciencias Aplicadas de la UNCa., en calidad de auxiliares de investigación.
- Dirección de tesinas de grado de la carrera de Ingeniería en Informática de la Facultad de Tecnología y Ciencias Aplicadas de la UNCa integrados al proyecto.
- Participación de los integrantes del proyecto en cursos de actualización y posgrado en el área de estudio.

Para garantizar la capacitación y actualización del equipo de investigación, así como la difusión de los avances y resultados logrados, se propuso la participación en eventos nacionales e internacionales de la especialidad, como congresos, simposios, seminarios y cursos.

### **Bibliografía**

“Software Engineering (9th Edition)” Ian Sommerville. Addison-Wesley; 9 edition (March 13, 2010). ISBN-10: 0137035152

“Software Engineering: A Practitioner's Approach” Roger Pressman : McGraw-Hill Science/Engineering/Math; 7 edition (January 20, 2009). ISBN-10: 0073375977.

“Software Architecture in Practice” (3rd Edition) (SEI Series in Software Engineering) by Len Bass, Paul Clements and Rick Kazman (Oct 5, 2012) Addison-Wesley Professional; 3 edition (October 5, 2012). ISBN-10: 0321815734.

“Software Management (Practitioners)”. Donald J. Reifer, Barry Boehm. Wiley-IEEE Computer Society Pr; 7 edition (August 30, 2006). ISBN-10: 0471775622

“Model Driven Architecture for Reverse Engineering Technologies: Strategic Directions and System Evolution”. Liliana Favre. IGI Global; 1 edition (January 31, 2010) ISBN-10: 1615206493.

“Real-Time Systems Design and Analysis: Tools for the Practitioner” Phillip A. Laplante, Seppo J. Ovaska. Wiley-IEEE Press; 4 edition (November 22, 2011) ISBN-10: 0470768649

## Un acercamiento en la integración entre BPMN y SOA

Diana Cruz<sup>1</sup>, Juan Fontana<sup>1</sup>, Silvia Rivadeneira<sup>1</sup>, Gabriela Vilanova<sup>2</sup>  
Departamento Ciencias Exactas y Naturales

<sup>1</sup>Unidad Académica Río Turbio, <sup>2</sup>Unidad Académica Caleta Olivia  
Universidad Nacional de la Patagonia Austral

[dianalrcruz@gmail.com](mailto:dianalrcruz@gmail.com), [jefontana30@yahoo.com.ar](mailto:jefontana30@yahoo.com.ar), [grivadeneira@uart.unpa.edu.ar](mailto:grivadeneira@uart.unpa.edu.ar),  
[vilanova@uolsinectis.com.ar](mailto:vilanova@uolsinectis.com.ar)

### Resumen

La automatización de los sistemas de información de las organizaciones presenta el desafío de asegurar que dichos sistemas respondan a los requerimientos cambiantes del entorno organizacional, tecnológico y de mercado, más aún, cuando se trata de organizaciones y sistemas complejos, que requieren de respuestas prácticamente inmediatas para seguir siendo competitivas.

Con el enfoque orientado a servicios se intenta integrar aplicaciones que permitan el reuso de componentes internos o externos a la organización (servicios web) para facilitar la comunicación de diferentes tipos de aplicaciones que se ejecuten en distintos tipos dispositivos y con el enfoque orientado a procesos permite que las organizaciones mejoren el rendimiento optimizando sus procesos para que sean flexibles y adaptables al entorno dinámico actual.

Al mismo tiempo, se están implementando sistemas de información basados en SOA, donde BPM juega el papel de orquestador de servicios, como es el caso de BPMS, pero que requieren ser evaluadas para determinar si la arquitectura es adecuada a la organización o al tipo de aplicación, etc., por tal motivo, nuestro trabajo pretende analizar los soportes tecnológicos y los marcos metodológicos que podrían aplicarse a los tipos de organizaciones que podemos encontrar en la Patagonia Austral, usando la arquitectura SOA.

**Palabras clave:** SOA, BPMN, Web Services, modelado de requerimientos, arquitectura de software

### Contexto

El presente artículo es una producción del Proyecto tipo 1 de Ciencia y Tecnología 29/B134 apoyo a formación de grado denominado “Modelado de Requerimientos y Diseño de Sistemas Complejos” financiado por la Universidad Nacional de la Patagonia Austral, iniciado en 2012.

La Universidad Nacional de la Patagonia Austral (UNPA) está ubicada en la provincia de Santa Cruz, está compuesta por cuatro Unidades Académicas dispuestas en las localidades de: Caleta Olivia, San Julián, Río Gallegos y Río Turbio.

El proyecto se encuentra radicado en la Unidad Académica Caleta Olivia (UACO), con la particularidad de que sus integrantes residen en varias localidades patagónicas, tales como Comodoro Rivadavia, Caleta Olivia, Río Turbio y Punta Arenas (Chile).

### Introducción

Las empresas necesitan poder interconectar sus procesos, personas e información tanto en la propia organización como con subsidiarias o socios comerciales. La inexistente integración entre los componentes de IT –sistemas, aplicaciones y datos- hace difícil obtener una respuesta rápida y efectiva ante los cambios que afectan a los negocios. La inflexibilidad genera costos, reduce la capacidad de respuesta ante los clientes, compromete el cumplimiento con las reglamentaciones vigentes, y afecta negativamente al personal de la empresa [10].

Ante el crecimiento de la complejidad de los sistemas, ya sea por el número de requerimientos o por el impacto de los mismos, se hace necesario establecer medios para el manejo de esa complejidad. En la



literatura base sobre diseño de sistemas, la mayoría de los autores coinciden en que una arquitectura de software define la estructura del sistema. La arquitectura de software juega un papel fundamental en el desarrollo a gran escala de software de calidad. Los ingenieros de software generalmente se refieren a la arquitectura de software como *la organización a nivel de un sistema de software integrado de los componentes, las relaciones entre ellos y las restricciones* [11,12]. Además, la arquitectura de un sistema es responsable de la captura de las abstracciones de la arquitectura esencial para garantizar un conjunto relevante de los factores de calidad. [13]

#### **a. Modelado de elicitación de requerimientos**

Sin dudas la elicitación de requerimientos ayuda a describir el problema que motivará a diseñar una solución de calidad. Es por ello que se debe garantizar que los requerimientos queden fuertemente alineados con las metas organizacionales, a partir de una notación de modelado de los procesos de negocio (BPMN) que involucren aquellas actividades que se deseen automatizar [1].

En sistemas complejos donde la probabilidad de cambios en las especificaciones funcionales es muy elevada, una opción efectiva ante este escenario es la combinación de BPMN con SOA (Service Oriented Architecture) para flexibilizar los cambios y reutilizar los componentes de procesos de negocio como servicios [2].

#### **b. Importancia del diseño arquitectónico**

El diseño arquitectural puede ser visto como un proceso de toma de decisiones, ya que implica tomar las decisiones correctas en el momento adecuado, con el fin de reducir riesgos, costos de desarrollo y mantenimiento elevando los niveles de productividad. Normalmente, estas decisiones de diseño no están explícitamente representadas en los artefactos que describen el diseño, sino que residen en la mente de los diseñadores. Una propuesta de solución

es la gestión del conocimiento "*Knowledge management*".

El enfoque de desarrollo de software orientado a servicios devino a un popular paradigma de arquitectura para el diseño y el desarrollo de sistemas distribuidos porque ofrece estandarización, independencia de plataforma, interfaces bien definidas, y herramientas de apoyo que habilitan a la integración del sistema.

La adopción de SOA requiere un enfoque en el que se reflejen un fuerte vínculo entre los objetivos de negocio y la estrategia de desarrollo [4] [10]. Esta arquitectura separa los procesos de negocio de las funciones automatizadas y las organiza como módulos catalogados en un diccionario de servicios que permiten su utilización por parte de toda la organización. Esta arquitectura ofrece la capacidad de añadir, modificar y optimizar los procesos de negocios [5].

La clave de esta arquitectura es la abstracción de los procesos, permitiendo que estos puedan ser transparentes y manejados a nivel negocio y de forma versátil. Así desde el punto de vista del negocio, SOA ayuda a resolver los siguientes requerimientos [6]:

- Mejorar la flexibilidad y agilidad de los sistemas.
- Proporcionar una visión integrada de los distintos "silos" de la organización.
- Mejorar la cobertura de las necesidades de negocio.
- Reducir el impacto de la evolución de la tecnología en las aplicaciones de negocio.

Para poder integrar el modelo de negocio con la arquitectura SOA, existe una forma de traspaso automática del modelo obtenido de BPMN mediante el uso de un framework denominado WS-BPEL, que es una tecnología de web service [5].

#### **c. Hacia una arquitectura orientada a servicios**

SOA es un tipo de arquitectura de software basado en servicios reutilizables, con interfaces públicas definidas, donde los proveedores y consumidores de servicios interactúan para realizar los procesos de

negocio. Se basa en cuatro abstracciones básicas: servicios, aplicación frontend, repositorio de servicios y bus de servicios. Un servicio consiste en una implementación que provee lógica de negocios y datos, un contrato de servicio, las restricciones al consumidor, y una interfaz que expone físicamente la funcionalidad. La aplicación frontend consume los servicios como procesos de negocios. Un repositorio de servicios almacena los contratos de servicios y el bus de servicios interconecta los application frontend y los servicios. [14]

Los problemas de heterogeneidad, interoperabilidad y requerimientos cambiantes, son solucionados con SOA que provee una plataforma para la construcción de aplicaciones basadas en servicios con las características de bajo acoplamiento, ubicación transparente de servicios e independencia de protocolos. Un consumidor de servicios no debe preocuparse por un servicio en particular con el que comunicarse debido a que la infraestructura por debajo puede hacer la elección en versión del consumidor.

### Líneas de investigación y desarrollo

En el marco de lo expuesto, desde un punto de vista general, nuestro estudio apunta al análisis de requerimientos y el diseño de software con un enfoque arquitectónico, evaluando los atributos de calidad que se deben tener en cuenta al adoptar una arquitectura de diseño en sistemas complejos.

Desde un punto de vista específico, se apunta a tres grandes líneas:

- a) el modelado de requerimientos con dos posibles enfoques combinando BPMN y UML
- b) el modelado de requerimientos con metodologías ágiles.
- c) la arquitectura de software con la integración de BPMN y SOA.

### Resultados y objetivos

Nos encontramos en pleno proceso de análisis de la literatura existente para conocer experiencias académicas, procesos y herramientas relacionadas con estas

temáticas tanto en Argentina como en el exterior.

El trabajo final de carrera de pre grado de uno de los alumnos integrantes del proyecto está orientándose a esta temática. Por otro lado se realizan aportes de la temática de proyecto en las asignaturas del área Ingeniería de software tales como, Ingeniería de requerimientos, Análisis y diseño de software, Arquitectura de software mediante la realización de trabajos finales de cátedra, trabajos de campo e informes técnicos. Se toman como base para modelado casos reales de la región en área turismo, energías alternativas y minería.

### Formación de Recursos Humanos

El proyecto actualmente cuenta con ocho integrantes de los cuales hay docentes cursando la Maestría de Informática y Sistemas – uno avanzado y otro iniciando -, dos de la Maestría en Educación en Entornos Virtuales – uno de ellos en trabajo de tesis -, un alumno en proceso de trabajo final en la Especialización en Management Tecnológico, dos alumnos de pre-grado de Analista de Sistemas en proceso de realización de trabajo final de carrera y, un académico perteneciente a la Universidad de Magallanes doctorado en el área de computación y automatización industrial por la Universidad Estatal de Campiñas (UNICAMP), quien forma docentes y alumnos en la Unidad Académica Río Turbio (UART).

### Referencias

- [1] Montserrat, C., Páez, A., Arias, C., Rivadeneira, S., Vilanova, G., Miranda, M. “El modelado de procesos como técnica de elicitación de requerimientos”. II EIPA. Santa Cruz. Argentina. (2012).
- [2] Arias, C., Vilanova, G, Rivadeneira, S., Miranda, M., Cruz, D. y Fontana, J. “Construcción de modelos de requerimientos a partir de modelos de procesos de negocio”. WBPM. JCC 2012. Valparaíso. Chile. (2012).
- [3] Lübke, D., Schneider, K. and Weidlich, M. “Visualizing Use Case

- Sets as BPMN Processes”. REV'08. IEEE. (2008).
- [4] Bianco, P., Kotermanski, R., Merson. P. “Evaluating a Service-Oriented Architecture”. Technical Report CMU/SEI-2007-TR-015. (2007).
- [5] Zhang L., Chen, S., Ge M., Bi, X. “A SOA-BPM\_Based Architecture for Intelligent Power Dispatching System”. IEEE. (2010)
- [6] “Arquitectura Orientada a Servicios (SOA): Cómo reformular la Arquitectura Corporativa para alcanzar el alto rendimiento”. Accenture. (2008)
- [7] Levinson, H., O'Brien, L. “Acquiring Evolving Technologies: Web Services Standards”. Technical Note CMU/SEI-2006-TN-001. (2006)
- [8] “Ultra-Large-Scale Systems. The Software Challenge of The Future”. SEI-CMU. (2006)
- [9] Erl, T. “Service Oriented Architecture: Concepts, Technology and Design”. Prentice Hall. (2005)
- [10] Microsoft Corporation, “La Arquitectura Orientada a Servicios (SOA) de Microsoft aplicada al mundo real”, Whitepaper, Diciembre de 2006.
- [11] L. Bass, P. Clements and R. Kazman. *Software Architecture in Practice*. Addison Wesley, 1099, ISBN 0-201-19930-0.
- [12] D. Garlan and M. Shaw. An Introduction to software architecture. In *Advances in Software Engineering and Knowledge Engineering*, pages 1-39, Singapore, 1993. World Scientific Publishing Company.
- [13] B. Boehm, P. Bose, E. Horowitz and M. J. Lee. Software requirements negotiation and renegotiation aids: A theory-W based spiral approach. In *Proc 17th International Conference on Software Engineering*, 1994.
- [14] Krafzig D., Banke, K., Dirck Slama, “Enterprise SOA, Service Oriented Architecture Best Practices” Prentice Hall Pearson Education.

# Metodología de Modelado de Aplicaciones Web Móviles Basada en Componentes

Pablo Vera, Claudia Pons, Carina González,  
Daniel Giulianelli, Rocío Rodríguez

## Universidad Nacional de La Matanza

Departamento de Ingeniería e Investigaciones Tecnológicas  
Florencio Varela 1903, San Justo, Buenos Aires, Argentina. Tel: 4480-8900 (int 8630)

## Universidad Nacional de la Plata

LIFIA – Laboratorio de Investigación y Formación en Informática Avanzada  
Calle 50 y 150, La Plata, Buenos Aires, Argentina. Tel: 221-422-8252

## Universidad de la Laguna

Departamento de Ingeniería de Sistemas y Automática y Arquitectura y Tecnología de  
Computadores. Área de Arquitectura y Tecnología de Computadores  
La Laguna, España. Tel: 922318284

pvera@ing.unlam.edu.ar, cpons@lifia.info.unlp.edu.ar, cjgonza@ull.es,  
dgiulian@ing.unlam.edu.ar, rrodriguez@ing.unlam.edu.ar

## Resumen

Esta línea de investigación busca crear una metodología de modelado utilizando una extensión conservativa de UML (Lenguaje Unificado de Modelado) que permita diseñar aplicaciones web móviles centrándose en el modelado de la interfaz de usuario mediante la utilización de componente configurables. El objetivo final es la generación automática del código fuente completo de la aplicación. Para lograr tal fin se extienden los diagramas de clases y componentes de UML. El diagrama de clases es extendido para poder generar la base de datos y sus relaciones. El diagrama de componentes es extendido mediante valores etiquetados que configuran cada componente permitiendo especificar su funcionalidad de forma clara y concisa pero a su vez conteniendo toda la información necesaria y suficiente para permitir generar una aplicación completa y funcional. Esta metodología se enmarca en el ámbito de MDA

(Arquitectura Dirigida por Modelos) y plantea una primera transformación de modelo a modelo para facilitar la tarea del diseñado, luego una segunda transformación de los modelos al código fuente para obtener la aplicación completa.

**Palabras clave:** MDA, UML, Hipermedia Móvil, XMI, MVC

## Contexto

Este proyecto es financiado por la Universidad Nacional de La Matanza, cuenta con asesoría y colaboración de la Universidad Nacional de La Plata y la Universidad de La Laguna (España). Se corresponde a una línea de investigación y desarrollo dentro del área de MDA.

## Introducción

El modelado de aplicaciones es un área muchas veces subestimada por la industria donde no se le da importancia, considerándose simplemente como una



pérdida de tiempo. En otros casos sólo se utiliza en etapas tempranas del desarrollo para hacer una primera definición del problema y obtener los requerimientos. Gran parte de los modelos realizados en estas etapas luego no son actualizados con los cambios realizados en la etapa de desarrollo haciendo que la documentación del sistema quede obsoleta. El modelado a bajo nivel de una aplicación también es una tarea ardua que agrega costos y tiempos que muchas empresas no están dispuestas a afrontar. Para subsanar todos estos problemas nace la arquitectura dirigida por modelos (MDA) [1] donde los modelos van evolucionando y transformándose hasta llegar a generar el código fuente o parte del mismo en forma automática. Si bien actualmente existen herramientas basadas en esta arquitectura que generan código fuente, ninguna de ellas permite generar una aplicación completa 100% funcional, e incluso el esfuerzo necesario para realizar un modelado detallado que permita lograr dicha meta es demasiado elevado.

Por este motivo se plantea la necesidad de generar una nueva metodología de modelado que esté orientada a la generación del código fuente y que a su vez permita a los analistas y arquitectos definir de una forma sencilla el comportamiento del sistema, sus pantallas y la información que el usuario deberá ver y manejar.

## Modelado de Aplicaciones web móviles

La metodología se plantea para el modelado de aplicaciones web móviles por las siguientes razones:

- **Interfaz reducida:** Las pantallas de los dispositivos móviles son pequeñas, más allá de la resolución que posean el tamaño es reducido, y por lo tanto la interfaz presentada al usuario debe ser simple, cómoda y

adecuarse al método de utilización. No es recomendable llenar de información una pantalla pequeña y tampoco es recomendable obligar al usuario a tener que realizar demasiado scroll para ver el contenido. Como esta metodología está basada en componentes, cada componente va a representar una pantalla que se mostrará al usuario. Por lo tanto cada componente está diseñado de forma que contenga la información mínima necesaria para cumplir su funcionalidad sin sobrecargar de información al usuario.

- **Sistema de navegación simple e intuitivo:** Al tener una pantalla reducida el sistema de navegación también debe minimizarse. Siguiendo las pautas del W3C (World Wide Web Consortium) sobre sitios web móviles [2]. Se incorpora a cada componente una barra de navegación reducida que será mostrada en la parte superior de la pantalla.
- **Aprovechamiento de características especiales:** Se incorporan ciertos elementos que permiten aprovechar las características de algunos dispositivos móviles. Por ejemplo el uso de la geolocalización mediante el GPS del dispositivo si está presente o la posibilidad de incorporar links especiales para realizar llamadas o enviar SMS si se está mostrando un número telefónico.

Sin embargo esta metodología puede aplicarse también a aplicaciones web tradicionales donde no se requieran layouts de pantallas demasiados complejos.

## Metodología de Modelado

El modelado de sistemas hipertexto es una práctica que data de varios años atrás cuando la web comenzaba a popularizarse. Uno de los trabajos más relevantes del área es OOHD (Object

Oriented Hypermedia Design Method) [3] que a su vez está basado en un trabajo previo llamado HDM (Hypermedia Design Method) [4]. OOHDM establece cuatro actividades principales para el diseño hipermedia. Estas actividades son: modelado conceptual, modelado de la navegación, diseño de la interfaz abstracta e implementación.

La metodología propuesta respeta estas actividades pero unifica el modelado de la navegación y de la interfaz abstracta en un único diagrama. Para llevar a cabo el modelado se toma como base UML y se extienden los diagramas de clases y componentes mediante una extensión conservativa del lenguaje.

A continuación se describen los pasos de la metodología, detallando las acciones del usuario y las transformaciones automáticas realizadas.

1. **Modelo Conceptual:** El usuario confecciona el diagrama de clases UML, especificando ciertas propiedades definidas en la extensión del modelo y además hace uso de tipos de dato especialmente definidos para aplicaciones móviles.
2. **Transformación de Modelo a Modelo:** Basado en el diagrama de clases se generan en forma automática un modelo de Navegación e Interfaz con los componentes básicos para poder administrar la información de la base de datos.
3. **Modelo de Navegación e Interfaz:** el usuario completa y/o modifica el diagrama generado en forma automática en el paso 2 para adaptarlo a la aplicación particular que está modelando.
4. **Modelo de Estados Válidos:** Si la aplicación lo requiere es posible especificar los estados por los que una entidad determinada puede pasar, dando una secuencia a los

mismos. Esto se especifica en un diagrama de estados UML.

5. **Transformación de Modelo a Código:** con los distintos modelos generados se genera en forma automática es script de la base de datos y el código fuente de la aplicación.

## Modelo de Navegación e Interfaz

La base de la metodología es el modelo de Navegación e Interfaz de usuario.

Este diagrama se define mediante una serie de componentes donde luego cada componente se transformará en una página web. Los componentes son controles estándares de la interfaz de usuario adaptados para dispositivos móviles.

Los componentes definidos son:

- **Login:** este componente permite autenticar un usuario en el sistema.
- **List:** listado de información que puede visualizarse como una grilla o tabla con filas y columnas que muestren información de las entidades. Para facilitar la visualización en pantallas reducida se incorpora la posibilidad de mostrar una segunda línea de información así como también el listado de acciones que se mostrarán en un menú contextual al pulsar sobre una opción.
- **Search:** es similar a List pero incorpora la posibilidad de agregar Filtros que serán ingresados por el usuario para acotar la búsqueda
- **Menu:** define un menú con links a otros componentes.
- **CRUD:** es el encargado de mostrar, actualizar, crear o eliminar un objeto. Permite definir valores por defecto para algunas propiedades al crear o actualizar y además permite omitir modificar ciertas propiedades

según si se crea o actualiza el objeto.

- **UpdateView:** operación de actualización con características especiales. Muchas veces la actualización de un objeto se realiza por partes donde algunas de sus propiedades se actualizan y otras no. Es por eso que este componente permite:
  - mostrar algunas propiedades sin modificarlas.
  - definir puntualmente cuales son las propiedades a actualizar.
  - crear entidades relacionadas al momento de actualizar un objeto como por ejemplo grabar una tabla adicional con información de seguimiento.

Todos los componentes incorporan además la configuración de la navegación donde se definen los links que aseguran que el usuario pueda moverse entre las distintas páginas de la aplicación.

Los componentes se modelan con el diagrama de componentes de UML donde se definen estereotipos para poder identificar cada tipo de componente. Luego para poder configurar cada componente y adecuarlo a la aplicación se utilizan una serie de valores etiquetados. Dichos valores están especialmente diseñados para brindar toda la información necesaria para que el componente pueda ser luego transformado en código fuente funcional. Para especificar la forma en que se configuran los componentes y los valores esperados en cada valor etiquetado se definió un lenguaje en formato BNF [5] que especifica las reglas de sintaxis.

Dicho BNF es completado en forma dinámica a partir del modelo conceptual ya que dentro de las reglas de escritura se validan los nombres de las clases y propiedades definidas para la aplicación.

## Herramienta de Modelado

Para brindar soporte a esta metodología se está desarrollando una herramienta que permita realizar el modelado completo, además incluye las transformaciones entre las distintas etapas.

Es una herramienta web que permitirá modelar la aplicación desde cualquier computadora con acceso a internet. También permitirá que varios usuarios puedan trabajar sobre el mismo proyecto. La figura 1 muestra una captura de pantalla correspondiente al menú de opciones de un proyecto particular.



**Figura 1:** Menú de opciones de un proyecto en la herramienta de modelado

La herramienta hace un seguimiento de las modificaciones realizadas en los modelos y da avisos al usuario para que realice las transformaciones necesarias para tener actualizado su proyecto. A modo de ejemplo, puede observarse en la figura 1 que se ha realizado una modificación en las clases del modelo conceptual pero no se han generado los componentes por eso se da una aviso al usuario de que los componentes están desactualizados.

## Líneas de Investigación y

### Desarrollo

Los temas que se desarrollan dentro de esta línea son los siguientes:

- Desarrollo Dirigido por Modelos
- Modelado de hipermedia móvil
- Modelado basado en componentes
- Perfiles de UML
- Generación automática de Código

### Resultados y Objetivos

El objetivo principal de esta línea de Investigación y Desarrollo es definir una metodología de modelado de aplicaciones web móviles que permita la generación de código automático.

Actualmente este proyecto se encuentra en proceso por lo cual ya se han obtenido algunos resultados:

- Se han definido todas las extensiones necesarias para todos los modelos UML.
- Se ha creado y definido el lenguaje de configuración de componentes.
- Se ha comenzado con el desarrollo de la herramienta de modelado y transformaciones.

Como objetivos futuros se planea:

- Formalizar la extensiones de los diagramas mediante un perfil UML
- Finalizar con programación de la herramienta de modelado y transformaciones.
- Agregar interoperabilidad a la herramienta incorporando la posibilidad de importar y exportar modelos al formato XMI [6].

## Formación de Recursos Humanos

En esta línea de investigación participan 6 docentes y 2 alumnos en formación.

Este trabajo tiene relación directa con una tesis Doctoral en desarrollo por uno de los docentes del grupo de investigación

### Referencias

- [1] Kleppe A., Warmer J., Bast W. "MDA explained: the model driven architecture: practice and promise". Addison-Wesley Professional (2003)
- [2] W3C: Mobile Web Best Practices 1.0, <http://www.w3.org/TR/mobile-bp/>
- [3] Schwabe D. y Rossi G. : An object oriented approach to Web-based applications design. Theor. Pract. Object Syst. Volume 4, Issue 4, pp 207-225. (1998)
- [4] Garzotto, D. Schwabe and P. Paolini: HDM - A Model Based Approach to Hypermedia Application Design. ACM Transactions on information Systems, pp. 1-26. (1993)
- [5] J. W. Backus: The Syntax and Semantics of the Proposed International Algebraic Language of the Zurich ACM-GAMM Conference. Proceedings of the International Conference on Information Processing, UNESCO, pp.125-132. (1959)
- [6] OMG, "MOF 2 XMI Mapping", Version 2.4.1, <http://www.omg.org/spec/XMI/>



## Reuso de Software Orientado a Dominios

Agustina Buccella, Juan Luzuriaga, Alejandra Cechich,  
Rodolfo Martínez, Rafaela Mazalu, Marcos Cruz,  
Matias Pol'la, Maximiliano Arias y Adriana Martin

GIISCO Research Group

Departamento de Ciencias de la Computación

Universidad Nacional del Comahue

Neuquen, Argentina

agustina.buccella@fai.uncoma.edu.ar, luzurj@yahoo.com.ar

Maria del Socorro Doldan y Enrique Morsan

Instituto de Biología Marina y Pesquera "Almirante Storni"

Universidad Nacional del Comahue

Ministerio de Producción de Rio Negro

San Antonio Oeste, Argentina

{msdoldan, qmorsan}@gmail.com

### 1 Resumen

Tomando como base los trabajos realizados en investigaciones anteriores, la línea de investigación actual de nuestro proyecto se enfoca en el estudio y uso de un conjunto de metodologías para mejorar el reuso de software y así reducir el tiempo, esfuerzo, costo y complejidad en el desarrollo de los productos. En particular se desea lograr un reuso efectivo orientado a dominios mediante la utilización del paradigma de *Ingeniería de Líneas de Productos de Software* (ILPS), el cual contempla la definición de partes comunes de un conjunto de productos dentro de un dominio específico y al mismo tiempo provee un mecanismo para modelar la variabilidad dentro de los mismos. De esta manera, se crea un conjunto de sistemas que comparten características comunes y ciertas partes variables. Este paradigma es la

base de nuestro proyecto en el cual se desarrollan nuevas técnicas y herramientas orientadas a mejorar el reuso en dos dominios específicos: el dominio geográfico y el de gobierno electrónico.

### 2 Contexto

La línea presentada se inserta en el contexto de los siguientes proyectos y acuerdos de cooperación:

- *Programa: Desarrollo de Software Basado en Reuso.* En proceso de acreditación por la Universidad Nacional del Comahue. Directora: Dra. Alejandra Cechich.

– *Proyecto: Reuso Orientado a Dominios.* En proceso de acreditación

por la Universidad Nacional del Comahue. Directora: Dra. Agustina Buccella, Codirector: Mg. Juan Manuel Luzuriaga.

- *PAE-PICT-2312: Métodos y Herramientas para Sistemas masivamente Distribuidos* Investigaciones conjuntas con ISISTAN-UNICEN
- *Acuerdo de Cooperación* entre el Laboratorio de Investigación en Ecología Bentónica y el Laboratorio de Parasitología e Histopatología de Moluscos del Instituto de Biología Marina y Pesquera Almirante Storni.

### 3 Introducción

La ingeniería de software orientada a reuso intenta mejorar los tiempos de desarrollo, la puesta en el mercado y los costos maximizando al mismo tiempo la calidad del producto resultante. En particular, el reuso orientado a dominios se refiere a la captura del conocimiento específico de un dominio mediante la construcción de artefactos de software. Dichos artefactos serán luego útiles para la construcción de un conjunto de programas o productos aplicables a diferentes organizaciones dentro del mismo dominio. De esta manera, la ingeniería de dominios se basa en la identificación y/o desarrollo de elementos de software reusables. Dentro de ésta área existen un conjunto de metodologías que ayudan a la identificación de dichos elementos y a la construcción de aplicaciones basadas en los mismos. Una de ellas es la Ingeniería de Líneas de Productos de Software [11] ya que permite la definición de las partes comunes de varios productos dentro de un dominio específico y al mismo tiempo permite insertar cierta variabilidad para adaptarse a las necesidades particulares de esos productos. Una de las definiciones más aceptadas en la literatura es la realizada en [11] en donde las líneas de productos se definen como *un conjunto de sistemas que comparten características comunes y*

*satisfacen necesidades específicas de un segmento del mercado.* Las principales características involucradas en esta disciplina son: la *variabilidad* en donde los sistemas individuales se deben considerar como variaciones de una parte común, *basada en arquitectura* en donde el software debe ser desarrollado considerando las similitudes entre sistemas individuales, y la *aproximación de dos ciclos de vida* en donde se deben considerar las dos ingenierías en el proceso de desarrollo - *la ingeniería de dominio*, en la cual se provee la definición de aspectos reusables con un rango de variabilidad y similitudes; y *la ingeniería de aplicación*, en la que se derivan las aplicaciones de la línea de productos a partir de la plataforma definida en la ingeniería del dominio.

Una Arquitectura de Línea de Productos de Software incorpora componentes reusables, definidos dentro de un dominio en particular, de modo que al desarrollar un nuevo producto, en lugar de hacerlo desde el principio, se instancian los componentes necesarios dentro de la arquitectura de LPS ya definida.

Existen diversos dominios en los cuales se pueden aplicar las LPS. Uno de ellos son los Sistemas de Información Geográficos ya que contienen un conjunto de requerimientos y herramientas especiales que permiten su uso en diferentes áreas. En la literatura existen varias definiciones para este concepto, cada una considerando la funcionalidad del sistema desde diferentes perspectivas [2, 4, 12]. De todas ellas, se puede concluir en la definición dada en [5], la cual introduce a *un GIS como un sistema basado en computadora diseñado para modelar, capturar, almacenar, manipular, consultar, recuperar, analizar y visualizar información eficientemente y donde parte de la misma es de naturaleza geográfica.*

Con esta definición se ve claramente que los GIS son más que herramientas para producir mapas. Mientras que en la cartografía tradicional el mapa es la base de datos, en un GIS el mapa solo es una proyección de una vista particular de una base de datos geográfica en un momento

dado. De esta manera el usuario de un GIS posee un número ilimitado de alternativas de análisis y de alternativas para realizar mapas desde puntos de vista diferentes de acuerdo a diferentes aspectos de la información [2].

Sin embargo, crear una línea de productos software que incluya a todo el dominio geográfico sería impráctico ya que habría que dejar demasiados aspectos abiertos a instanciar por cada producto. Esto equivaldría a desarrollar la funcionalidad casi desde cero perdiendo los beneficios del reuso de una línea de productos de software. Por lo tanto, es necesario acotar el dominio geográfico en el que se está trabajando y luego identificar un conjunto de requerimientos comunes con su variabilidad; la cual surgirá de las particularidades de los sistemas geográficos y del dominio específico en el cual se basa la información. En particular para el dominio de ecología marina, existen servicios geográficos comunes a cualquier organización que realiza sus investigaciones en ese dominio. Por ejemplo, un servicio que permita comparar características geográficas en distintos intervalos de tiempo, o generar estadísticas a partir de la información sobre la distribución poblacional de distintas especies. Al identificar estos servicios geográficos, una LPS permite implementarlos una única vez como componentes reusables facilitando el desarrollo de nuevos productos a partir de la línea. El desarrollo de componentes reusables permite ahorrar tiempo y recursos mientras que a la vez brinda flexibilidad para adaptarse a los diferentes requerimientos pertenecientes a diferentes organizaciones (instanciando la variabilidad ya definida en la LPS). Así, implementar estos servicios una única vez facilita además el desarrollo de un GIS que será utilizado en sus variantes por diferentes organizaciones que comparten servicios dentro de un dominio en particular.

Otro dominio interesante para ser analizado es

el dominio de gobierno electrónico ya que posee aspectos con respecto a la interoperabilidad que todavía no han sido profundizados en la literatura y para los cuales el reuso es una herramienta válida [3]. Es sabido que la interoperabilidad es uno de los problemas que hoy enfrentan los estados, específicamente al querer compartir e intercambiar información para sus procesos de negocios. La interoperabilidad dentro del estado es hoy en día un nudo operacional si se quiere mejorar la eficiencia incidiendo directamente en tareas asociadas a una mejor atención al ciudadano. En algunos países se están implementando plataformas de interoperabilidad únicas dentro del estado, como es el caso de Chile, denominada Plataforma de Servicios Electrónicos del Estado (PISEE<sup>1</sup>). Así, se desprende la necesidad de una Ingeniería de Dominios cuidadosamente orientada a soportar reuso de servicios en pos de una interoperabilidad efectiva<sup>2</sup>.

#### 4 Líneas de investigación y desarrollo

Dentro del dominio de ecología marina, se han realizado avances con respecto a la identificación de un conjunto de servicios geográficos comunes. Para esto se ha estado trabajando con dos organizaciones dedicadas al estudio de este dominio: el Instituto de Marina y Pesquera Biología "Almirante Storni" (IBMPAS) y el Centro Nacional Patagónico (CENPAT-CONICET). El trabajo con dichas organizaciones ha permitido conocer a fondo el conjunto de actividades y objetivos que poseen. Así, se obtuvo un conjunto de servicios adaptados al dominio y basados en los estándares geográficos definidos por el Open Geospatial Consortium<sup>3</sup> (OGC) y la Inter-

<sup>1</sup>[http://wikigob.cl/mediawiki/index.php?title=Plataforma\\_Integrada\\_de\\_Servicios\\_Electr%C3%B3nicos\\_del\\_Estado](http://wikigob.cl/mediawiki/index.php?title=Plataforma_Integrada_de_Servicios_Electr%C3%B3nicos_del_Estado)

<sup>2</sup>Interoperability Solutions for European Public Administrations - <http://ec.europa.eu/isa/>

<sup>3</sup><http://www.opengeospatial.org/>

national Organization for Standardization<sup>4</sup> (ISO). En particular, el estándar de Arquitectura de Servicios<sup>5</sup> (definido en OpenGIS Service Architecture) y la norma ISO/DIS 19119<sup>6</sup> definen una taxonomía de servicios geográficos en la que cada servicio de un sistema debería clasificarse en una o más categorías (en función de si se trata de un servicio simple o agregado). También se estudiaron y analizaron varias metodologías para el diseño de arquitecturas de LPS ampliamente referenciadas en la literatura. Se identificaron los distintos procesos propuestos, y a partir de ellos, se creó una nueva metodología aprovechando los beneficios de cada uno. La metodología propuesta resultó de la incorporación de algunas actividades específicas de los métodos y de la combinación de algunas actividades. Finalmente, se diseñó una Arquitectura de Referencia en base a un estilo arquitectónico en capas, principalmente para maximizar los requerimientos de flexibilidad y capacidad de evolución. La arquitectura diseñada consta de tres capas principales: modelo geográfico, procesamiento geográfico e interface de usuario. Dentro de cada capa, se definieron los componentes reusables que forman parte de la plataforma, donde cada componente se encarga de implementar un conjunto de servicios. También se implementaron parcialmente algunos de los componentes que conforman la plataforma, reusando partes de diferentes herramientas GIS existentes de código abierto y adaptando su código fuente. Estos trabajos han sido publicados en [8, 9, 10]. En trabajos futuros, dentro del dominio de ecología marina, se continuarán desarrollando nuevos servicios reusables y específicos de manera de validar el reuso efectivo en cada producto siendo desarrollado para cada organización. A su vez, se propondrán nuevas técnicas para la administración de la variabilidad que permitan asistir

<sup>4</sup><http://www.iso.org>

<sup>5</sup>The OpenGIS Abstract Specification: Service Architecture, 2002.

<sup>6</sup>Geographic information. Services International Standard 19119, ISO/IEC, 2005.

al proceso de creación de nuevos productos dentro de la SPL y a la definición de componentes que ayuden al descubrimiento inteligente de patrones para un mejor análisis y uso de la información almacenada.

Al mismo tiempo, se ha explorado la problemática de gobierno electrónico en los subdominios de las notificaciones, gestión de recursos y construcción de sitios accesibles. Para ello, hemos propuesto procesos normalizados basados en modelos de mejora y en estándares para incorporación de firma digital. También hemos diseñado un método para incluir accesibilidad web en etapas tempranas del diseño y una herramienta de soporte que utiliza estándares (pautas WCAG 1.0/2.01). Este último trabajo ha requerido la traducción de recomendaciones escritas en lenguaje natural a una especificación estructural de datos que pueda ser utilizada desde una perspectiva arquitectónica. Estos trabajos han sido publicados en [6, 7]. Dentro de este dominio se continuará con el desarrollo de técnicas para la mejora de la madurez de gobierno electrónico incluyendo sus prácticas más características, por ejemplo, la interoperabilidad basada en dominios.

## 5 Resultados y Objetivos

El objetivo principal de la línea de investigación es *Desarrollar técnicas y herramientas que mejoren el desarrollo de software basado en reuso a nivel de dominios, considerando los dominios específicos de los sistemas de información geográficos y las aplicaciones para gobierno electrónico*. Dentro de los primeros avances, se ha publicado un artículo [1] enfocado en la instanciación y validación de la LPS creada para dos organizaciones que trabajan en este dominio, el IBMPAS y el CENPAT-CONICET.

## 6 Formación de Recursos Humanos

El proyecto reúne aproximadamente a 10 investigadores, entre los que se cuentan docentes



y alumnos de UNComa, y asesores externos. El proyecto cuenta actualmente con 3 doctores, y un magister. Varios de los docentes-investigadores de GIISCo-UNComa han terminado o se encuentran próximos a terminar carreras de postgrado. A su vez, algunos alumnos están finalizando tesis de grado para continuar sus estudios de posgrado sobre las mismas líneas investigadas. Un alumno de estos ya posee una beca del Conicet Tipo I aceptada. Por último, se cuenta con dos alumnos dentro del programa de Becas TICS.

## References

- [1] A. Buccella, A. Cechich, M. Arias, M. Polía, S. Doldan, and E. Morsan. Towards systematic software reuse of gis: Insights from a case study. *Computers & Geosciences*, To Appear, 2013.
- [2] P. Burrough and R. McDonnell. *Principles of Geographical Information Systems*. Oxford University Press, 1998.
- [3] Giorgos Laskaridis, Konstantinos Markellos, Penelope Markellou, Angeliki Panayiotaki, Evangellos Sakkopoulos, and Athanasios Tsakalidis. E-government and interoperability issues. *International Journal of Computer Science and Network Security*, 7(9):28–38, 2007.
- [4] R. Laurini and D. Thompson. *Fundamentals of spatial information systems*. The APIC Series N 37, Academic Press, 1992.
- [5] M.A. Rodríguez Luaces. *A Generic Architecture for Geographic Information Systems*. PhD thesis, Universidade da Coruña, 2004.
- [6] J. Luzuriaga, R. Martínez, and A. Cechich. Design and implementation of an electronic signature solution in the justice area. In *Proceedings International Conference on Theory and Practice of Electronic Governance*, pages 299–304, USA, 2009. ACM (Association for Computing Machinery).
- [7] A. Martín, G. Rossi, A. Cechich, and S. Gordillo. Engineering accessible web applications: An aspect-oriented approach. *IEEE Transactions on Software Engineering*, 13(1):419–440, 2010.
- [8] P. Pernich. *Diseño e Implementación de una Arquitectura de Línea de Productos para Servicios SIG* Bachelor's thesis, Departamento de Ingeniería de Sistemas, Facultad de Informática, Universidad Nacional del Comahue, Neuquen, Argentina, May 2011.
- [9] P. Pernich, A. Buccella, A. Cechich, S. Doldan, E. Morsan, M. Arias, and M. Polía. Developing a subdomain-oriented software product line. In *Proceedings of the CACIC'11: 17th Congreso Argentino en Ciencias de la Computación*, La Plata, Argentina, 2011.
- [10] P. Pernich, A. Buccella, A. Cechich, S. Doldan, E. Morsan, M. Arias, and M. Polía. Product-line instantiation guided by subdomain characterization: A case study. *Journal of Computer Science and Technology, Special Issue 12(3)*. ISSN:1666-6038, 12(3):116–122, 2012.
- [11] Klaus Pohl, Günter Böckle, and Frank J. van der Linden. *Software Product Line Engineering: Foundations, Principles and Techniques*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.
- [12] P. Rigaux, M. Scholl, and A. Voisard. *Spatial Databases With Application To GIS*. Academic Press, 2001.

## Desarrollo de un modelo basado en servicios digitales comunes reutilizables para Gobierno Electrónico

SÁNCHEZ RIVERO David, CASTRO Marcelo, VARGAS Alejandro, ARAGÓN Fabiana, REINOSO Elizabeth, APARICIO María, FARFÁN José, CÁNDIDO Andrea, CASTRO Daniel & CAZÓN Liliana

Investigación + Desarrollo en Gobierno Electrónico / Facultad de Ingeniería /  
Universidad Nacional de Jujuy

Av. Italia y Av. Martiarena / S. S. de Jujuy / Provincia de Jujuy

Tel. 388-4221591

vdsanchezrivero@fi.unju.edu.ar, mcastro@fi.unju.edu.ar, lavargas@fi.unju.edu.ar,  
fraragon@fi.unju.edu.ar, edrreinoso@fi.unju.edu.ar, mcaparicio@fi.unju.edu.ar,  
jhfarfan@fi.unju.edu.ar, agcandido@fhycs.unju.edu.ar, ddcastro@fi.unju.edu.ar,  
lbcazon@fce.unju.edu.ar

### Resumen

Teniendo en cuenta investigaciones anteriores, se pudo observar que el proceso de Gobierno Electrónico (GE) se encuentra fuertemente orientado a los servicios. Estos servicios son brindados a la ciudadanía por diferentes organismos del Estado.

Es necesario formalizar el proceso GE a través de un modelo que satisfaga una serie de precisiones relacionadas con el desarrollo, la implementación, la sostenibilidad en el tiempo, la reutilización de componentes y sobre todo que cumpla con los requerimientos de la organización, al momento de desarrollar soluciones en el ámbito de GE. La falta de marcos conceptuales y políticas de administración de las TIC, sumadas a la carencia, que existe en la Administración Pública resultan ser inconvenientes importantes a tener en cuenta.

De allí que el proceso de GE, se convierte en un conjunto de acciones informales y poco consensuadas, lo cual hace indispensable contar con instrumentos metodológicos que permitan formalizarlo. Este proyecto intenta aplicar una metodología de modelado y diseño orientada a servicios; y la Arquitectura Orientada a Servicios (SOA) será la base para el diseño de un modelo de Servicios digitales Comunes Reutilizables (SCR), que permita elaborar una matriz de interacción entre los diferentes servicios, o entre servicios y componentes reutilizables ya desarrollados.

**Palabras clave:** Gobierno Electrónico. Tecnologías de Información y Comunicación (TIC). Tecnología orientada a servicios, Formalización del proceso. Arquitectura Orientada a Servicios (SOA).

### Contexto

El proyecto “Desarrollo de un modelo basado en servicios digitales comunes reutilizables para Gobierno Electrónico” incorpora conceptos y experiencias de proyectos anteriores, desarrollados por nuestro equipo de investigación, denominados “Tic’s: Automatización y Estandarización del Proceso de Gobierno Electrónico”, concretado con un avance del 100 %, entre los años 2007 al 2009 y “Desarrollo de una herramienta para automatizar el proceso de Gobierno Electrónico”, cuyo periodo de trabajo comprendió los años 2010 al 2012. Ambos proyectos constituyen la base conceptual para desarrollar la idea de trabajar sobre un modelo que permitiera formalizar el proceso de GE, a través de servicios digitales.

El proyecto se lleva a cabo en la Facultad de Ingeniería, de la Universidad Nacional de Jujuy, aprobada por Resolución del Concejo Superior 0167/12, el mismo posee Categoría “A” y es subsidiado por la Secretaría de Ciencia y Técnica y Estudios Regionales, dependiente de la U.N.Ju.

Es importante destacar que la Facultad de Ingeniería cuenta con convenios firmados con el Gobierno de la Provincia de Jujuy, para

efectuar tareas de investigación y transferencia, referentes al tema.

## Introducción

Este proyecto se encuentra circunscrito en el campo de la Ingeniería de Software, particularmente se centra en la aplicación de metodologías formales y semi-formales para el desarrollo y evaluación de aplicaciones; especialmente dirigido al ámbito de GE.

La incorporación de las TIC's, en ámbitos tan diversos, es consecuencia de lo imprescindible que resulta para todos los participantes del proceso de transmisión de información en tiempo y forma. Esto se encuentra sustentado principalmente, en la cada vez más sencilla forma de implementar y utilizar este tipo de tecnología.

Uno de los recursos más demandados es el posicionamiento en un entorno web, de toda clase de procesos de negocio para alcanzar ventajas competitivas, como puede ser un proceso de GE.

El Estado como generador de servicios orientados al ciudadano, también necesita tener presencia en Internet. Sin embargo la incorporación de herramientas y estándares tecnológicos en la esfera gubernamental, como es el caso de GE, no resulta una tarea sencilla.

De acuerdo con la "Metodología para la automatización y estandarización del proceso de GE", [Castro et al, 2008], es imprescindible construir soluciones de GE basadas en una serie de estándares y modelos de calidad tecnológicos que aseguren la calidad de los procesos y de los productos.

Sin perder de vista el objetivo principal que es facilitar la misión que el Estado brinde y ofrezca servicios digitalizados; se llevó a cabo una investigación de herramientas afines y se llegó a la conclusión que SOA (Service Oriented Architecture), propone un marco de trabajo conceptual que permite a las organizaciones, particularmente al Estado, unir los objetivos de negocio con la infraestructura de las TIC integrando los datos y la lógica de negocio de sus sistemas separados [Marsili, 2007].

Thomas Erl, define a "SOA, como un paradigma abstracto, tradicionalmente ha

representado una arquitectura de base distribuida sin ninguna referencia hacia la implementación". Este modelo representa sólo un subconjunto de SOA en su forma más común y contemporánea, junto con la plataforma de servicios Web y un conjunto de servicios de orientación comúnmente aceptados, los principios de SOA han surgido como una plataforma de arquitectura diferenciada, explícitamente, de sus predecesores. Introduce nuevos conceptos apoyados por tecnologías selectas que aumentan significativamente las características de las plataformas tradicionales de computación distribuida, tanto es así, que los entornos orientados al servicio a menudo terminan la redefinición de la infraestructura de Tecnología Informática (TI)" [Erl, 2006].

La World Wide Web Consortium (W3C) define a SOA como "un conjunto de componentes que pueden ser invocados, y cuyas descripciones de la interfaz se pueden publicar y exponer". Sprott & Wilkes, ampliando el concepto anterior, señalan que SOA son "las políticas, prácticas y marcos que permiten la funcionalidad de las aplicaciones que se proporcionan y se consume como un conjunto de servicios publicados con una granularidad relevante para el consumidor del servicio. Los servicios pueden ser invocados, publicados y expuestos, y se abstrae de la aplicación utilizando una interfaz basada en estándares" [Sprott, 2012].

Teniendo en cuenta lo expresado precedentemente, podemos concluir que SOA es una arquitectura y que resulta ser más que una composición de servicios web. No obstante, se piensa que la implementación de una arquitectura SOA, en una organización, significa una inversión monetaria importante en plataformas de servicios web para obtener beneficios sustanciales de su aplicación; lo cual es un error.

Una arquitectura SOA es una filosofía de diseño independiente de cualquier tendencia de proveedores, productos, tecnología o industria. Se puede implementar haciendo uso de los servicios web, pero estos servicios no necesariamente requieren implementar SOA. Y

es importante destacar que no es una metodología.

Es de aplicación gradual y se construye sobre las inversiones actuales. Es un medio y no el objetivo, es decir que lo importante es la entrega de la solución y no cómo se la desarrolle.

Como toda arquitectura, SOA se basa en la interacción de distintos componentes básicos como ser mensajes, operación, servicio y proceso. Es decir:

- Mensajes: representan los datos requeridos para completar todas o algunas partes de una unidad de trabajo.
- Operación: representa la lógica requerida para procesar los mensajes para completar una unidad de trabajo.
- Servicio: representa un conjunto de operaciones lógicamente agrupadas que puedan realizar unidades de trabajo afines.
- Proceso: contiene las reglas de negocio que determinan cuáles son las operaciones de servicio utilizado para completar una unidad de automatización [Erl, 2006].

La interrelación del trabajo de estos componentes permite, a SOA, establecer un comportamiento orientado a los servicios.

Para la Administración Pública, la finalidad de incorporar TIC es brindar servicios al ciudadano de manera rápida y eficiente; por ello distintas Administraciones están utilizando o se encuentran evaluando la posibilidad de emplear Arquitecturas Orientadas a Servicios (SOA) o Enterprise Architectures (EA)/SOA.

Estas arquitecturas están enfocadas a dar apoyo a los procesos de negocio con una infraestructura independiente y tienen en cuenta elementos centrados en la implantación de web services junto con sus adecuadas especificaciones (procesos de negocio, gestión, presentación, interoperabilidad, lenguaje XML, metadatos, fiabilidad, seguridad, transacciones, recursos o mensajería) [Criado, 2010].

Un ejemplo de la implementación de SOA en la Administración Pública, en la República Argentina, es la Administración Nacional de la Seguridad Social (ANSES) de acuerdo a su Plan Estratégico 2009-2011 [Plan, 2009].

## Líneas de Investigación y Desarrollo

El equipo de trabajo se encuentra desarrollando actividades de investigación y desarrollo, fundamentalmente en el área de la Ingeniería de software y específicamente en la formalización del proceso de Gobierno Electrónico, Ingeniería Web y Calidad del Software a partir de los proyectos “TIC’s: automatización y estandarización del proceso de Gobierno Electrónico” y “Diseño y Desarrollo de una herramienta para automatizar el proceso de Gobierno Electrónico”.

En el nuevo desafío que presenta el proyecto “Desarrollo de un modelo basado en servicios digitales comunes reutilizables para Gobierno Electrónico”, se trabajará sobre ejes disciplinares relacionados a Arquitecturas Avanzadas de Software.

Se investigarán, en principio, las ventajas e inconvenientes sobre la utilización de la arquitectura multicapa y los patrones relacionados a la misma. Posteriormente se procederá a estudiar los objetivos y beneficios de SOA, servicios web asociados y su aplicación específica a soluciones de GE. Simultáneamente se evaluará la conveniencia de diseñar el modelo utilizando UML Web Application Extension (UML WAE) o UML based Web Engineering (UWE).

## Resultados y Objetivos

Para el inicio del desarrollo del proyecto de investigación, se planteó como primera etapa el relevamiento de documentación y bibliografía en la temática referida a las tecnologías orientadas a servicios, lo cual fue ampliamente cumplido y resumido en el trabajo denominado “Servicios digitales comunes reutilizables para gobierno electrónico”, presentado en el año 2012 [Castro et al., 2012].

En el citado trabajo se analizaron aspectos conceptuales de diferentes autores sobre la Arquitectura Orientada a Servicios (SOA), estableciendo claramente una serie de nociones que la distinguen. En primer lugar como filosofía de diseño independiente y luego como arquitectura de referencia, además se establece que no se trata de una metodología. Posee un



cierto grado de madurez en su desarrollo, que construye en forma gradual sobre lo existente y representa un medio para alcanzar la solución deseada.

Por otra parte, se analizaron en forma detallada los diferentes componentes que la conforman, mensajes, operaciones, servicios, procesos y sus interrelaciones; concepciones imprescindibles para la concreción del actual proyecto de investigación.

A posteriori se procedió a establecer la necesidad de utilizar SOA, como arquitectura conceptual orientada a los servicios que poseen las soluciones de GE. Para esto se procedió a analizar minuciosamente los diversos enfoques que resultan sumamente necesarios para poder llevar a cabo una arquitectura basada en servicios [Chaix, 2008], tales como la necesidad del ciudadano, la intervención del organismo responsable, las estrategias de apoyo y la participación de los diferentes actores del proceso.

Actualmente, el equipo de investigación se encuentra abocado a analizar y diseñar el modelo de servicios digitales comunes reutilizables, realizando el análisis y diagnóstico de un sistema en producción, y de la arquitectura tecnológica que posee la unidad de organización bajo estudio.

Es importante señalar que la implementación de SOA en organismos gubernamentales debe estar lo suficientemente justificada ya que el proceso tiene un alto costo desde varios puntos de vista, lo que podría desembocar en un fracaso si el mismo no se encuentra lo suficientemente sustentado en el conocimiento de la parte tecnológica y de los problemas que se plantean en el ámbito gubernamental, perdiendo así las múltiples ventajas que ofrece el uso de esta arquitectura de referencia.

### Formación de Recursos Humanos

El proyecto de investigación se encuentra conformado por once docentes investigadores, cinco de los cuales se encuentran categorizados y seis en proceso de formación. Además cuenta con cuatro alumnos que están siendo iniciados en tareas de investigación y actividades de desarrollo.

En el transcurso del año anterior, se incorporó una alumna becaria perteneciente al Programa de Estímulo a las Vocaciones Científicas, el cual depende del Consejo Interuniversitario Nacional. La becaria se encuentra desarrollando actualmente actividades relacionadas al proyecto de investigación en la Université de Technologie de Belfort – Montebéliard- Francia.

Por otra parte el desarrollo de las tareas de investigación, ha generado en los años 2012, tres anteproyectos de tesis en la Maestría en Ingeniería de Software, de la Universidad Nacional de San Luis todas pertenecientes a integrantes del equipo de trabajo, con una inminente presentación de un plan de tesis que lleva por título “Creación y evaluación de modelos LSP en un contexto MDA”.

Además se realizó la presentación y aprobación de una propuesta técnica de Tesis Doctoral para cursar el Doctorado en Ciencias Informáticas en la Facultad de Informática de la Universidad Nacional de la Plata.

Simultáneamente se encuentran en desarrollo cinco proyectos finales de carrera de grado, pertenecientes a alumnos de Ingeniería Informática, de la Facultad de Ingeniería de la Universidad Nacional de Jujuy.

### Referencias

- [Castro et al., 2008] Castro, M., Farfán J., Sánchez Rivero D, Castro D.; “Tic: sistematizando el proceso de gobierno electrónico”, Actas del “III Congreso da Academia Trinacional de Ciencias”, Universidade Estadual do Oeste do Paraná – Unioeste, Brasil; 2008.
- [Marsili, 2007] Marsili, Diego; “¿Qué es SOA, la arquitectura orientada a servicios?”. Disponible en: <http://www.iprofesional.com/notas/46399-Que-es-SOA-laarquitectura-orientada-a-servicios.html>, Última visita: Mayo de 2012.
- [Erl, 2006] Erl, Thomas; “Service-Oriented Architecture: Concepts, Technology & Design”. Pearson Education, 2006; Disponible en:

- [http://www.soabooks.com/ctd/sample\\_chapters](http://www.soabooks.com/ctd/sample_chapters). Última visita: Abril de 2012.
- [Sprott, 2012] Sprott, David & Wilkes, Lawrence, “Understanding Service-Oriented Architecture”, MSDN Library. Disponible en: <http://msdn.microsoft.com/en-us/library/aa480021.aspx>; Última visita: Abril de 2012.
- [Criado, 2010] Criado, Ignacio; Gascó, Mila & Jiménez, Carlos; “Bases para una Estrategia Iberoamericana de Interoperabilidad”; Disponible en: [http://www.sgp.gov.ar/contenidos/cofefup/documentos/docs/ESTRATEGIA\\_INTEROPERABILIDAD\\_2010.pdf](http://www.sgp.gov.ar/contenidos/cofefup/documentos/docs/ESTRATEGIA_INTEROPERABILIDAD_2010.pdf); Última visita: Mayo de 2012.
- [Plan, 2009] Plan Estratégico Anses 2009-2011; Disponible en: [www.anses.gov.ar/prensa/.../2009/Plan\\_Estrategico2009-2011.pdf](http://www.anses.gov.ar/prensa/.../2009/Plan_Estrategico2009-2011.pdf); Última visita: Mayo de 2012.
- [Castro et al., 2012] Castro M., Sánchez Rivero D., Reinoso E., A, Aparicio M., Aragón F., Cazón L., “Servicios digitales comunes reutilizables para gobierno electrónico”, 41 JAIIO-Jornadas Argentinas de Informática en la Facultad, La Plata, Argentina, 2012.
- [Chaix, 2008] Yves Chaix, “Propuesta de arquitectura para los gobiernos municipales electrónicos”, 2008. Disponible en: [http://www.gobenic.gob.ni/eventos/gobierno-digital-municipal/Propuesta\\_de\\_arquitectura\\_para\\_el\\_gobierno\\_electronico\\_municipal.V2.pdf/view](http://www.gobenic.gob.ni/eventos/gobierno-digital-municipal/Propuesta_de_arquitectura_para_el_gobierno_electronico_municipal.V2.pdf/view). Última visita Mayo de 2012.

## Herramientas de Modelado y Simulación para Sistemas de Gran Escala

Jair Lobos<sup>1</sup>, Veronica Gil-Costa<sup>2,3</sup>, Andrea Giubergia<sup>2</sup> and Marcela Printista<sup>1,3</sup>

<sup>(1)</sup>Departamento de Computación  
Facultad de Ciencias Físico  
Matemáticas y Naturales  
Universidad Nacional de San Luis  
Ejército de los Andes 950, 1° piso.  
(02652-420823)

<sup>(2)</sup>Departamento de Minería  
Facultad de Ciencias Físico  
Matemáticas y Naturales  
Universidad Nacional de San Luis  
Chacabuco y Pedernera.  
(02664-4436531)

<sup>(3)</sup> CONICET San Luis  
Almirante Brown 907  
(02664 421654 )

### CONTEXTO

La línea de investigación presentada en este trabajo recurre a un proyecto que vincula estrechamente dos temas que han cobrado gran interés en los últimos años debido al avance de la tecnología y a los costos excesivos que requieren las pruebas y ejecuciones sobre plataformas reales. Nos referimos a las líneas de Modelado y Simulación.

En particular, nos enfocamos en el modelado de aplicaciones de gran escala para plataformas paralelas que no pueden ser probadas en sistemas y hardware reales debido al costo de los mismos. Para ello, es posible utilizar diferentes herramientas como las Petri Nets [Petri62], Devs [Zeig76], Análisis Operacional [Den78] y UML. Otra ventaja de las técnicas de modelado y simulación utilizadas en este proyecto, es que permite obtener estimaciones de las métricas utilizadas en las aplicaciones para determinar el costo-beneficio de implementar y desplegar la aplicación en un hardware real.

### RESUMEN

El crecimiento constante de la tecnología, diseño e implementación de nuevas arquitecturas de computadoras, placas aceleradoras como las GPU, redes de alta velocidad, etc., son el resultado de la gran demanda introducida por los usuarios de sistemas de gran escala que requieren no sólo procesar sus requerimientos rápidamente (fracciones de segundo) sino de la gran cantidad de datos utilizados en estos sistemas.

Algunos de estos datos (los más cercanos al usuario de redes) se encuentran en las páginas o documentos Web que son procesadas e indexadas por grandes motores de

búsqueda Web como Google, Yahoo! o Bing. Pero existen datos más complejos como las imágenes satelitales que son utilizados por especialistas en otras disciplinas como la geología y la minería. Estos datos (ya sean imágenes, información referente a perforaciones realizadas en la superficie de la tierra, etc.) deben ser eficiente y eficazmente procesadas por sistemas de información en un tiempo razonable.

En ambos casos, al utilizar documentos Web como datos de sistemas industrializados de propósito específico, existe la complejidad de que dichos datos tienden a ser modificados a lo largo del tiempo. Esto último introduce una dimensión adicional que debe ser analizada al momento de diseñar estos sistemas complejos.

En este trabajo, se presentan los objetivos y los desafíos que se pretenden abordar desde el grupo interdisciplinario de investigación de la Universidad Nacional de San Luis, para abordar los desafíos que involucra el modelado y diseño de sistemas de gran escala que deben ser capaces de procesar grandes volúmenes de datos e información.

**Palabras clave:** *Sistemas de Gran Escala, Herramientas de modelado: Petri Nets, Devs, UML.*

### 1.INTRODUCCION

La gran demanda de procesamiento y análisis eficiente de grandes volúmenes de datos e información (conocida recientemente como “Big Data” y “stream processing”) ha introducido el desafío del desarrollo de sistemas que sean capaces de resolver los requerimientos de

procesamiento realizados por diferentes usuarios sobre estos datos. En este trabajo se describen los lineamientos de investigación que se llevan a cabo sobre dos tipos de sistemas que involucra la administración de datos masivos. El objetivo principal es poder modelar eficientemente (a través de herramientas existentes) estos sistemas para luego mediante técnicas de simulación poder verificar el comportamiento de los mismos sobre diferentes escenarios de carga de trabajo.

El primer sistema consiste en motores de búsqueda Web de gran escala, que deben ser capaces de procesar miles de consultas por segundo. Estos motores de búsqueda utilizan servicios enfocados a realizar diferentes tareas en simultáneo. Algunas de estas tareas involucran (a) la administración de memorias caches que almacenan las consultas mas referenciadas, (b) el ranking de los top-k documentos más relevantes para las consultas de los usuarios, (c) selección de publicidades que mejor se ajustan a las consultas y a los usuarios, (d) gestión de historial de usuarios, etc..

El segundo sistema comprende el procesamiento de información y datos (generalmente imágenes) que son requeridos para evaluar proyectos de gran escala como lo son los proyectos de extracción de minerales. En particular, en este caso, se utiliza el lenguaje de modelado UML y se lo extiende mediante el uso de perfiles para soportar características propias de los sistemas bajo estudio.

### 1.1 Petri Nets

Las Redes de Petri corresponde a un concepto introducido por Carl Adam Petri en 1962 [Petri62]. Las Redes de Petri son una representación matemática y gráfica para describir y estudiar sistemas a eventos discretos en el cual se puede describir la topología de un sistema concurrente, paralelo o distribuido [Murata89, Molloy81]. Algunos trabajos introductorios se pueden encontrar en [Peterson77, Peterson81] y algunos libros en [Aalst11, Silva85]

Una Red de Petri es un tipo particular de grafo dirigido que posee un estado inicial, denominado  $M_0$ . El grafo generado  $N$  de una Red de Petri es un grafo dirigido, con peso y bipartito que posee 3 tipos de elementos: (1) nodos, también llamados lugares (places); (2) transiciones (transitions), que reflejan las acciones o eventos y (3) arcos (arcs), que conectan un lugar/nodo con una transición o conectan una transición a un lugar/nodo.

Gráficamente los lugares se representan con un círculo, las transiciones con barras o cajas. Los arcos son etiquetados con sus pesos (enteros positivos). Las etiquetas para pesos unitarios usualmente son omitidas. Un arco que esté etiquetado con  $k$  puede ser interpretado como  $k$  arcos paralelos. Una marca (estado) se asigna a cada lugar como un entero no negativo. Si se marca un lugar  $p$  con un valor  $k$  no negativo, se dice que  $p$  se marca con  $k$  tokens, lo cual representa a  $k$  recursos los cuales se mueven entre los lugares de la red. Los tokens o fichas se representan, en una Red de Petri, por medio de un punto negro e indican los recursos que posee un lugar.

A diferencia de los simuladores de eventos discretos, un modelo basado en Redes de Petri [Petri62] es más simple y eficiente, y es más fácil de extender de forma tal de incluir nuevas características o cambiar el comportamiento de algunos componentes del modelo utilizando un lenguaje gráfico y un test de verificación del modelo.

Sin embargo, una limitación de las Redes de Petri tradicionales es que al modelar un sistema real, se debe generar una gran cantidad de lugares (places) y transiciones (transitions), por lo cual su análisis se hace complejo. Además, en los sistemas reales, a menudo, se presentan procesos similares que se producen en paralelo o de forma simultánea, y se diferencian unos de los otros por sus entradas y sus salidas. Al utilizar Redes de Petri coloreadas, la cantidad de lugares, transiciones y arcos, en general se reducen. Lo cual hace que el modelo tenga una mayor facilidad de entendimiento.

### 1.2 DEVS

La "Especificación de Sistemas de Eventos Discretos (DEVS)" es un formalismo para modelar sistemas y realizar y simulaciones. Se basa en conceptos de teoría de sistemas y fue desarrollada por Zeigler en [Zeig76].



DEVS fue creado para modelar y simular sistemas dinámicos de eventos discretos, de manera que permite especificar sistemas cuyo estado se altera ya sea por la recepción de un evento de entrada o por el vencimiento de una demora de tiempo. Como una manera de rebajar la complejidad del sistema a estudiar es que el modelo se organiza jerárquicamente, en que los componentes de alto nivel del sistema se descomponen en elementos más simples. La separación entre modelo y simulador, y su naturaleza de módulos jerárquicos ha permitido llevar a cabo pruebas formales en las diferentes entidades bajo estudio [Wai09].

Por medio de DEVS, un sistema real se representa mediante la composición de componentes atómicos y acoplados (un componente acoplado corresponde a la composición de dos o más componentes atómicos). Un componente atómico está definido por el conjunto de eventos y puertos de entrada, el conjunto de puertos y eventos de salida, el conjunto de estados, una función de transición de estados externa, una función de transición de estados interna, la función de salida y la función de avance de tiempo. Un estado para el cual la función de avance del tiempo es cero se denomina "estado transiente" y dispara una transición interna de forma inmediata. Por el contrario, si el tiempo es infinito, se denomina "estado pasivo". El sistema continuará en dicho estado hasta que reciba un evento externo [Zeig76, Zeig00].

Un modelo DEVS acoplado se compone de uno o más modelos atómicos y/o submodelos acoplados. Está definido formalmente por el conjunto de puertos y eventos de entrada, el conjunto de puertos y eventos de salida, el conjunto de nombres de los componentes, el conjunto de componentes que lo forman, el conjunto de acoples externos de entrada, el conjunto de acoples externos de salida, el conjunto interno de acoples y, finalmente, por la función de selección. Esta función de selección permite resolver la ambigüedad que surge cuando hay más de un evento que debe realizar una transición interna al mismo tiempo [Zeig76].

### 1.3 Análisis Operacional

Por medio del análisis operacional se puede realizar predicciones respecto del rendimiento de sistemas representados mediante modelos de redes de colas. En análisis Operacional

[Den78], todas las ecuaciones derivan al menos los siguientes tres principios:

- Todas las cantidades deben ser definidas de manera que sean precisamente medibles, y las suposiciones de forma que sean directamente comprobables.
- El sistema debe ser balanceado en su flujo.
- Los dispositivos deben ser homogéneos.

En base a estos tres principios del análisis operacional se pueden derivar ecuaciones que permiten caracterizar el comportamiento y rendimiento de sistemas tales como los computacionales. Desde el punto de vista del análisis operacional, siempre deben existir los elementos "sistema" y "periodo de tiempo" en un problema. El sistema puede ser uno real o supuesto y el periodo de tiempo puede ser pasado, presente o uno futuro.

Las variables del análisis operacional poseen el mismo valor durante dicho periodo de tiempo. Estas pueden ser obtenidas directamente mediante mediciones o derivadas de estas. Entre las variables básicas - obtenidas durante el periodo de observación - se tienen:

- T: el periodo de observación.
- A: el número de arribos durante T.
- B: el tiempo total en que el sistema estuvo ocupado durante T.
- C: El numero de "trabajo" completos durante T.

A partir de las cuales se derivan las siguientes métricas de rendimiento:

- $\lambda = A/T$ , es la tasa de arribos medida en trabajos por segundo.
- $X = C/T$ , es la tasa de salida también medida en trabajos por segundo (throughput).
- $U = B/T$ , es la utilización, o fracción del tiempo T en que el sistema estuvo ocupado.
- $S = B/C$ , es el tiempo de servicio promedio por trabajo completado.

Todas estas variables pueden cambiar de valor de un periodo de observación a otro.

### 1.4 UML y Simulación

El Lenguaje de Modelado Unificado, UML (Unified Modelling Language) [Booch98], fue construido para diseñar sistemas de información y para facilitar el desarrollo y mantenimiento de sus procesos. Con el pasar de los años, el uso de UML se ha extendido y actualmente es el estándar más utilizado para especificar y documentar sistemas.

UML es una notación de propósito general, pero no siempre satisface las necesidades que requiere una aplicación en particular. Es por ello que UML permite extender su sintaxis y su semántica a través de mecanismos propios, para que lo conviertan en un lenguaje con características más específicas orientadas a ciertos dominios.

Estos mecanismos de extensión, inherente al mismo UML, que permiten extender y adaptar las metaclasses de un metamodelo cualquiera a las necesidades concretas de un dominio de aplicación, se denominan Perfiles UML (UML Profiles) [Deb06].

Por otro lado, la técnica de Simulación [Gold07], [Sad07] implica replicar artificialmente las características de un sistema a través de un modelo e imitar su operación a medida que transcurre el tiempo. En base al análisis de comportamiento del modelo, luego será posible inferir las características operacionales del sistema real. El modelado y la simulación brindan la posibilidad de estudiar nuevas estrategias y de predecir el efecto de la aplicación de nuevas políticas, que de otra manera, serían excesivamente costosas o incluso imposibles de reproducir y de estudiar.

Ambos enfoques, modelado UML y simulación, son ampliamente usados en ingeniería de sistemas. Si bien estas dos teorías han evolucionado por separado, existen pocos trabajos [Teil08] que integran las herramientas de modelado con las herramientas de software de los lenguajes de simulación para desarrollar parcial o totalmente un sistema.

## 2. LINEAS DE INVESTIGACION y DESARROLLO

La línea de investigación descrita en la sección anterior involucra una serie de desarrollos individuales que en su conjunto logran obtener el objetivo planteado. Para ello es necesario estudiar formas.

## 3. RESULTADOS OBTENIDOS ESPERADOS

### Modelado y Simulación mediante Análisis Operacional

Los resultados obtenidos hasta el momento son:

- Modelado de un motor de búsqueda Web basado en servicios [Gil13].
- Diseño e implementación de un simulador basado en procesos.
- Estudio de formulas que permitan determinar la cantidad de recursos requeridos por el motor de búsqueda.

Los resultados esperados son:

- Desarrollo de una metodología que permita evaluar la capacidad computacional de un motor de búsqueda.
- Validación de las fórmulas y metodología desarrollada.

### Modelado a través de UML

Los resultados obtenidos hasta el momento son:

- Mostrar que UML proporciona una alta flexibilidad a la hora de modelar sistemas [Giuber12].
- Crear un metamodelo que incluya el perfil específico, que luego pueda ser trasladado a un ambiente gráfico de simulación

Los resultados esperados son:

- Mostrar que UML puede emplearse como una etapa previa o de transición hacia la implementación de la simulación.

### Modelado y simulación a través de DEVS

Los resultados obtenidos hasta el momento son:

- Estudio de métodos de formalización como DEVS
- Modelo y diseño de un motor de búsqueda Web (sin jerarquías de cache) utilizando DEVS

• Los resultados esperados son:

- Extender el primer modelo para incluir características más relevantes como jerarquías de cache y costos de consultas individuales.
- Estudiar técnicas de integración de módulos de software.

### Modelado y simulación con Petri Nets

Los resultados obtenidos hasta el momento son:

- Modelado de un motor de búsqueda Web mediante petri Nets[Gil12]
- Análisis y validación del modelo.

Los resultados esperados son:

- Diseño de un simulador de petri nets paralelo que se ajuste a las características de un motor de búsqueda.

#### 4. FORMACION DE RECURSOS HUMANOS

Actualmente, se cuenta con dos doctores en ciencias de la computación realizando la investigación teórica y dirección de los algoritmos propuestos. También se cuenta con un alumno de doctorado que se encuentra iniciando su carrera doctoral, un segundo alumno de doctorado que se encuentra realizando una estadía en Ottawa, Canada con un grupo de investigación especializado en DEVS; y una alumna de maestría próxima a finalizar su tesis.

Mediante este trabajo de investigación se podrán formar profesionales que puedan modelar, diseñar e implementar algoritmos eficientes que se ejecuten en sistemas de gran escala y requieren el procesamiento de datos masivos.

#### 5. BIBLIOGRAFIA

- [Aalst11] W.M.P. van der Aalst and C. Stahl. Modeling Business Processes - A Petri Net-Oriented Approach. The MIT Press, 2011.
- [Booch98] Booch G.; Rumbaugh J.; Jacobson I., (1998), The Unified Modeling Language User Guide (Addison-Wesley Object Technology Series).
- [Buzen76] J. P. Buzen. Fundamental operational laws of computer system performance. Acta Inf., 7:167-182, 1976
- [Deb06] Debnath N.C., Garis A., Riesco D., Montejano G., (2006), Defining Patterns Using UML Profiles.
- [Den78] P. J. Denning and J. P. Buzen. The operational analysis of queueing network models. ACM Computing Surveys, 10:225-261, 1978.
- [Gil12] "Capacity Planning for Vertical Search Engines: An approach based on Coloured Petri Nets". Gil-Costa Veronica, Lobos Jair, Inostroza-Psijas Alonso and Mauricio Marin. Petri Nets 2012.
- [Gil13] "Service Deployment Algorithms for Vertical Search Engines". Alonso Inostroza-Psijas, Veronica Gil-Costa, Mauricio Marin and Esteban Feustein. PDP 2013.
- [Giub12] "Estereotipos UML para Aplicar en un Ambiente de Simulación de Procesos Mineros". Andrea Giubergia, Daniel Riesco, Marcela Printista y Veronica Gil Costa. CACIC 2012, Argentina.
- [Gold07] Goldsman D., (2007), Introduction to simulation. Proceedings of the 2007 Winter Simulation Conference. IEEE.
- [Molloy81] M.K. Molloy. On the integration of delay and throughput measures in distributed processing models. Technical report, Phd Thesis, UCLA, 1981.
- [Murata89] Murata, T. 1989. Petri nets: Properties, analysis and applications. Proceedings of the IEEE 77, no. 4: 541-580.
- [Peterson77] J.L. Peterson. Petri nets. Computing Surveys, 9:223-252, 1977.
- [Peterson81] J.L. Peterson. Petri net theory and the modeling of systems. Prentice Hall, Englewood Cliffs, 1981.
- [Petri62] C.A. Petri. "Communication with Automata" New York: Griffiss Air Force Base. Tech. Rep. RADC-TR-65-377, vol.1, Suppl. 1, 1962.
- [Sad07] Sadowski D.A., (2007), Tips for successful practice of simulation. Proceedings of the 2007 Winter Simulation Conference.
- [Silva85] M. Silva. Las Redes de Petri en la Automatica y la Informatica. AC, Madrid, 1985.
- [Teil08] Teilans A.; Kleins A.; Merkurjev Y.; Grinbergs A., (2007), Design of UML models and their simulation using ARENA. WSEAS TRANSACTIONS ON COMPUTER RESEARCH. Issue 1, Volume 3, January 2008.
- [Wai09] Wainer, G. 2009. Discrete-Event Modeling and Simulation: A Practitioner's Approach.
- [Zeig76] Zeigler, B. P. 1976. Theory of Modeling and Simulation. Wiley-Interscience.
- [Zeig00] Zeigler, B. P, H. Praehofer, and T.G. Kim. 2000. Theory of Modeling and Simulation, 2nd. ed. New York: Academic Press.

## Semántica Oculta en Modelos de Requisitos

Marcela Ridaó <sup>(1)</sup> Jorge H. Doorn <sup>(1)(2)</sup>

(1) INTIA, Universidad Nacional del Centro de la Provincia de Buenos Aires, Argentina

(2) Universidad Nacional de La Matanza, Argentina  
e-mail: {mridao, jdoorn}@exa.unicen.edu.ar

### Resumen

La Ingeniería de Requisitos (IR) es un área dentro de la Ingeniería de Software que tiene por objetivo mejorar la forma en que se comprenden y definen sistemas de software complejos. Los modelos construidos en el proceso de la Ingeniería de Requisitos han sido concebidos procurando maximizar su expresividad en relación con su propósito. Este proyecto se enmarca en una estrategia de la IR basada en dos modelos principales: el LEL (Léxico Extendido del Lenguaje) y los Escenarios. Estos modelos constituyen la base de una estrategia para obtener una especificación de requisitos de software. Sin embargo, estos modelos parecen contener más información que la que puede obtenerse en una primera lectura. En este artículo se propone, justamente, realizar nuevas lecturas de estos modelos para extraer esa información adicional. En particular, se propone obtener información sobre la estructura semántica de glosarios del proceso del negocio, a partir de su visualización por medio de grafos.

**Palabras clave:** Ingeniería de Requisitos, Glosarios, Trazado de Grafos, Métodos Dirigidos por Fuerzas.

### Contexto

Esta línea de I/D se enmarca en el proyecto acreditado Bases de Datos y Procesamiento de Señales, que se desarrolla en la Facultad de Ciencias Exactas de la Universidad Nacional del Centro de la Provincia de Buenos Aires, INTIA (INvestigación en Tecnología Informática Avanzada).

### Introducción

En los últimos años, se ha prestado especial atención a problemas tales como las redes organizacionales o redes sociales organizacionales, redes de referencias bibliográficas, redes de grupos de interés, etc. En este contexto, han adquirido importancia las disciplinas dedicadas al estudio de fenómenos donde el aspecto dominante es la complejidad estructural y no la complejidad esencial de los elementos involucrados en la estructura [1] [2].

Algunos de los modelos de la Ingeniería de Requisitos pueden ser estudiados desde el punto de vista estructural. En particular, el LEL puede ser visualizado bajo una óptica estructural, si se lo representa mediante un grafo, donde los símbolos sean representados mediante nodos, y las menciones a otros símbolos como arcos conectando dichos nodos. De esta manera, el LEL puede visualizarse como una suerte de red lingüística con una estructura claramente compleja. Esta estructura contiene información implícita empotrada, adicional a la información explícita que cada nodo almacena.

Construir y analizar el grafo de los símbolos del LEL se constituye así en una suerte de minería de información y de conocimiento, ya que se obtiene información sintáctica acerca

de la estructura del LEL y al mismo tiempo se adquiere parte del conocimiento subyacente bajo esa estructura.

El LEL, como se verá más detalladamente en la siguiente sección, es una representación de los símbolos del lenguaje del dominio del problema, cuyo objetivo principal es que el ingeniero de requisitos conozca el lenguaje que habla el usuario, posponiendo la comprensión del problema a etapas posteriores de la estrategia de IR. Sin embargo, se puede suponer que si el Universo de Discurso contiene áreas de interés diferenciadas, los términos correspondientes a cada una de estas áreas deberían tener un mayor grado de acoplamiento que aquellos que corresponden a áreas diferentes.

Por lo anterior, en este trabajo se propone analizar el grafo construido a partir del LEL, con el fin de estudiar los mencionados agrupamientos.

### LEL: Léxico Extendido del Lenguaje

La construcción de un vocabulario que capture la jerga usada por los expertos del dominio ha sido propuesta por distintos autores [3] [4]. De hecho, varias experiencias han mostrado que un glosario del vocabulario de los clientes-usuarios es, en sí mismo, una fuente de información para elicitar información del Universo de Discurso [5] [6] [7] [8] [9].

En este trabajo, se estudiará un modelo de glosario en particular: el Léxico Extendido del Lenguaje (LEL). El LEL es una representación de los símbolos del lenguaje del dominio del problema, cuyo objetivo principal es que el ingeniero de requisitos conozca el lenguaje que habla el usuario, sin preocuparse por entender el problema [10] [11].

Este léxico se construye utilizando lenguaje natural y está compuesto por símbolos que pueden ser Sujetos (realizan acciones), Objetos (las acciones se realizan sobre ellos), Verbos (acciones del sistema) y Estados significativos del sistema [12].

Cada símbolo tiene uno o más nombres o frases que lo identifican y dos tipos de descripciones, la noción y el impacto. La noción describe la denotación de la palabra o frase. Indica quién, cuándo ocurre, qué procesos involucra, qué significado tiene el símbolo, etc. El impacto describe la connotación del símbolo, es decir, su repercusión en el sistema. Cada entrada puede contener una o más nociones y uno o más impactos.

### Trazado de Grafos: Métodos dirigidos por fuerzas

La Teoría de Grafos tiene diversidad de aplicaciones. La representación mediante nodos y conexiones es usada para representar redes físicas como circuitos eléctricos, carreteras, moléculas orgánicas, y también interacciones menos tangibles



como relaciones sociológicas, bases de datos, o el flujo de control de un programa computacional [13] [14].

El Trazado de Grafos aplica topología y geometría para derivar representaciones de grafos en dos dimensiones. Básicamente, consiste en una representación gráfica del grafo en el plano, usualmente destinada a una visualización conveniente de ciertas propiedades del grafo en cuestión o de los objetos modelados [15] [16].

En este trabajo, al representar el grafo de un LEL, el énfasis se pone en la estructura de la red, y no en los criterios estéticos utilizados generalmente en el trazado de grafos [17], como distribución uniforme de los nodos, minimización de cruces de arcos, uniformidad en la longitud de los arcos, simetría, etc.

Existen diferentes estilos de representación, adecuados a diferentes tipos de grafos o diferentes propósitos de representación [18]. Entre una gran variedad de algoritmos se destaca una familia de métodos conocidos como “dirigidos por fuerzas”. Estos métodos son muy usados hoy en día para dibujar grafos, porque dan buenos resultados, son sencillos de implementar, y son muy flexibles, por lo que pueden ser fácilmente adaptados a aplicaciones concretas con requerimientos de visualización específicos [19] [20]. Estos algoritmos usan analogías físicas para dibujar el grafo. Tienen como denominador común las siguientes características:

- Modelan al grafo como un sistema físico.
- El trazado del grafo es obtenido buscando el equilibrio del sistema físico.

Los modelos físicos más comunes son los que consisten de un sistema de fuerzas (donde generalmente se definen fuerzas que actúan entre los vértices del grafo), en cuyo caso el objetivo del algoritmo es encontrar una posición para cada vértice, de manera que el total de la fuerza ejercida en cada vértice sea cero.

Entre los primeros autores aplicando analogías con sistemas físicos para el trazado de grafos, se destaca el “Spring Embedder” propuesto por Eades [21], que se basa en reemplazar los nodos por anillos de acero y cada arco con un resorte para formar un sistema físico. Los nodos son ubicados en alguna disposición inicial, y se dejan actuar las fuerzas de los resortes hasta lograr un estado de energía mínima. La implementación de Eades, sin embargo, no siguió al pie de la letra la ley de Hooke, sino que incorporó al cálculo de las fuerzas resultantes, fuerzas repulsivas calculadas entre los nodos no conectados.

Otros autores proponen algoritmos derivados del Spring Embedder de Eades, como Kamada y Kawai [22], Davidson y Harel [23], Chernobelskiy et al [24], Kobourov [25], Bannister et al [26], Fruchterman y Reingold [17], entre otros.

En este trabajo, se utilizará el algoritmo propuesto en [17], que se basa en los siguientes principios:

- Los nodos conectados por un arco deberían ser dibujados cerca.
- Los nodos no deberían ser dibujados *demasiado* cerca uno de otro.

Cuán cerca se deberían ubicar los nodos, depende de cuántos haya y cuánto sea el espacio disponible. El algoritmo se basa en simulaciones moleculares o planetarias. Si los nodos se comportan como partículas atómicas o cuerpos celestes, ejerciendo fuerzas atractivas y repulsivas sobre los demás, las fuerzas inducen movimiento. Sin embargo, no se propone una simulación exactamente fiel a la realidad. Del

mismo modo que en el algoritmo de Eades, sólo los nodos que son vecinos se atraen entre sí, mientras todos los vértices se repelen unos a otros. Esto es consistente con la asimetría propuesta por los dos principios antes enunciados.

## Objetivos

Con el fin de detectar agrupamientos de símbolos, se propone aplicar el algoritmo propuesto por Fruchterman y Reingold [17] en la visualización de los grafos correspondientes a los Léxicos de diferentes casos de estudio.

Para un LEL dado, entonces, se propone construir un grafo donde los símbolos son los nodos y las menciones a otros símbolos, los arcos.

Para la configuración inicial, los nodos se ubican al azar en el marco de trabajo, y luego se aplica el algoritmo, modificando la ubicación de los nodos en forma iterativa. Cada iteración consta de tres pasos:

- calcular el efecto de las fuerzas atractivas sobre cada nodo,
- calcular el efecto de las fuerzas repulsivas y
- modificar la posición de los nodos.

Inicialmente, se efectuaron pruebas utilizando el algoritmo con las fuerzas propuestas por Fruchterman y Reingold. En este trabajo, se propone modificar el sistema de fuerzas, comparando los resultados obtenidos.

## Las fuerzas

$f_a$  y  $f_r$  son las fuerzas de atracción y de repulsión, respectivamente.

Las pruebas iniciales, fueron efectuadas utilizando el sistema de fuerzas propuesto por [17]:

$$f_a(d) = c1 * \frac{d^2}{k} \quad k = C * \sqrt{\frac{Area}{NúmeroNodo \ s}}$$

$$f_r(d) = c2 * \frac{k^2}{d}$$

$d$  es la distancia entre los vértices y  $k$  el radio vacío alrededor de un nodo

Posteriormente, se trabajó con el par de fuerzas propuesto por Eades [21]:

$$f_a(d) = c3 * \log\left(\frac{d}{c4}\right)$$

$$f_r(d) = c5 * \sqrt{d}$$

Las constantes C, c1, c2, c3, c4 y c5 son obtenidas experimentalmente.

## Marco de trabajo

El grafo debe ser confinado al marco especificado por el usuario.

El algoritmo propuesto por [17] considera la ubicación de nodos ficticios en el perímetro del marco de trabajo, que ejercen fuerzas repulsivas sobre los nodos del grafo, pero ellos mismos permanecen fijos. De este modo, el marco se modela como cuatro paredes que contienen al grafo dentro de ellas.

Sin embargo, las experiencias realizadas permitieron determinar que no era necesaria la ubicación de nodos en las paredes del marco de trabajo, ya que, con la elección

adecuada de las constantes que intervienen en el cálculo de las fuerzas, los nodos se mantienen dentro del marco definido.

**Resultados obtenidos**

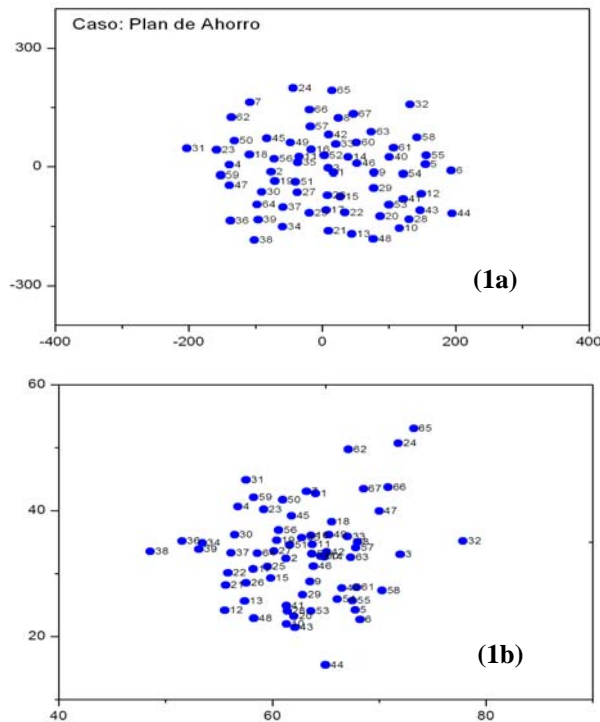
Con el fin de verificar si el algoritmo propuesto permite detectar agrupamientos de símbolos, se lo aplicó a diferentes casos de estudio cuyos Léxicos habían sido verificados y validados previamente.

**Caso 1. Sistema de Planes de Ahorro Previo para la Adquisición de Vehículos 0Km [27]**

Para este caso se prefirió no analizar en forma semántica la presencia o ausencia de agrupamientos de símbolos, con el fin de comprobar si éstos eran detectados por el algoritmo.

En primera instancia, se aplicó el algoritmo, utilizando el primer sistema de fuerzas. En la figura 1a, se presentan los resultados obtenidos.

Luego, se aplicó el algoritmo con el segundo par de fuerzas al mismo caso de estudio, obteniendo los resultados que aparecen en la figura 1b.



**Fig. 1. Distribución de nodos caso 1**

Cabe destacar que, tanto en éstas como en las siguientes figuras, no se muestran los arcos correspondientes a los vínculos entre los símbolos, para permitir una mejor visualización de la distribución de los nodos.

En el grafo presentado en la figura 1a no se observan agrupamientos. En la figura 1b, en cambio, se puede observar que algunos símbolos tienden a agruparse.

Un análisis semántico preliminar del léxico utilizado no mostró agrupamientos claramente definidos, aunque se distinguieron grupos de símbolos que podrían constituir clusters.

Para este caso de estudio, se contaba con los modelos de LEL y Escenarios desarrollados por 9 grupos de trabajo diferentes. Los grafos presentados en la figura 1 corresponden al LEL de uno de esos grupos.

Los modelos creados por cada grupo difieren en cuanto a los símbolos y Escenarios creados, en mayor o menor grado. Sin embargo, existen áreas de interés en el Universo de Discurso estudiado, que la mayoría de los grupos observó.

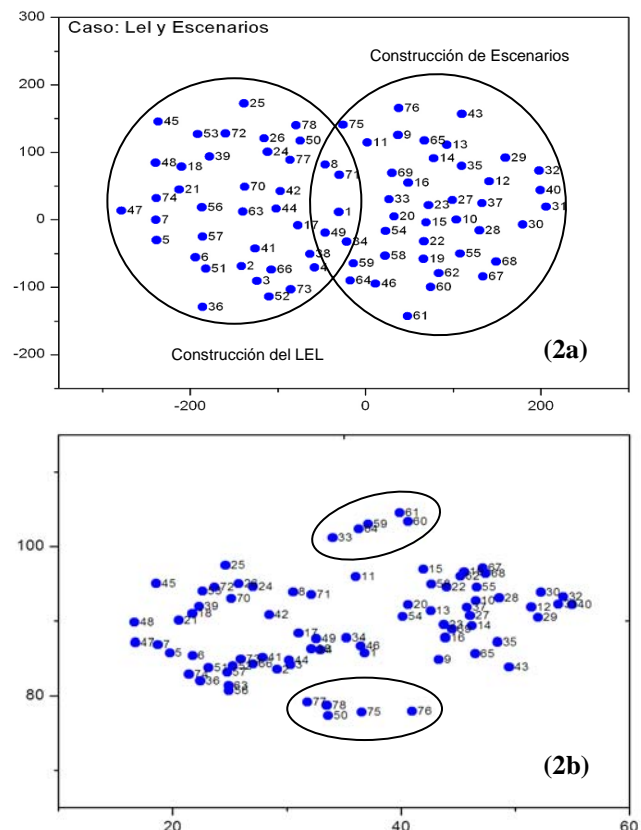
Por ello, se planea efectuar un análisis semántico detallado de los Léxicos disponibles para este caso de estudio, para comprobar si los agrupamientos que se observan en los grafos, efectivamente existen en el Universo de Discurso.

**Caso 2. LEL del proceso de Construcción de LEL y Escenarios [28].**

Para este caso de estudio, se conocía de antemano la existencia de clusters. En este caso, existen al menos dos agrupamientos: uno constituido por los símbolos correspondientes a la Construcción del LEL, y otro constituido por los símbolos correspondientes a la construcción de Escenarios.

En la figura 2a se presenta la distribución de los nodos obtenida utilizando el primer conjunto de fuerzas.

Se verifica que todos los símbolos que corresponden a la construcción del LEL se agrupan a la izquierda del grafo, mientras los que corresponden a la construcción de Escenarios se agrupan a la derecha. En la zona central, se ubican aquellos símbolos relacionados tanto con un proceso como con el otro. Por ejemplo, el nodo 8 corresponde al símbolo Cliente-Usuario, y el nodo 49 corresponde al símbolo Contexto del Problema.



**Fig. 2. Distribución de nodos para el caso 2**

En el grafo de la figura 2b, obtenido con el segundo sistema de fuerzas, se observa también que los símbolos correspondientes a la construcción del LEL tienden a agruparse de un lado, mientras los correspondientes a la construcción de escenarios, lo hacen al otro. Pero, en este caso, se distinguen además otros agrupamientos más

pequeños, que no se observaban en la figura anterior. Cada uno de estos grupos fue comprobado en el análisis semántico del caso de estudio.

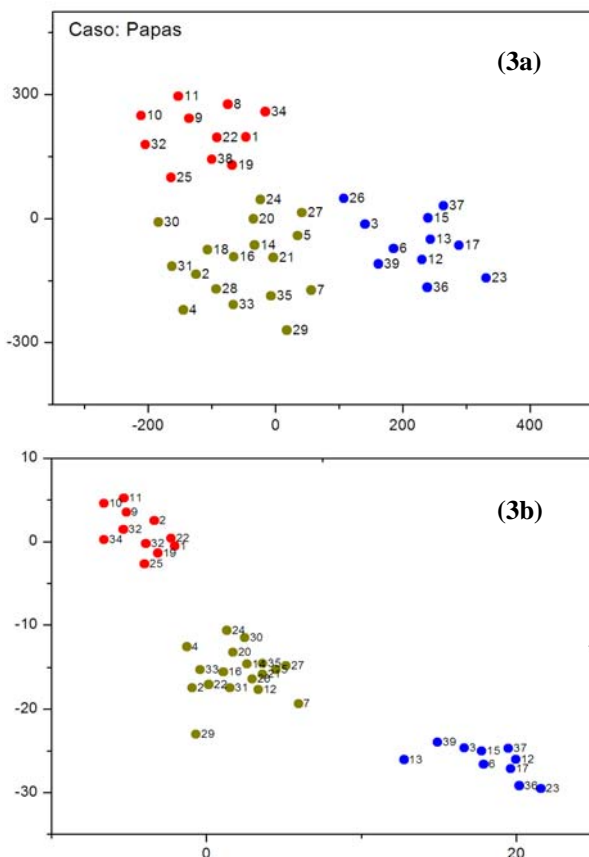
A modo de ejemplo, en la figura 2b se marcaron dos de los agrupamientos, y en la tabla 1 se presentan los nombres de los símbolos pertenecientes a los mismos.

**Tabla 1. Símbolos correspondientes a dos de los clusters visibles en la figura 2b.**

Vista de Hipertexto en los modelos de LEL y Escenarios	Extracción de los Requisitos del Sistema desde los Escenarios
50 Navegación	33 Escenario
75 Vínculo	59 Requisito del sistema
76 Vínculo a escenario	60 Requisito Funcional
77 Vínculo a símbolo del LEL	61 Requisito no Funcional
78 Vínculo arbitrario	64 Sistema de Software

**Caso 3. Relación entre productores de papa y una empresa acopiadora.**

Un análisis semántico preliminar de este caso de estudio, determinó la existencia de dos agrupamientos de símbolos; un grupo correspondiente a Producción y Entrega de Papas y otro a Canje de Semillas.



**Fig. 3. Distribución de nodos para el caso 3**

Sin embargo, en la visualización del grafo luego de la aplicación del algoritmo, tanto con el primer sistema de fuerzas, como con el segundo, se observa claramente la presencia de tres agrupamientos, como puede verse en las figuras 3a y 3b respectivamente.

Con el fin de detectar los límites de los clusters visualizados, se aplicó el clasificador K-Means [29] a los datos obtenidos, indicando la presencia de 3 grupos. En la figura 3 se muestran los clusters detectados con diferentes

colores, y en la tabla 2, se presentan algunos de los símbolos pertenecientes a cada agrupamiento.

**Tabla 2. Fragmento de la lista de símbolos correspondientes a cada uno de los clusters**

<b>Producción y Entrega</b>
2 Bonificación
4 Causa justificada
5 Contrato de adquisición y producción
14 El Productor
16 Recibir papa
18 Establecer Programa de Entregas
20 Inspeccionar papa
21 MC S.A.
24 Pagar por la papa
...
<b>Canje de Semilla</b>
3 Cancelar la operación de canje
6 Contrato de canje de semilla
12 División de Semillas MC S.A.
13 El Comprador
15 El Vendedor
17 Entregar semillas
23 Orden de carga
26 Papa Grado MC Consumo
37 Semilla de Papa
...
<b>Calidad</b>
1 Agrietadura
8 Corazón hueco
9 Defecto de calidad
10 Defecto externo
11 Defecto interno
19 Grado MC Consumo
22 Malformación
25 Papa chica
38 Tolerancias establecidas
...

Un análisis semántico más profundo del caso de estudio, confirmó la presencia de estos tres agrupamientos. En particular, el tercer grupo, correspondiente a Calidad de la Papa, no había sido detectado en el análisis inicial, ya que los símbolos incluidos en él, se habían considerado parte del cluster correspondiente a Producción y Entrega de Papa.

Es importante destacar que estos resultados realzan la capacidad predictiva de la estrategia, ya que, más allá de confirmar la presencia de agrupamientos conocidos, permite descubrir la presencia de otros.

**Conclusiones y trabajo futuro**

Se ha propuesto un algoritmo que permite detectar agrupamientos de símbolos en un modelo del proceso de Requisitos: el Léxico Extendido del Lenguaje. El análisis de los resultados obtenidos hasta el momento, permite concluir que es posible efectuar una segunda lectura de algunos de los documentos de Ingeniería de Requisitos mediante el estudio de las estructuras semánticas subyacentes. El uso de grafos y el análisis de su estructura son herramientas apropiadas para dichos estudios.

Se ha comprobado que las fuerzas de atracción y repulsión que se apliquen a los nodos, inciden en la calidad de los resultados. Por ello, se propone considerar la aplicación de otros sistemas de fuerzas con el fin de verificar si es posible obtener una mejor visualización de los agrupamientos.

Los resultados obtenidos se acoplan en forma muy natural con estrategias genéricas de clusterización como K-Means. El



conocimiento previo del número de clusters es imprescindible para el uso de esta estrategia. Por ello, uno de los objetivos a lograr es la determinación del número de clusters detectados.

Entre otras posibilidades, se plantea estudiar la simetría existente en los grafos resultantes de la aplicación del algoritmo, así como la relación entre las variancias sobre los ejes X e Y. Esto permitiría descartar la presencia de clusters en aquellos casos donde exista simetría en los ejes X e Y, así como variancias similares sobre ambos ejes. Del mismo modo, la simetría o asimetría, según sea el caso, en forma conjunta con las variancias, podrían contribuir a la determinación del número de agrupamientos.

Se planea analizar, también, si la ubicación inicial de los nodos puede afectar la disposición final por el efecto que determinado nodo pudiera causar sobre sus vecinos, evitando su normal desplazamiento.

Además, se pretende aplicar el algoritmo a otras formas de representación del grafo. Por ejemplo, se planea estudiar la representación en tres dimensiones.

Se prevé avanzar en la correlación entre los resultados obtenidos mediante la aplicación de este algoritmo, con un análisis semántico detallado de cada uno de los casos de estudio. Para ello, se propone estudiar semánticamente los léxicos desarrollados por diferentes grupos de trabajo para un mismo caso de estudio, con el fin de detectar los agrupamientos semánticos detectados por cada grupo. Los resultados del algoritmo presentado serán un complemento muy útil para dicho estudio, ya que permitirán visualizar los agrupamientos que se observen en la lectura de los modelos.

## Formación de Recursos Humanos

El grupo de investigación que participa en el proyecto Bases de Datos y Procesamiento de Señales está compuesto por 7 integrantes, todos ellos docentes de la Fac. de Ciencias Exactas categorizados en el Programa de Incentivos.

La línea de I/D forma parte de la tesis doctoral de la integrante del grupo Marcela Ridao, para el doctorado en Ciencias Informáticas de la Universidad Nacional de La Plata. El tema de la tesis es Técnicas Cuantitativas Orientadas al Reuso Semántico de Modelos de Requisitos.

## Referencias

- Barabasi, A.: *Linked, The New Science of Network..* Perseus publishing (2002)
- Dorogovtsev, S., Mendes, J.: *Evolution of networks: From biological nets to the Internet and WWW.* Oxford University Press, Oxford (2003)
- Arango, G., Schafer, W., Prieto, R.: *Domain Analysis Methods – Software Reusability.* Ellis Horwood Ltd (1993)
- Leite, J., Franco, A.: *O Uso de Hipertexto na Elicitação de Linguagens da Aplicação.* En: IV Simpósio Brasileiro de Engenharia de Software, SBC, pp. 134-149. Brazil (1990)
- Ben Achour, C., Rolland, C., Maiden, N., Souveyet, C.: *Guiding Use Case Authoring: Results of an Empirical Study.* In: International Symposium on Requirements Engineering, pp. 36-43. IEEE Computer Society Press, Limerick-Ireland (1999)
- Rolland, C., Ben Achour, C.: *Guiding the construction of textual use case specifications.* *Data & Knowledge Engineering* 25, 125-160 (1998)
- Oberg, R., Probasco, L., Ericsson, M.: *Applying Requirements Management with Use Cases.* Rational Software Corporation (1998)
- Regnell, B.: *Requirements Engineering with Use Cases – a Basis for Software Development,* Doctoral Thesis. Department of Communication Systems. Lund University (1999)
- Prakash, S., Aurum, A., Kox, K.: *Requirements Engineering Practice in Pharmaceutical and Healthcare Manufacturing.* In: 11th Asia-Pacific S.E. Conference, pp. 402-409 (2004)
- Leite, J., Franco, A.: *A Strategy for Conceptual Model Acquisition.* In: IEEE International Symposium on RE, pp. 243-246. IEEE Computer Society Press (1993)
- Leite, J., Rossi, G., Balaguer, F., Maiorana, V., Kaplan, G., Hadad, G., Oliveros, A.: *Enhancing a Requirements Baseline with Scenarios Requirements Engineering Journal* 2(4), 184-198 (1997)
- Leite, J., Doorn, J., Kaplan, K., Hadad, G., Ridao, M.: *Defining System Context Using Scenarios.* In: Leite, J., Doorn, J (eds.) *Perspectives on Software Requirements.* Kluwer Academic Press, pp. 169-199 (2004)
- Gross, J., Yellen, J.: *Editors. Handbook of Graph Theory.* CRC Press (2003)
- Gross, J., Yellen, J.: *Editors. Graph Theory and Its Applications, Second Edition (Discrete Mathematics and Its Applications).* Chapman & Hall/CRC (2006)
- Di Battista, G., Eades, P., Tamassia, R., Tollis, I.: *Graph Drawing: Algorithms for the Visualization of Graphs.* Prentice Hall (1999)
- Brandes, U., Kenis, P., Wagner, D.: *Communicating centrality in policy network drawing.* *IEEE Transactions on visualization and computer graphics* 9(2), 241-253 (2003)
- Fruchterman, T., Reingold, E.: *Graph Drawing by Force-directed Placement.* *Software-Practice and Experience* DOI 10.1002/spe.4380211102 Wiley Online Library (2006).
- Kaufmann, M., Wagner, D.: (eds.) *Drawing graphs: methods and models,* LNCS, vol 2025. Springer-Verlag (2001)
- Walshaw, C.: *A multilevel algorithm for force-directed graph-drawing.* *Journal of Graph Algorithms and Applications* 7(3), 253-285 (2003)
- Aiello, A., Silveira, A.: *Trazado de grafos mediante métodos dirigidos por fuerzas: revisión del estado del arte.* Tesis de Licenciatura en Ciencias de la Computación (2004)
- Eades, P.: *A heuristic for graph drawing.* *Congressus Numerantium* 42, 149-160 (1984)
- Kamada, T., Kawai, S.: *An algorithm for drawing general undirected graphs.* *Information Processing Letters* 31, 7-15 (1989)
- Davidson, R., Harel, D.: *Drawing graphs nicely using simulated annealing,* *ACM Transactions on Graphics* 15(4), 301-331 1996.
- Chernobelskiy, R., Cunningham, K., Goodrich, M. T., Kobourov, S. G., Trott, L.: *Force-directed Lombardi-style graph drawing,* *Proc. 19th Symposium on Graph Drawing,* pp. 78-90. 2011
- Kobourov, Stephen G.: *Spring Embedders and Force-Directed Graph Drawing Algorithms,* 2012.
- Bannister, M. J., Eppstein, D., Goodrich, M. T., Trott, L.: *Force-directed graph drawing using social gravity and scaling.* *Proc. 20th Int. Symp. Graph Drawing.* 2012.
- Rivero, L., Doorn, J., del Fresno, M., Mauco, V., Ridao, M., Leonardi, C.: *Una Estrategia de Análisis Orientada a Objetos basada en Escenarios: Aplicación en un Caso Real.* En: WER'98 - Workshop en Engenharia de Requisitos, pp. 79-90. Maringá, Brasil (1998)
- García, O., Gentile, C.: *Diseño de una herramienta para construcción de LEL y Escenarios,* Graduation dissertation. Universidad Nacional del Centro de la Pcia. de Bs. As. (1999)
- Peña, J., Lozano, J., Larrañaga P.: *An empirical comparison of four initialization methods for the K-Means algorithm.* *Pattern Recognition Letters* 20, 1027-1040 (1999)



## Reuso de Conocimiento en Foros de Discusión Técnicos

Nadina Martinez Carod, Gabriela Aranda, Alejandra Cechich,  
Pamela Faraci, Carina Noda, Mauro Sagripanti

Grupo de Investigación en Ingeniería de Software del Comahue (GIISCo)

<http://giisco.uncoma.edu.ar>

Facultad de Informática. Universidad Nacional del Comahue

Buenos Aires 1400, (8300) Neuquén

Contacto: {nadina.martinez, gabriela.aranda, alejandra.cechich}@fai.uncoma.edu.ar

### Resumen

Los técnicos informáticos suelen utilizar frecuentemente los foros de discusión en Internet para realizar una pregunta sobre un problema o una dificultad para llevar a cabo alguna tarea en particular. Dicha pregunta suele ser respondida por uno o más usuarios, proponiendo una serie de soluciones al problema en cuestión. Bajo este prisma, los foros de discusión técnicos se pueden considerar como plataformas colaborativas donde es posible plantear dudas y compartir soluciones probadas que pueden ser reutilizadas por otras personas en situaciones similares.

Sin embargo, la dinámica de Internet permite que muchas preguntas y respuestas similares, o al menos relacionadas al mismo ítem planteado, se encuentren diseminadas en diversos foros de discusión. Ante la necesidad de encontrar una solución para resolver un problema en particular, un técnico necesitar navegar por varios foros y conversaciones hasta descubrir el lugar donde el problema tratado es el más semejante a su dilema, y encontrar una solución que pueda satisfacerle.

A partir de este escenario, nuestro proyecto tiene como objetivo la definición de un modelo de calidad que pueda ser aplicado a foros de discusión técnicos así como la generación de conocimiento procedente de la información contenida en ellos, complementado con la implementación y puesta en marcha de un portal donde dicho modelo sea utilizado para facilitar la búsqueda de soluciones a problemas técnicos recurrentes.

### Palabras Clave

Foros de discusión, Modelos de calidad, Gestión de Conocimiento.

### Contexto

La línea presentada se inserta en el contexto de los siguientes proyectos y acuerdos de cooperación:

- UNCo-FAI: “Reuso de Conocimiento en Foros Técnicos”, subproyecto del Programa “Desarrollo Orientado a Reuso” que se encuentra en proceso de acreditación por la Universidad Nacional del Comahue.
- PAE-PICT-2312: Métodos y herramientas para sistemas masivamente distribuidos.

Investigaciones conjuntas con  
ISISTAN-UNICEN

- Acuerdo de cooperación con el Grupo Alarcos, Escuela Superior de Informática, Universidad de Castilla-La Mancha, Ciudad Real, España

## Introducción

El reuso del conocimiento ha ido evolucionando a lo largo de las últimas décadas. En particular, en las organizaciones actuales existe una tendencia cada vez más marcada a reutilizar el conocimiento, definiendo estrategias para el reuso de soluciones ya probadas en casos de problemas recurrentes[8].

Más aún, en los últimos años se están produciendo cambios importantes en la disciplina, debido a los avances en las tecnologías Web y al crecimiento desmedido (y muchas veces solapado en el tiempo) de las fuentes de información en el mundo entero.

En este contexto, si bien la Web abre las barreras físicas gracias a su naturaleza pervasiva, hay una serie de desafíos que necesitan ser considerados [9]:

- Las barreras cognitivas parecen cada vez mayores en lo que respecta a la “confiabilidad” del conocimiento que se encuentra en la web, esto es debido a que al ser de naturaleza libre no existen restricciones en la información, la cual en un gran número de veces no es debidamente verificada.
- La distribución desmedida en un ambiente sin límites, hace difícil la “usabilidad” (que sea fácil de usar), y por consiguiente la “utilidad” (que sea útil), de la información o del conocimiento

subyacente, enfocándose en el desafío referente a la captura y recuperación de información precisa y actualizada.

Teniendo en cuenta dichos desafíos, se puede considerar a la Web actual como un punto de encuentro de ideas, que posibilita la creación de debates técnicos y científicos mediante varias herramientas colaborativas (foros de discusión, blogs, wikis, etc.), y que son, en cierta forma, una nueva y exhaustiva forma de revisión de conocimiento para la comunidad participante.

De las herramientas mencionadas, nuestro proyecto se enfoca en los foros de discusión, cuya característica principal es su capacidad para:

- permitir preguntar sobre un problema particular o pedir opiniones en ambientes informales,
- compartir conocimiento creado por una comunidad de aprendizaje

En los foros de discusión las opiniones y soluciones vertidas por los distintos usuarios pueden ser reutilizadas tantas veces como sea necesario. Cuando un usuario tiene un problema en especial y no encuentra una conversación directamente relacionada con él en el foro, escribe una nueva entrada (pregunta o planteo de un problema), que en general, es un pedido de aporte de soluciones. Luego, las respuestas (propuestas) de diferentes participantes van formando un hilo (thread) de discusión. Dichos hilos de discusión se convierten en una fuente de información que puede ser reutilizada en situaciones similares ya sea por los mismos técnicos que la han generado o por otros que acceden al foro en busca de respuestas.

Sin embargo, dada la naturaleza de Internet, muchas preguntas y respuestas similares se encuentran diseminadas en distintos foros de discusión, por lo que generalmente es necesario navegar por varios foros hasta dar con la solución correcta.

Nuestro proyecto tiene como objetivo principal favorecer el reuso de la información contenida en dichas conversaciones existentes en la Web.

De manera complementaria, nuestro proyecto planea aprovechar el trabajo de investigación realizado sobre foros de discusión públicos para proponer técnicas en pos de mejorar la enseñanza en las materias específicas del área de programación. Este último objetivo está proyectado cumplirse bajo la participación de los alumnos de nuestras materias en los foros de discusión de la Plataforma de Educación de la Universidad Nacional del Comahue (PEDCo). De esta manera los alumnos podrán mejorar su aprendizaje utilizando el conocimiento significativo producto de la información analizada

Una característica a destacar de nuestro proyecto es que esta relacionado con más de un área de investigación. Esta apertura nos permite trabajar en forma colaborativa e interdepartamental enriqueciendo las áreas involucradas. Por ejemplo, más allá de nuestra participación en el grupo de investigación de Ingeniería del Software, es muy interesante el intercambio realizado con los investigadores del Departamento de Teoría de la Computación, relacionados al estudio de algoritmos de análisis de lenguaje natural y sentiment analysis.

En cuanto a trabajos relacionados con propuestas de reuso de conocimiento en foros de discusión, debemos mencionar la de un sistema recomendador para

conocimiento colaborativo, desarrollado por Chen et al [2], que analiza automáticamente los mensajes de un foro de discusión de un curso de Inteligencia Artificial, en el objeto de proponer mensajes de contenido similar a aquellos de estudiantes de dictados anteriores del mismo curso. Respecto a los puntos en común con nuestro trabajo, esta propuesta evalúa la relevancia de los mensajes existentes en función de su similitud con el mensaje recientemente ingresado por un estudiante. El enfoque de Helic y Scerbakov [4] propone un método de clasificación de los mensajes de un foro de discusión de acuerdo a una jerarquía de temas. La principal diferencia entre nuestra perspectiva y las propuestas de Chen et al y Helic et al, es que mientras ambos fueron pensados para un dominio de aprendizaje colaborativo, nuestro recomendador apunta a un contexto más amplio, involucrando usuarios con distinto conocimiento previo del tema (background). Finalmente, y más importante, en los trabajos mencionados el foro utilizado es único, por lo tanto se puede asegurar que la información a analizar se encuentra en un formato estándar, contrariamente a nuestra propuesta la cual recolecta información de diferentes foros, por lo cual la heterogeneidad de formatos de la información a capturar es un desafío extra.

## **Líneas de investigación y desarrollo**

Este proyecto de investigación está enmarcado dentro del programa de investigación “Desarrollo de Software Basado en Reuso” conducido por el grupo GIISCo. El objetivo de dicho grupo es trabajar para ofrecer soporte en investigación y transferencia de tópicos

relacionados con la Ingeniería de Software.

Además de la línea de este subproyecto, que recibe el nombre de “Reuso de Conocimiento en Foros de Discusión Técnicos”, el programa actualmente desarrollado por el Grupo GIISCo aborda distintos aspectos del Reuso, en otras dos grandes líneas (Reuso Orientado al Dominio y Reuso Orientado a Servicios).

Específicamente dentro del subproyecto de Reuso de Conocimiento en Foros de Discusión Técnicos, se destacan dos líneas de investigación relacionadas: La primera está enfocada en la captura, análisis y procesamiento de la información disponible en foros de discusión técnicos y un modelo de calidad relacionado que permita reutilizar el conocimiento disponible en Internet por parte de la comunidad de técnicos informáticos. La segunda línea está enfocada en la aplicación del conocimiento adquirido como parte de nuestra investigación, para hacer uso de la información de los foros de discusión de la Plataforma de Educación a Distancia de nuestra Universidad (PEDCO) y aprovecharla para proponer mejoras en la enseñanza de los alumnos de nuestra Facultad.

## Resultados y objetivos

En [10] se ha presentado la visión de nuestra línea de investigación basada en el reuso como una actividad sistemática de ingeniería. Durante el año 2012, nuestro grupo GIISCo ha avanzado en el tratamiento del desarrollo de software basado en reuso desde los pilares de servicios, de productos, de datos y reuso de información, en colaboración con investigadores de ISISTAN, UNICEN; *reuso de conocimiento*, abordando el tema de información en foros técnicos.

Parte del trabajo realizado hasta el momento es la definición de un modelo de calidad para foros de discusión técnicos (basado en estándares para la calidad de datos software [1] y otros modelos de calidad de datos en la Web [3][6][7]), aplicable a foros no restringidos disponibles en Internet. Dicho modelo está en proceso de ser publicado. Además, a partir de dicho modelo se ha comenzado a proponer un conjunto de métricas de calidad en cuya validación se trabajará a lo largo del proyecto.

De manera complementaria y paralela, se planea extender esta investigación al reuso de conocimiento en foros de discusión privados (de uso restringido a una comunidad en particular), específicamente en el contexto de la enseñanza de lenguajes de programación. En este sentido, uno de los objetivos de nuestro proyecto será proponer técnicas para mejorar la enseñanza en las materias del Departamento de Programación, a partir de la aplicación de técnicas de captura de información y generación de conocimiento desde los foros de discusión de la Plataforma de Educación a Distancia PEDCO, utilizada en las materias de dicho Departamento.

La información acumulada a partir de la interacción de los alumnos en los foros de discusión de la plataforma será analizada para generar conocimiento que pueda ser utilizado para mejorar el aprendizaje de los mismos alumnos en los siguientes cursos y de otros alumnos en el mismo curso. En este objetivo se combinan (en un nuevo dominio relacionado al campo de la gestión del conocimiento), experiencias obtenidas durante la ejecución del proyecto anterior, mediante la aplicación de técnicas cognitivas, conocidas como Modelos de Aprendizaje, que clasifican a las personas de acuerdo a



la manera que ellas perciben y procesan la información. Dichos modelos serán muy útiles para el objetivo de este proyecto, en la línea de investigación referida a la enseñanza de programación en la Facultad de Informática de nuestra Universidad.

## Formación de recursos

### humanos

El proyecto inicia una nueva línea de investigación respecto a la definición de un modelo de calidad y de gestión de conocimiento a partir de información contenida en foros de discusión técnicos. Para llevar a cabo estas metas, el proyecto está conformado por dos docentes del Departamento de Programación que han concluido en el año 2009 su Doctorado en Informática, y a dos docentes con dedicación simple, del Departamento de Ingeniería de Sistemas y de Programación que comienzan a formarse en investigación. También involucra a dos estudiantes que están desarrollando sus tesis de Licenciatura en Ciencias de la Computación. Además, se cuenta con la colaboración de una docente del Departamento de Teoría de la Computación de la misma Facultad, que está desarrollando su tesis de Doctorado sobre técnicas de análisis de lenguaje natural y provee asesoramiento sobre algoritmos de aprendizaje automático. Finalmente se cuenta con la asesoría externa de una docente e investigadora de la Universidad de Castilla La Mancha, lo que permite asociar desarrollos y producciones entre ambas universidades. De esta manera el proyecto está conformado por un grupo interdisciplinario garantizando la pluralidad de puntos de vista y la cooperación entre las partes.

## Referencias

- [1] ISO/IEC 25012:2008, Software product Quality Requirements and Evaluation (SQuaRE): Data quality model. 2008.
- [2] Weiqin Chen, Ricard Persen (2009), "A Recommender System for Collaborative Knowledge".
- [3] Coral Calero, Angélica Caro, Mario Piattini (2008), "An Applicable Data Quality Model for Web Portal Data Consumers", World Wide Web, vol. 11, no. 4, pp. 465-484.
- [4] D.Helic, N. Scerbakov (2003), "Reusing Discussion Forums as Learning Resources in WBT Systems".
- [5] J. Dorn (2010), "Social Software (and Web 2.0)".
- [6] Irfan Rafique, Philip Lew, Maissom Qanber Abbasi, Zhang Li (2012), "Information Quality Evaluation Framework: Extending ISO 25012 Data Quality Model", International Journal of Computer and Information Sciences, vol. 6, no. 1.
- [7] Richard Y. Wang, Diane M. Strong (1996), "Beyond accuracy: What data quality means to data consumers", Journal of Management Information Systems, vol. 12, no. 4, pp. 5-33.
- [8] Smith y A. Duffy (2001), Re-using knowledge: why, what and where. En Proceedings de 2001 International Conference on Engineering Design, Glasgow.
- [9] P. Di Maio (2009), Toward Pragmatic Dimensions of Knowledge Reuse and Learning on the Web. Proceedings of I-KNOW'09 and I-SEMANTICS'09, Graz, Austria.
- [10] Alejandra Cechich, Agustina Buccella, Andrés Flores, Gabriela Aranda, Nadina Martínez Carod, Juan Luzuriaga, Rodolfo Martínez, Marcelo Moyano, Rafaela Mazalu, Adriana Martin, Martín Garriga, Natalia Huenchuman. Desarrollo Basado en Reuso, WICC 2012, XIV Workshop de Investigadores en Ciencias de la Computación, Posadas, 2012

# MODELADO ESPECÍFICO DE DOMINIO (DSM) PARA UNA FAMILIA DE APLICACIONES DE SOFTWARE DE SIMULACIÓN DE ECOSISTEMAS NATURALES

Urciuolo Adriana, Iturraspe Rodolfo, Moyano Ezequiel, Gel Matías  
Instituto de Desarrollo Económico e Innovación, UNTDF

Dir.: Onas 450, (9410) Ushuaia. Tierra del Fuego. Tel: ++54-2901-443533

[aurciuolo@untdf.edu.ar](mailto:aurciuolo@untdf.edu.ar), [riturraspe@untdf.edu.ar](mailto:riturraspe@untdf.edu.ar), [emoyano@untdf.edu.ar](mailto:emoyano@untdf.edu.ar), [mgel@untdf.edu.ar](mailto:mgel@untdf.edu.ar)

## Resumen

*El Dominio de aplicaciones de simulación de ecosistemas naturales (DSEN) presenta características propias que requieren de un manejo específico, tales como la complejidad inherente a la información ambiental y a la aplicación de modelos matemáticos para simular los fenómenos naturales. La evolución del software del dominio y los requerimientos del mismo: manejo de complejidad, facilidades de integración y variabilidad en la configuración de aplicaciones particulares hacen necesaria la utilización de técnicas apropiadas de ingeniería de dominio, que faciliten reuso y permitan elevar el nivel de abstracción de los lenguajes utilizados acercando al desarrollador a los conceptos del dominio.*

*El modelado específico del dominio (DSM) propone elevar el nivel de abstracción, expresando las soluciones directamente en términos del dominio. Utiliza lenguajes específicos de dominio en combinación con generadores de código y frameworks de dominio, pasando de especificaciones de muy alto nivel a código ejecutable. Provee verdaderos beneficios cuando el dominio es bien acotado y las posibles variaciones están previstas.*

*En el marco de la presente línea de investigación se propone utilizar DSM para el desarrollo de software de una familia de aplicaciones DSEN definida mediante un enfoque de reuso sistemático.*

**Palabras clave:** reuso, dominio específico, familia aplicaciones

## Contexto

La línea de investigación se desarrolla en el Instituto de Desarrollo Económico e Innovación (IDEI) de la UNTDF, por parte de un Grupo de docentes-investigadores que lleva adelante proyectos en la temática de Sistemas de Información Ambiental – Hidroinformática desde hace más de 10 años (acreditados por la UNPSJB previamente a la creación de la UNTDF). Actualmente esta línea se enmarca en el Proyecto recientemente iniciado: “Red para la conservación de Ecosistemas fluviales de Patagonia - CONICET (Feb/2013-Feb/2018 - coordinado por el CENPAT-CONICET, con nodos en Chubut, Neuquén y Tierra del Fuego) siendo coordinado el Nodo Tierra del Fuego por la UNTDF. El Proyecto se lleva adelante con un enfoque multidisciplinario por parte de organismos de investigación y gestión en la Patagonia y consta de diversos componentes. El grupo responsable de la presente línea de investigación desarrolla sus actividades en el Nodo Tierra del Fuego de la Red Ecofluvial, en los siguientes componentes del Proyecto: Sistema integrado de información de cuencas y Modelado de ecosistemas en cuencas patagónicas.

## Introducción

El estudio de los sistemas de información ambiental adquiere cada vez mayor importancia, dada la función esencial que los mismos cumplen en la toma de decisión para el manejo y aprovechamiento sustentable de los recursos naturales y el medio ambiente [5]. Las particularidades de este dominio hacen necesario un permanente estudio, análisis y reflexión acerca de metodologías de desarrollo de software adecuadas para modelar los ecosistemas naturales [8]. Aplicaciones típicas del dominio son las relativas a la modelación de procesos físicos (por ej. crecidas de ríos), como es el caso del software de modelos de simulación hidrológica, climática, calidad de aguas, etc. Existen varias generaciones de estos modelos, los cuales proveen medios de extrapolación cuantitativa y predicción de fenómenos naturales a partir de variables físicas y parámetros (por ej. modelos lluvia-caudal). Uno de los requerimientos básicos de estas aplicaciones es el *manejo de la complejidad*, debida principalmente a ciertas características de la información ambiental (datos espacio-temporales) y a la definición de escenarios de simulación con diferentes procesos naturales y métodos intercambiables. La integración de información proveniente de diferentes subdominios y la simulación de procesos físicos de la naturaleza a través de modelos matemáticos constituyen problemas que caracterizan estas aplicaciones y deben resolverse repetidas veces ante las diferentes situaciones naturales a simular. El software en muchos casos ha sido desarrollado y utilizado principalmente por expertos del dominio, con grandes dificultades para su extensión, su ajuste y/o adaptación a distintos escenarios, así como para la incorporación de nuevas o diferentes formas de simular un mismo proceso, sin posibilidad de reuso del esfuerzo realizado. Por otra parte los expertos no cuentan con facilidades para trabajar en un nivel de abstracción apropiado.

No obstante los problemas mencionados, el desarrollo de modelos reusables utilizando medios de especificación y lenguajes apropiados a las particularidades del software en el dominio de simulación de ecosistemas naturales (DSEN), aún es incipiente. La evolución del software del dominio y los requerimientos del mismo: manejo de complejidad, facilidades de integración y variabilidad en la configuración de aplicaciones particulares, hacen necesaria la utilización de técnicas apropiadas de ingeniería de dominio, que permitan la construcción de una familia de productos, los cuales compartiendo ciertos artefactos de software, puedan ser adaptados a diferentes configuraciones. Asimismo resulta necesario contar con lenguajes específicos del dominio que permitan a los desarrolladores del software trabajar con mayor nivel de abstracción.

Existen numerosas líneas de investigación en la actualidad en las cuales se analiza el problema de la complejidad de las aplicaciones ambientales, abordando la cuestión del reuso y modelado específico de dominio. En el pasado se han desarrollado frameworks, modelos conceptuales y microarquitecturas para el problema de la representación de objetos geográficos y campos continuos, así como modelos conceptuales y microarquitecturas OO para modelos de simulación hidrológica [10] [11] [12][13] que abordaban el problema de reuso de diseño. En cuanto al modelado apropiado al dominio se ha propuesto un Perfil UML 2 para aplicaciones de monitoreo ambiental [14]. No obstante, aún no se ha trabajado bajo el criterio de reuso sistemático [3] y son escasos los DSL para la modelación del comportamiento de ecosistemas naturales [1]

Los DSL (Lenguajes específicos de dominio) se han desarrollado en diversas áreas para facilitar la construcción de modelos en un nivel más cercano al conceptual a los expertos del dominio. El modelado de software con adecuadas herramientas para la definición de

los conceptos específicos del dominio es muy conveniente para el desarrollo de sistemas complejos. En DSM [4] los modelos se construyen usando conceptos que representan objetos del dominio de aplicación en lugar de conceptos de un lenguaje de programación determinado. El lenguaje de modelado sigue las abstracciones y semántica del dominio, permitiendo que los desarrolladores se perciban trabajando directamente con los conceptos del dominio. Para estos fines, DSM utiliza lenguajes específicos de dominio en combinación con generadores de código y frameworks de dominio, pasando de especificaciones de muy alto nivel a código ejecutable.

El desarrollo de una plataforma DSM requiere una considerable inversión de recursos, la necesidad de expertos del dominio, así como desarrolladores experimentados. Este trabajo se ve facilitado en la actualidad por cuanto existen entornos especializados que permiten desarrollar los lenguajes y generadores específicos de la solución DSM sobre la arquitectura que ofrecen (“Language Workbenches” s/Fowler)

Mientras un lenguaje de modelado tradicional tiende a ser general, un lenguaje DSM resulta lo más específico posible, elevando el nivel de abstracción de los modelos. Mientras más específico sea el dominio, mayor es el beneficio en productividad de aplicar DSM.

Al analizar la viabilidad de una solución DSM se debe considerar el beneficio devuelto durante su uso: cuántas variaciones de un producto permite y el número de características en que varían los productos permitidos. En el dominio DSEN, si bien pueden realizarse numerosas variaciones de un mismo producto, la gran dispersión de modelos de simulación de ecosistemas naturales existentes, puede disminuir la utilidad del DSL y las posibilidades de reuso. De allí la importancia de definir claramente una Familia de aplicaciones con un enfoque de reuso

sistemático. Las Líneas de Productos de Software (LPS) proveen reuso sistemático y un modo conveniente de diseñar e implementar una variedad de sistemas fuertemente relacionados. [3]. En el marco de este enfoque es posible trabajar en familias de aplicaciones que comparten características comunes, maximizando los beneficios de DSM [6].

El diseño sistemático de una arquitectura y un conjunto de recursos (assets) reusables constituyen la base para construir una familia de aplicaciones relacionadas [2]. Esta definición es importante para la posibilidad de acotar el trabajo en un dominio a un subconjunto del mismo, llamado “familia de aplicaciones”

Se debe tener en cuenta al adoptar este enfoque, que no es necesario construir todos los productos de una Línea de Productos de software, sino que una LPS puede pensarse como la definición de un conjunto virtual de productos, que comparten características comunes a ser construidos a partir del mismo conjunto de recursos centrales. Solo aquellos productos que satisfagan cierta necesidad serán construidos, mientras que los otros representan la capacidad potencial de construcción de productos. Así se espera desarrollar la arquitectura y recursos para una familia DSEN, implementando sólo aquellos componentes de interés al caso de estudio del Proyecto.

Los productos en una familia de aplicaciones DSEN pueden variar en términos de su comportamiento, configuración física, factores de escala, tipo de parámetros, etc.

En el marco de esta línea de investigación se espera avanzar en el desarrollo de los recursos centrales (arquitectura y componentes básicos) necesarios para la construcción de software en una familia de aplicaciones DSEN, generando además un Framework de aplicación y un lenguaje específico para el dominio que facilite a los expertos el desarrollo de nuevas aplicaciones.



## Líneas de Investigación y Desarrollo

La línea de investigación actual se enfoca en DSM para familias de aplicaciones DSEN. De acuerdo a lo expuesto, y a los fines de lograr los objetivos del proyecto, se hace necesario el estudio de diferentes tópicos vinculados al tema Ingeniería de Dominio, principalmente en materia de reuso sistemático y DSM. Se describen las líneas de investigación y desarrollo del grupo en estos dos aspectos:

1) Reuso sistemático: En relación e este tema, se plantea el estudio de métodos para las distintas etapas del Desarrollo de Líneas de Productos de Software [3]: Análisis de dominio, Modelado de la variabilidad, Desarrollo de arquitecturas referenciales para dominios específicos, Definición de recursos centrales, Definición familias de aplicaciones (dado que se trabajará con una familia específica)

En cuanto a desarrollo, se obtendrá en esta línea una arquitectura referencial (a partir del análisis de dominio), donde desarrollar los componentes principales de un modelo de simulación de cuenca básico (incluyendo simulación hidrológica), teniendo previstas distintas variaciones para futuros trabajos.

La arquitectura definida deberá entonces proveer facilidades para ser instanciada en aplicaciones de la familia DSEN seleccionada

2) DSM (Modelado específico de dominio)

Los temas centrales de estudio en este aspecto, para dar soporte a un alto nivel de abstracción, son: DSLs (lenguajes específicos de dominio), frameworks específicos de dominio y generadores de código [9].

En cuanto a desarrollo se prevé la creación de un ambiente integrado de desarrollo y de un DSL específicos para la familia de aplicaciones del dominio DSEN seleccionada, permitiendo se eleve el nivel de abstracción, ocultando a los desarrolladores la complejidad en implementación.

Se trabaja con expertos del dominio, para obtener el conocimiento necesario acerca de

métodos de simulación de procesos físicos y la validación continua de resultados.

## Resultados y Objetivos

Si bien la línea de investigación en el marco del Proyecto de Red aprobado por CONICET (febrero/2013) es incipiente, el grupo de investigación ya ha obtenido resultados de proyectos anteriores que constituyen un insumo para las actuales investigaciones. Entre ellos pueden citarse:

- Microarquitecturas de diseño OO para sistemas de modelación hidrológica [1].
- Patrones conceptuales para sistemas de información hídrica [2]
- Perfil UML2 para aplicaciones de Monitoreo ambiental [14]
- Análisis de dominio para sistemas complejos (Tesis de grado Lic. Ezequiel Moyano)
- Framework de soporte a aplicaciones de Glaciología (Tesina de grado en curso Matías Gel)

Los resultados mencionados constituyen aportes al conocimiento del dominio, sus características y avances en el desarrollo de una arquitectura referencial.

## Objetivo general

Contribuir al desarrollo de modelos y lenguajes específicos de dominio para una familia de aplicaciones DSEN que permitan incrementar productividad y reuso, así como elevar el nivel de abstracción de lenguajes utilizados.

## Objetivos específicos

- ✓ Adaptación de técnicas de análisis de dominio apropiadas para el modelado de sistemas que simulan el comportamiento de ecosistemas de la naturaleza.
- ✓ Definición de características comunes, ámbito y variabilidad para aplicaciones de simulación de ecosistemas naturales.
- ✓ Definición de la familia de aplicaciones de software DSEN y sus recursos centrales (core assets).

- ✓ Definición de una arquitectura referencial para aplicaciones DSEN, que considere el conjunto de variaciones permitidas para la familia de productos seleccionada.
- ✓ Desarrollo de componentes centrales para la implementación de productos de la familia de aplicaciones
- ✓ Definición de un lenguaje específico de dominio y de un Framework de aplicación para expertos del dominio.

La validación de resultados será realizada en la construcción de un modelo de simulación para la cuenca del río Grande, correspondiente al caso del estudio del Nodo Tierra del Fuego de la Red Ecofluvial.

### Formación de Recursos Humanos

En el marco de esta línea de investigación se han concluido dos tesis de grado de Licenciatura en Informática de integrantes del Grupo de investigación (Lic. Ezequiel Moyano, APU Matías Gel) y se está formulando la propuesta para la tesis del posgrado (Magister en Ing. de Software – UNLP) (Lic Ezequiel Moyano). Asimismo se encuentra en desarrollo la tesis de Doctorado en Ciencias Informáticas (UNLP) sobre Componentes de Dominio para aplicaciones ambientales

El grupo de investigación es multidisciplinario, por cuanto, si bien se consideran fundamentales las actividades relativas al análisis y diseño de sistemas, es necesario el aporte de expertos del dominio, a los fines obtener conocimiento del mismo.

### Referencias

- [1] Andrew Fall, Joseph Fall *A domain-specific language for models of landscape dynamics* Original Research Article *Ecological Modelling, Volume 141, Issues 1–3, 1 July 2001, Pages 1-18*
- [2] Aurangzeb Khan, Farooque Azam and Jahanzaib Khan. *Architecture Centric Development in Software Product Lines. International Journal of Computer Applications* 30(2):6-12, September 2011.

[3] Clements P. and Northrop. *Software Product Lines: Practices and Patterns*. Addison-Wesley, Boston, MA, 2001.

[4] Fowler M. *Domain-Specific Languages – USA* - Addison-Wesley Professional - 2011.

[5] Green D., Klomp N. *Environmental informatics - a new paradigm for coping with complexity in nature*. Complexity International Vol 6, 1998.

[6] Ilker Altintas N., Semih Cetin, Ali H. Dogru, Halit Oguztuzun, "Modeling Product Line Software Assets Using Domain-Specific Kits," *IEEE Transactions on Software Engineering*, vol. 38, no. 6, pp. 1376-1402, Nov.-Dec., 2012.

[8] Parent C., S. Spaccapietra E. Zimányi, *Spatio-temporal conceptual models: Data structures + space + time*. In C. B. Medeiros, editor, Proc. ACM-GIS, pp 26-33, 1999.

[9] Steven K. y Juha-Pekka, Tolvanen - *Domain-Specific Modeling Enabling Full Code Generation* - USA - IEEE Computer Society / John Wiley & Sons – 2008.

[10] Urciuolo Adriana, Iturraspe Rodolfo, Villarreal Martín. "Microarquitecturas de Diseño OO para Sistemas de Modelación Hidrológica". La Plata, Octubre de 2003. Publicado en *Proceedings del Congreso Argentino de Ciencias de la Computación 2003 – CACIC 03 (12 páginas)*.

[11] Urciuolo Adriana, Iturraspe Rodolfo. 2002. "Conceptual Patterns for Water Resources Information Systems". *Journal of Computer Science and Technology* Vol. 3 - No. 1 - April 2003 pp 20-26.

[12] Urciuolo A., R. Iturraspe, A. Parsón, "Conceptual Microarchitectures for Hydrologic Simulation Models", En: *CLEI Electronic Journal*, ISSN 0717-5000, Cecilia Bastarrica ed. <http://www.clei.cl>. Vol 7, 1, Jun 2004, paper 6, 18 pp.

[13] Urciuolo A., R. Iturraspe, "Conceptual Patterns for Water Resources Information Systems", En: *Journal of Computer Science and Technology* Vol. 3 - No. 1 - April 2003 - ISSN: 1666-6038, pp 20-26.

[14] Urciuolo Adriana, Iturraspe Rodolfo, Moyano Ezequiel. *Perfil UML 2.0 para Aplicaciones de Monitoreo Ambiental*. Actas de las VI JIISIC'07, 31 de Enero al 2 de Febrero del 2007, Lima, Perú. Facultad de Ciencias e Ingeniería and Departamento de Ingeniería, Pontificia Universidad Católica del Perú 2007, ISBN 978-9972-2885-1-7. pp. 393-401.

# “Aproximación a los Métodos Formales: una experiencia con un método de desarrollo liviano.”

Fernando Aguirre, Edgardo Belloni

Dirección de Investigación y Desarrollo - Departamento de Ingeniería y Cs. de la  
Producción - Universidad Gastón Dachary

Campus Urbano. Av. Lopez y Planes y Av. Jauretche. Posadas. Misiones

Tel. (0376) 447-6999

faguirre@ugd.edu.ar, ebelloni@ugd.edu.ar

## RESUMEN

El presente trabajo tiene como objetivo determinar cuáles son las bondades de la implementación de técnicas formales para el desarrollo de software, enfocándose en los métodos denominados semi formales, ya que éstos suponen una aplicación más al "alcance" de la industria, debido a su menor complejidad y costo. Para ello dos equipos desarrollarán un mismo componente de software. Uno de los cuales utilizará una técnica de desarrollo semi formal y se compararán los resultados, basados en la medición de las métricas previamente definidas.

Se utilizará la herramienta de Microsoft Spec#, y el lenguaje de programación C# de la Suite Visual Studio .Net.

## CONTEXTO

El presente proyecto se desarrolla en el ámbito de la Dirección de Investigación y Desarrollo de la Universidad Gastón Dachary,

con el aporte del Departamento de Ingeniería y Cs. de la Producción.

El proyecto fue presentado y aprobado en la convocatoria de Proyectos de Investigación del año 2.012, que fuera realizada por la mencionada Dirección.

El proyecto además se encuentra en consonancia con el objetivo de la tesis en realización de uno de los investigadores, para la obtención del título de la Maestría en Ingeniería de Software, dictada en la Universidad de La Plata, cuyo título es "Un aporte empírico al desarrollo formal liviano en base a Spec#"

## INTRODUCCIÓN

¿Cómo aseguramos que nuestros programas sean correctos y libres de errores?

A diferencia de otras disciplinas formales, la Ingeniería de Software no ha

desarrollado técnicas precisas para el modelado de los objetos a construir, antes que estos fueren efectivamente desarrollados. No existen actualmente, o al menos no ampliamente difundidos, técnicas tales como los modelos matemáticos, que permiten predecir el resultado de construcciones del mundo real con exactitud.

Ahora entonces, veamos con que mecanismos cuenta la Ingeniería del Software para evitar construcciones sin errores:

Lamentablemente todas las aproximaciones existentes, distan bastante de ser soluciones exactas. “No existe la bala de plata” [F. Brooks 87] reza Brooks, dado que el modo con que cuentan los informáticos para respaldar la confianza de los programas es la verificación empírica [JuristoMoreno 06]. Con lo cual, el principal problema consiste en la imposibilidad de garantizar la no existencia errores en una aplicación, dado que depende primero de la porción de programa que se esté probando y del grado de dependencia que tenga con otras porciones de programa (acoplamiento), y segundo, del grado cobertura de casos del universo de datos de pruebas suministrado.

Por lo tanto, usualmente se realiza la detección y corrección de errores en simultáneo con la construcción de la aplicación en sí. Tal actividad requiere un esfuerzo igual o superior al esfuerzo de la programación propiamente dicha.

Muchos errores están relacionados con la complejidad inherente del software, tales como errores en la definición del dominio, en la especificación de los flujos de trabajos, etc., para lo cual se requiere soluciones orientadas a atacar las dificultades esenciales del software.

Ahora bien, los demás tipos de errores, que son los que están más cercanos a las denominadas complejidades no esenciales, mayormente inherentes al trabajo de programación en sí, son los más atacados por las técnicas de validación y verificación de software.

Pretender generar una técnica de validación y verificación del software con una efectividad cercana al 100% hoy en día, constituye una empresa con un éxito altamente improbable, sin embargo, las líneas de investigación actuales consisten en generar aportes para la reducción del esfuerzo en las pruebas y lograr la conversión paulatina de las técnicas manuales y artesanales de validación a procesos cada vez más automáticos.

#### Verificación formal de software.

Los métodos formales en para verificación de software constituyen el intento de la Ingeniería de Software de parecerse a las demás ingenierías, proveyendo herramientas matemáticas para la detección y corrección de errores.

Constituye la principal referencia dentro de las técnicas de verificación estáticas. Implican una especificación formal del sistema y un análisis matemático detallado de la especificación y pueden desarrollar argumentos formales para que un programa se ajuste a su especificación formal. [HuthRyan 04]

#### Spec#, un aporte al desarrollo de software de alta calidad.

El sistema de programación Spec# provee un conjunto de herramientas para la verificación formal del software. Su principal ventaja es



que permite al programador abstraerse de las complejas especificaciones matemáticas, y por ello se considera como un sistema formal “liviano”.

Se fundamenta conceptualmente en la lógica de Hoare [Hoare 69], y es una extensión del lenguaje de programación C#, creado por Microsoft e incluido por primera vez en la Suite de Desarrollo Visual Studio 2005.

Spec# se encuentra en plena etapa de investigación, con algunos resultados prometedores. Consiste en estructuras de especificación tipo pre y pos condición, tipos “no nulos” y algunas otras funcionalidades para abstracciones de datos de alto nivel. [Microsoft 08]

## LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Esta línea de investigación se encuadra dentro de los estudios relacionados a la aplicación de los métodos formales en la industria de software.

Se centra en la aplicación de un método formal *semi liviano* para la construcción de un componente de software y cuáles son sus efectos en el desarrollo de software. Para ello se trabajará con distintos grupos de desarrolladores que implementarán las especificaciones del componente, por un lado utilizando una metodología tradicional y por otro lado, incluyendo técnicas de verificación semi formal.

Existen otros trabajos relacionados, como [R. Khun 98], el cual aporta un análisis del costo de implementación de los métodos formales, haciendo referencia a que los métodos formales pueden ser aplicados

productivamente mediante técnicas formales livianas, analizando especificaciones parciales de software o inclusive en los requerimientos, de forma temprana.

La línea principal de trabajo se centra en la aplicación real de técnicas formales livianas, utilizando el paradigma Orientado a Aspectos, mediante la herramienta de Microsoft Spec#.

## RESULTADOS OBTENIDOS/ESPERADOS

Se espera del proyecto: Aportar mayor evidencia empírica en la utilización de los métodos formales de verificación y validación de software con el objetivo de ratificar las bondades que brindan, aún teniendo en cuenta las limitaciones asociadas a que Spec# todavía se encuentra en etapas tempranas de evolución.

A la vez, se espera también aportar mecanismos y/o procedimientos, que contribuyan a facilitar y/optimizar la implementación de dicha herramienta colaborando en la difusión y masificación del uso de la misma.

## FORMACIÓN DE RECURSOS HUMANOS

La estructura del proyecto está principalmente conformada por dos docentes de la UGD, uno de los cuales se encuentra desarrollando la tesis de la Maestría en Ingeniería de Software de la Universidad Nacional de La Plata, que desarrolla su trabajo en el ámbito del proyecto, contribuyendo de esta manera a su formación académica.

El equipo del proyecto se complementa con la incorporación de Auxiliares de Investigación.

Se incorporarán al proyecto 1 (un) Auxiliar de Investigación de 1ª, quien preferentemente se encontrará realizando su tesis para la obtención de título de grado y 4 (cuatro) Auxiliares de Investigación de 2ª.

Se pretende que los demás auxiliares generen sus tesis de grado en el ámbito del proyecto.

## REFERENCIAS

[F. Brooks 87] F. Brooks. *No Silver Bullet, Essence and Accidents of Software Engineering*. Computer Magazine of University of North Carolina at Chapel Hill. Abril 1987.

[JuristoMoreno 06] N. Juristo, A. Moreno, A. Vegas. *Técnicas de Evaluación de Software*. Universidad Politécnica de Madrid. Octubre 2006.

[HuthRyan 04] M. Huth and M. Ryan. *Logic in Computer Science. Modeling and Reasoning about Systems*. Cambridge University Press. 2004.

[Hoare 69] Hoare, C. A. R. *An axiomatic basis for computer programming*. Communications of the ACM, 12(10):576-585, October 1969.

[Microsoft 08] Microsoft Research. *Program Verification Using the Spec# Programming System*. ETAPS Tutorial. Marzo 2008.

[Woodcock 99]. Woodcock, P. Gorm Larsen, J. Bicarregui, J. Fitzgerald., *Formal Methods: Practice and Experience*. ACM Computer Survey 41(4). 2009.

[R. Kuhn 03] R. Kuhn, D. Craigen, M. Saaltink. *Practical Application of Formal Methods in Modeling and Simulation*. 2003

[JacksonWing96] D. Jackson and J. Wing, *Lightweight Formal Methods* IEEE Computer, Abril 1996, pp. 21-22

[M. Barnett 11] M. Barnett, M. Fahndrich, K. Rustan, M. Leino, P. Muller, W. Schulte, H. Venter, *Specification and Verification: The Spec# Experience*. 2011

[R. Khun 98] R. Kuhn, R. Chandramouli, R. Butler. *Cost Effective Use of Formal Methods in Verification and Validation*. 1998

# Validación de Especificaciones No Funcionales de Aplicaciones Web a Través de Técnicas de Testing de Usabilidad

Esp. Ing. Juan Carlos Moreno, Mgter. Ing. Marcelo Martín Marciszack  
Dpto. de Ing. en Sistemas de Información / Facultad Regional Córdoba / Universidad Tecnológica Nacional  
Dirección: Av. Marcelo López esq. Cruz Roja Argentina – Córdoba – Argentina  
E-mails: { marciszack, jmoreno33}@gmail.com

## Resumen

La Calidad en el desarrollo del software es hoy en día un hecho, que no se puede dejar de lado, cuando se desarrolla un producto. La usabilidad es una de las características importantes a tener en cuenta, si queremos brindar calidad en nuestros productos. La usabilidad pasa a ser un atributo fundamental, cuando nuestro software interactúa con humanos y compite con otros productos en el mercado, sobre todo cuando se encuentra en un entorno web. Si un sistema web es altamente usable, harán más productiva y eficiente las tareas, y se verá reflejado en rentabilidad económica para la organización. Al mismo tiempo los usuarios aceptarán favorablemente a los sistemas web. El trabajo presentado en este artículo, tiene como objetivo el desarrollo de una propuesta metodológica, que permita gestionar y validar especificaciones no funcionales para aplicaciones web, a través de técnicas de testing de usabilidad.

**Palabras clave:** Especificaciones No Funcionales, Usabilidad Web, Validación de Requerimientos No Funcionales.

## Contexto

El presente proyecto se encuentra circunscripto dentro del ámbito de Investigación de los Sistemas de Información en el Dpto. de Ingeniería en Sistemas de Información de la Universidad Tecnológica Nacional de la Facultad Regional Córdoba. A través del mismo se busca desarrollar y proponer una manera metodológica e ingenieril de validar la usabilidad en el desarrollo software orientado a entornos web, aplicado desde el punto de vista de la ingeniería de requisitos no formales. Ésto permitiría brindar cierto grado de calidad al proceso de desarrollo del software y al mismo tiempo asegurar la construcción de software más útil y aceptable a nivel humano.

Esta investigación se lleva a cabo dentro del ámbito académico Universitario, formando parte de un proyecto de investigación, denominado “Validación

de Requerimientos a Través de Modelos Conceptuales”, la cual contempla los requerimientos funcionales. Este proyecto es parte del desarrollo de un proyecto de I+D consolidado, donde se busca resolver problemas relacionados a la elicitación y especificación de requerimientos que vincula las distintas etapas del proceso de desarrollo del software. De esta forma se contempla el desarrollo del software en forma integral a través de todas sus etapas.

## Introducción

Diariamente se desarrollan nuevos paquetes de software, que usan como soporte toda la infraestructura y tecnología de la red de Internet. Estas aplicaciones, comúnmente se denominan aplicaciones Web. Se caracterizan por ser interactivas, y muchas de ellas implementan procesos y técnicas de seguridad adecuados para el intercambio de información en entornos web seguros, con usuarios ubicados en cualquier lugar del planeta. Si a lo anterior se le suma: el avance tecnológico en las comunicaciones, el uso masivo de computadoras, el incremento de la demanda de aplicaciones web, vemos que el estudio de la interacción del software web con los usuarios, pasa a ser relevante y debe ser tenido en cuenta en la Ingeniería del Software. El estudio de esta característica se denomina usabilidad web. La cuestión se suscita debido a que en la construcción del software, muchas veces no se trata de forma adecuada la interacción del software con el usuario final, y se hace más énfasis en los aspectos de la arquitectura y procesos internos del mismo.

Con el avance de Internet y de las aplicaciones desarrolladas para el entorno web, la demanda de software más usable ha ido en crecimiento. Este incremento en la demanda de aplicaciones web, no solo ha traído como consecuencia el desarrollo de nuevas características multimediales (especialmente diseñadas para la interacción humana), sino también de aspectos críticos que tienen que ver con la seguridad y el intercambio de información, que ayudan a la

confianza e incentivan el uso de los sitios web. Detrás de estos sitios web, siempre está la imagen de una organización que desea brindar fiabilidad, transparencia y seguridad a sus potenciales clientes, a través del entorno web. De este modo la organización ahorra costos, se expande geográficamente, incrementa sus operaciones. Como contrapartida el medio exige una gran organización interna y sitios Web más usables que respondan a sus necesidades. Surge como una necesidad la Ingeniería Web, cuyo objetivo ha sido construir aplicaciones para sistemas basados en la web, con calidad y aplicando principios de la ingeniería del software (Sistemas de negocios críticos basados en el desarrollo web) [22]. Muchas veces la calidad de las aplicaciones web, y por lo tanto de la usabilidad web, ha sido evaluada basándose en el sentido común, experiencia y lógica, que surge de la experiencia de los desarrolladores [1].

Se exige entonces, que los sistemas sean más usables y fáciles de aprender, teniendo en cuenta la calidad inherente del software [10], como así también la calidad de uso [12] del mismo en cuanto a eficiencia y eficacia para la consecución de las tareas. Esto adquiere mayor relevancia cuando hablamos de sistemas que interactúan con humanos, que es lo que sucede en internet, cada vez con mayor frecuencia. Según Offutt, uno de los tres criterios principales que conduce el desarrollo de aplicaciones Web es la usabilidad [19].

Como consecuencia de lo anterior, se debe definir, como primer paso, que es usabilidad. El concepto de usabilidad ha sido definido por varias normas de Organizaciones Internacionales de Estándares de Calidad como ISO y la IEEE. En cada norma se presenta a la usabilidad como un atributo del software y está relacionado a la calidad del mismo.

En el estándar ISO 9241-11, se hace referencia a la usabilidad como “la medida en que un producto se puede usar por determinados usuarios para conseguir objetivos específicos con efectividad, eficiencia y satisfacción en un determinado contexto de uso” [12].

La norma ISO/IEC 9126-1 define a la usabilidad como un parámetro de calidad del software y conforma una de las características de calidad relevantes del producto. En esta norma se reconoce que el objetivo de esos atributos es satisfacer las necesidades del usuario a través de seis atributos: funcionalidad, fiabilidad, utilidad, eficiencia, mantenibilidad y portabilidad, que a su vez se descomponen en subatributos. Esta norma contempla la calidad interna, externa y la calidad en uso del software [11].

La norma ISO/IEC 14598 describe un proceso para evaluar la calidad de los productos de

software, el cual es compatible con la norma anterior, y donde la usabilidad conforma la calidad del software [10] [3].

Una de las definiciones referenciadas en el SWEBOK [24] por Watts Humphrey [8], define a la calidad de software como “alcanzar excelentes niveles de adecuación para el uso”. Cuando se utiliza la palabra adecuación, se refiere a la aplicación que va a tener un producto en un dominio dado. Es aquí donde aparece el usuario, como un elemento, que conforma una de las partes importantes de este dominio. El usuario es el destinatario del producto fabricado y es quien juzga la calidad del mismo, a través del nivel de satisfacción, que percibe por la interacción y uso del mismo. El software es un producto, que no es ajeno a este concepto. El software debe estar bien definido en el dominio o medio en el que va a ser usado, para poder alcanzar los niveles de calidad requeridos. El usuario deberá percibir de que los sistemas lo ayudaran a realizar sus tareas, en caso contrario serán adversos a su utilización [16]. La usabilidad tiene que ver con las condiciones específicas del dominio o medio de aplicación, con las necesidades del mismo y con los usuarios que hagan uso de los sistemas. Es decir que la usabilidad no es un atributo propio del software y depende del dominio o medio donde se quiere implementar el software. Por lo tanto la usabilidad no se puede especificar independientemente del dominio de uso o aplicación [6]. La aceptación de las aplicaciones que se desarrollan para el entorno web, se fundamenta estrictamente en su alto grado de usabilidad [14]. Una aplicación web con una pobre facilidad de uso, será reemplazado por otra más fácil de usar, tan pronto como el usuario sepa de su existencia [15]. Un buen nivel de usabilidad, hace que el usuario perciba a un sistema, como uno de mayor calidad, sin que esto signifique un aumento excesivo en el costo de desarrollo. Por lo tanto, la usabilidad web tendrá dos objetivos: desarrollar aplicaciones Web bajo el punto de vista de la Ingeniería del Software para crear un producto funcionalmente correcto, y evaluar los criterios de usabilidad de las aplicaciones Web, que haga al software más amigable con el usuario. Entonces la definición de las técnicas y métodos para asegurar la facilidad de uso, es uno de los objetivos de los investigadores de ingeniería web [21].

El término usabilidad es un término abstracto y es necesario descomponerlo en atributos para poder trabajar con el mismo y observar su presencia en los sistemas [16]. Estos distintos aspectos denominados atributos, se describen a continuación:

- **Recuerdo en el tiempo:** Se refiere a usuarios eventuales, que no utilizan a diario el sistema y en específicamente se refiere a la capacidad de



poder usar el sistema sin tener que aprender cómo funciona nuevamente.

- **Facilidad de aprendizaje:** Este atributo considera, que facilidad existe para aprender la funcionalidad básica del sistema, como para poder realizar correctamente la tarea que desea realizar el usuario y poder volver a utilizarlo en el futuro.
- **Eficiencia:** Este atributo toma en cuenta el número de transacciones por unidad de tiempo que el usuario puede realizar usando el sistema. Lo que se busca es la máxima velocidad de realización de tareas del usuario.
- **Tasa de error:** Este atributo es totalmente adverso a la usabilidad de un sistema. Se mide el número de errores cometidos por el usuario, realizando una tarea específica. Los errores disminuyen la eficiencia.
- **Satisfacción Subjetiva:** Muestra la impresión subjetiva obtenida por el usuario al usar el sistema.

Todos estos atributos son considerados a nivel de evaluación de la usabilidad web, donde se transforma en un factor relevante de la calidad del software, que se diseña para este entorno. Esto es reconocido en una publicación de Olsina (2006) [18], donde expresa que: "Las aplicaciones web son interactivos, centrados en el usuario, basadas en hipermedia, donde la interfaz de usuario juega un papel central".

### La interacción con el usuario

Con frecuencia se identifica la usabilidad con las características de los elementos de la interfaz gráfica del usuario [5]. Sin embargo, la usabilidad no sólo tiene que ver con la interfaz gráfica del usuario, sino al modo de realizar las operaciones y como interactúa con el sistema. La interacción del sistema no se encuentra definida en la interfaz que se presenta al usuario, sino que se encuentra embebido en el código que implementa la funcionalidad del mismo [17] [13]. La parte crítica es la lógica del sistema, la cual es considerada durante el diseño de la arquitectura del sistema [25]. La interacción forma parte de la lógica de negocio. Esto quiere decir que el diseño y la interacción del software deben tenerse muy en cuenta desde etapas tempranas de la creación del software. La interfaz es lo que hace posible la interacción del usuario con la aplicación [Duan 2007]. Un buen diseño de la interfaz gráfica hace posible y ayuda a que un sistema aumente su usabilidad, pero si a nivel funcional interno su respuesta no es la adecuada, no

se podrá mejorar su grado de usabilidad cambiando la interfaz únicamente [13] [6]. Como ejemplo de lo antes mencionado se puede citar: saturación de información en una misma página, el uso de links redundantes, dificultad del usuario para encontrar la información que busca, usar links inválidos o que ya no están activos, demoras en las bajadas de datos, etc, de páginas que visualmente son atractivas, pero poseen problemas de usabilidad [17].

### La Ingeniería de Software

La ingeniería del software ha venido estudiando los procesos de desarrollo de software durante años. Estos procesos son un conjunto organizado de actividades que transforman las entradas en salidas. De allí surgieron modelos de procesos como el de cascada, el modelo en "V" o el modelo en espiral [2], que han sido muy aplicados en el desarrollo del ciclo de vida del software [23] [9]. Muchos autores han propuesto varios modelos de procesos para la implementación y evaluación de la usabilidad de los sistemas. Entre ellos se pueden citar los siguientes modelos: Evaluación Eurística, Diseño Contextual, Casos Esenciales de Uso, Desarrollo de Escenarios, o Modelos Genéricos [7]. Todos ellos son distintos a los modelos de la Ingeniería de Software. Entonces la idea es encontrar un modelo que abarque la mayoría de los casos posibles y se adecue a la Ingeniería del Software. La ingeniería del software debe tener en cuenta la usabilidad, desde etapas tempranas de desarrollo del software [6]. Una falla de usabilidad puede ocasionar que mi sistema tenga altos costos de mantenimiento y reproceso de trabajo. Todo el esfuerzo invertido será en pos de buscar la aceptación por parte del usuario, para que vuelva a tener confianza en el sistema y haga uso del mismo. Un sistema con fallas de usabilidad, provoca que el tiempo estimado para el desarrollo del proyecto se prolongue, se incrementen los costos y se genere desconfianza en los usuarios finales. A esto se le debe sumar la amenaza de que un competidor desarrolle un sistema más usable, que pueda llegar a reemplazar a otro sistema con fallas de usabilidad [17]. Según el Dr. Jakob Nielsen, los competidores de todo el mundo están a un click de distancia, refiriéndose a la navegación de páginas web en internet. Un sistema con alta usabilidad ayuda al usuario a realizar sus tareas más eficientemente, y a cumplir los objetivos para los que fue creado. Por esta razón se debe tener en cuenta en los presupuestos de los proyectos, los costos destinados a usabilidad, para poder mantener y quizás acrecentar el mercado para el cual se diseñó el producto.

## La Usabilidad en el Ciclo de Vida del Software

La usabilidad del sistema debe ser considerada, desde etapas tempranas de desarrollo del sistema. El Ingeniero en Sistemas, no se puede ocupar de la usabilidad cuando el desarrollo del sistema, se encuentra casi concluyendo, y esperar que los usuarios ejecuten las pruebas ellos mismos [4]. Para ello el arquitecto del software debe pensar en la usabilidad del software y en el dominio en el que el mismo será utilizado desde el inicio del diseño del sistema, sobre todo por el uso de los recursos que en la primera etapa no son tan altos, como en las etapas posteriores del ciclo de vida de desarrollo del software [23]. Sin embargo es necesario conocer el ciclo de vida del software. Muchos autores plantean varios modos para verificar la usabilidad, en etapas tempranas como por ejemplo: el uso del prototipado, la participación de usuarios representativos durante el proceso de desarrollo, aún incluyendo personas que son usuarios comunes, pero que pueden ayudar al sistema a ser más usable [20] [6].

## Líneas de Investigación y Desarrollo

Las Líneas de Investigación de Desarrollo involucradas son:

- Ingeniería de Software en la construcción de aplicaciones informáticas destinadas al desarrollo de programas web.
- Ingeniería de la Usabilidad: Técnicas y análisis de la Interacción Persona-Ordenador
- Métodos y técnicas para evaluar calidad del software de Requerimientos No Funcionales

## Resultados Esperados y Objetivos

El objetivo general del proyecto de investigación es el siguiente:

- Establecer un marco teórico metodológico de técnicas de usabilidad para validar especificaciones de requerimientos no funcionales de aplicaciones web.

Asimismo entre los resultados que se esperan como fruto del presente trabajo, son los siguientes:

- Establecer requisitos de usabilidad básicos, que debe poseer cualquier aplicación web.
- Describir métricas de usabilidad, que permitan medir y evaluar los resultados obtenidos de la

aplicación de distintas herramientas, para la evaluación de los aspectos de usabilidad.

- Demostrar aplicaciones reales de los principales conceptos de usabilidad, en el entorno web.

## Formación de Recursos Humanos

Este proyecto al estar circunscripto dentro de otro proyecto de investigación de I+D consolidado y al desarrollarse dentro del ámbito académico de la Universidad, prevé la transferencia de los conocimientos que se vayan adquiriendo, como así también la formación de los recursos humanos que participan del mismo. El mismo cuenta con tres becarios de investigación, dos tesis de la maestría en Ingeniería en Sistemas de Información y un doctorando de Ingeniería de Software que desarrollan sus trabajos en el ámbito del proyecto. Se prevé vinculaciones con otras redes de investigación, tratando de contribuir el conocimiento científico-técnico para su posterior transferencia a nivel social.

## Referencias

- [1] Abrahao S., Condori-Fernandez N., Olsina L., Pastor O., Defining and validating metrics for navigational models. Proceedings Ninth International Software Metrics Symposium, Australia, 2003
- [2] Boehm B., A Spiral Modelo for Software Development and Enhancement, Computer, Vol 21 n° 5, 1998
- [3] Bevan Nigel, Quality and usability: A new framework, National Physical Laboratory, Teddington, Middlesex, Netherlands. 1997
- [4] Costabile Maria Francesca, Usability In The Software Life Cycle, Dipartimento di Informatica, Università di Bari, Via Orabona 4, 70125 Bari, Italy, 2000
- [5] Duan, Jianli; Zhang, Ning, Research on Visualization Techniques for Web Usability Analysis, IEEE, 2007
- [6] Xavier Ferré Grau, Principios Básicos de Usabilidad para Ingenieros Software, Facultad de Informática de Universidad Politécnica de Madrid, Boadilla del Monte, Madrid, 2005.
- [7] Granollers Toni; Lorés Vidal, Sendin Montse, Perdris Ferran, Integración de la IPO y la Ingeniería

- del Software: MPIu+a”, Depto. Informática Ingeniería Industrial, Universidad de Lleida,
- [8] Watts Humphrey, *Managing the Software Process*, Addison Wesley, 1989.
- [9] Ian Sommerville, *Ingeniería del Software*, 7° Ed., Pearson Educación, Madrid, España, 2005
- [ISO 13407] ISO 13407: 1999, “Human-Centred Design Processes for Interactive Systems, ISO, Genova, Suiza, 1999
- [10] International Standard (1998) ISO 14598-1:1998, “Information Technology – Evaluation of Software Products – General Guide”. 1998
- [11] ISO/IEC 9126-1:1991, “Software Product Evaluation – Quality characteristics and guidelines for their use”, 1991
- [12] International Standard (1998) ISO 9241-11:1998, “Ergonomic requirements for office work with visual display terminals(VDTs), part 11: Guidance on usability.
- [13] Krug Steve, *No me hagas pensar: Una aproximación a la usabilidad en la web*, 2ª ed. Madrid, Pearson, 2006.
- [14] Matera M., Rizzo F., Carughi G.: *Web usability: principles and evaluation methods: Web engineering*, Springer, 2006 , pp. 143–179
- [15] Mendes E., Mosley N., Counsell S., ‘The need for web engineering: an introduction: Web engineering’, Springer, 2006
- [16] Nielsen Jakob, *Usability Engineering*, Academic Press. 1993.
- [17] Nielsen Jakob, *Designing Web Usability*, Pearson Educación, 2002
- [18] Olsina L., Covella G., Rossi G., ‘Web quality’: ‘Web engineering’, Springer, 2006, pp. 109–142
- [19] Offutt J., *Quality attributes of web software applications*, IEEE Software, 2002
- [20] Otaiza Renato, Rusu Cristian, Roncagliolo Silvana, *Evaluating the Usability of Transactional Web Sites*, Pontificia Universidad Católica de Valparaiso, Chile.
- [21] Pfleeger S.L., ‘Albert Einstein and empirical software engineering’, IEEE Comput., 1999
- [22] Pressman R., *What a tangled Web we weave*, IEEE Software, 2000
- [23] Pressman R., Jesús Elmer Murrieta, *Ingeniería del software: un enfoque práctico*, 6°Ed., McGraw-Hil Interamericana, España, 2005
- [24] IEEE Computer Society Professional Practices Committee, “Guide to the Software Engineering Body of Knowledge – 2004 Version”, IEEE Computer Society, Los Alamitos – California USA, 2004.
- [25] Rafla, Tamer ; Robillard, Pierre N.; Desmarais, Michel , *Investigating the impact of usability on software architecture through scenarios: A case study on Web systems*, Software Engineering Research Lab, Ecole Polytechnique de Montreal, Station Centre-Ville, Montreal, Canada, 2005

## CONSTRUCCIÓN DEL LEL DE REQUISITOS

**Gladys Kaplan<sup>1,3</sup>, Jorge Doorn<sup>1,2</sup>, Nora Gigante<sup>1</sup>**

<sup>1</sup>Departamento de Ingeniería e Investigaciones Tecnológicas, UNLaM

<sup>2</sup>INTIA, Departamento de Computación y Sistemas, Facultad de Ciencias Exactas, UNCPBA

<sup>3</sup>LINTI, Facultad de Informática, UNLP

Florencia Varela 1903 – San Justo – Pcia. Bs As

[gladyskaplan@gmail.com](mailto:gladyskaplan@gmail.com), [jdoorn@exa.unicen.edu.ar](mailto:jdoorn@exa.unicen.edu.ar), [noragigante@gmail.com](mailto:noragigante@gmail.com)

*Palabras Clave: evolución del vocabulario, LEL, LEL<sub>r</sub>, proceso de requisitos.*

### RESUMEN

La línea de investigación “Estabilidad de los documentos en el proceso de Requisitos” concentra su atención en la ambigüedad que pueden contener algunos documentos finales del proceso de requisitos. El decremento en la calidad de estos documentos, específicamente en los escenarios futuros y en la especificación de requisitos del software, se debe a la disminución del cubrimiento del LEL. Dicha brecha se genera en el mismo momento en el que se piensa incorporar un nuevo sistema de software, ya que para describir las necesidades y las posibles soluciones es necesario incorporar nuevo léxico y modificar algunos existentes. Esta falta de cubrimiento del LEL en los documentos finales de requisitos genera ambigüedad perjudicando la comprensión de la solución. Por lo tanto, es necesario identificar la evolución del vocabulario para registrarla de una manera ordenada con el objetivo de asegurar la calidad y la comprensión de todos los documentos generados. La identificación y registro de este nuevo léxico puede ser realizado en forma concurrente a la descripción de los escenarios futuros o luego de ser construidos. El proceso de construcción de este nuevo glosario denominado *LEL de requisitos* ha motivado esta nueva línea de investigación.

### CONTEXTO

El proyecto “Estabilidad de los documentos del proceso de requisitos” (código C144) que se desarrolla en el Departamento de Ingeniería e Investigaciones Tecnológicas de la Universidad Nacional de La Matanza (UNLaM). Este proyecto es la continuación de varios proyectos anteriores a lo largo de los cuales se ha definido un proceso de requisitos [Leite 04] y se lo ha aplicado reiteradas veces a diferentes realidades.

### INTRODUCCIÓN

Es ampliamente conocido [Rolland 98] [Oberg 98] [Wiedenhaupt 98] [Alsplough 99] [Robertson 06] que la construcción de glosarios como parte del proceso de requisitos, es una actividad de gran valor ya que permite asegurar una buena comunicación entre:

- los clientes-usuarios con el equipo de desarrollo;
- el propio grupo de clientes-usuarios;
- el propio equipo de desarrollo.

Los glosarios se construyen con el fin de esclarecer y determinar el vocabulario utilizado en los documentos de requisitos,



por lo tanto sirve como ancla en toda la documentación generada, principalmente en la especificación de requisitos de software (ERS). El uso de los glosarios facilita la elicitación de conocimiento y mejora la validación con los clientes y usuarios.

Es frecuente encontrar que en el propio universo de discurso existan ambigüedades sobre la semántica de algunos términos. Es así que suele encontrarse que diferentes actores usen el mismo término con significados diferentes. El glosario puede registrar este hecho distinguiendo explícitamente los diferentes significados de estos homónimos de manera que al usar el término en otro documento se pueda identificar el significado utilizado.

Sin embargo, cuando estos glosarios son construidos tempranamente en el proceso de requisitos, se agrega otro nivel de ambigüedad: la evolución no registrada del vocabulario del dominio. Esto provoca que los glosarios tiendan a quedar rápidamente obsoletos y en el peor de los casos esta ambigüedad queda oculta en el proceso, pudiendo generar requisitos de software mal comprendidos y erróneos.

En el presente proyecto de investigación se utiliza el proceso [Leite 04] que genera, como primera actividad del proceso, el Léxico Extendido del Lenguaje (LEL) [Leite 90], [Hadad 08] que es un glosario que representa el vocabulario de los clientes y usuarios. El LEL se retroalimenta durante la construcción de los escenarios actuales, los cuales describen los procesos del negocio tal como se realizan hasta ese momento [Leite 00]. Cuando se describen los escenarios futuros, que contienen la solución para el nuevo sistema de software [Doorn 02], se genera una ampliación de la semántica de los términos existentes, o sea una evolución de contexto que se manifiesta en el vocabulario con la incorporación de términos nuevos, la modificación de algunos existentes y el

desuso de otros. Esta evolución del vocabulario es necesaria para describir como serán los nuevos procesos del negocio cuando el sistema de software se encuentre en ejecución.

Existe una fuerte permeabilidad conceptual entre los clientes y usuarios con los desarrolladores, dado que los documentos que crean los desarrolladores motivan, por un lado, a los clientes a comprender la interacción entre el nuevo sistema de software y el proceso del negocio y, por el otro, a los desarrolladores en las problemáticas específicas de los clientes. Esto hace que se produzca una evolución en el vocabulario propio del contexto, la que debe ser registrada en los documentos que se construyan.

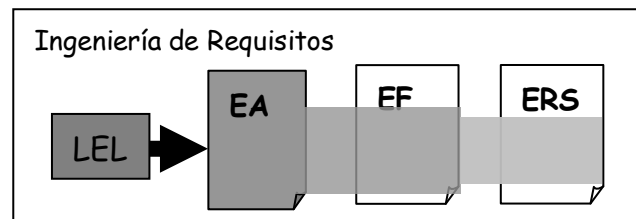


Figura 1 – Cubrimiento del LEL en los documentos de requisitos

En la Figura 1 se puede ver como el LEL cumple inicialmente con los objetivos previstos y su cubrimiento es total en las descripciones de los escenarios actuales (EA), mientras que en las descripciones de la planificación de los procesos de negocio, o sea en los escenarios futuros (EF) y en la ERS, este cubrimiento disminuye. Cuanto más significativos sean los cambios en la planificación de los procesos del negocio, existirá más necesidad de incorporar nuevo vocabulario o modificar el existente.

Este nuevo glosario es utilizado para mejorar la legibilidad y comprensión de los EF y deberá ser nuevamente actualizado para cubrir la ERS, ya algunas transformaciones desde los EF a los requisitos del software requerirán

de algunas ampliaciones léxicas. Por tal motivo, este glosario se denomina “LEL de los Requisitos o LEL<sub>r</sub>”.

## LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Uno de los objetivos propuestos para esta línea de investigación “Estabilidad de los documentos de requisitos” se centra en mejorar la calidad, consistencia y vigencia de los documentos de requisitos generados durante todo el proceso de construcción del software.

Un estudio profundo sobre la consistencia semántica existente entre el LEL y el resto de los documentos de todo el proceso de requisitos [Leite 04] reafirmo que los procesos del negocio evolucionan cuando se planifica la solución. Esta evolución queda claramente representada en los EF. Pero el LEL, en el cual se anclan las descripciones de los EF y la ERS, permanece estático en el tiempo o con mínimas modificaciones. Este congelamiento del léxico deja parte de las descripciones de los EF y de los requisitos del software con ambigüedad ya que las definiciones del LEL no corresponden semánticamente con el significado utilizado. Esta línea de investigación propone evolucionar el LEL y seleccionar un proceso de construcción apropiado.

Existen diferentes opciones a la hora de construir el LEL<sub>r</sub>.

Por un lado es necesario determinar si es conveniente evolucionar el LEL existente o si es mejor construir un glosario independiente. Por otro, hay que incluir este proceso de construcción del LEL<sub>r</sub> dentro del proceso de requisitos.

El LEL<sub>r</sub> comparte muchas de las particularidades del LEL. Algunas diferencias entre ambos glosarios son las siguientes:

El LEL describe el vocabulario del UdeD actual. El LEL <sub>r</sub> describe el vocabulario de los requisitos.
--

El LEL se genera muy tempranamente en el proceso de requisitos, antes de construir los EA. El LEL <sub>r</sub> se crea al trabajar con la planificación de la solución.
---

El LEL se utiliza en los modelos del UdeD actual. El LEL <sub>r</sub> se genera para todos los modelos de la planificación del UdeD futuro, o sea para los EF y la ERS
--

Cabe destacar que el LEL<sub>r</sub> deberá permitir incluir términos no pertenecientes al vocabulario del dominio, por lo tanto será parcialmente artificial. Estos términos provienen exclusivamente del ingeniero de requisitos con el propósito de minimizar la ambigüedad de los EF y de la ERS y mejorar la comprensión de los requisitos del software.

## RESULTADOS Y OBJETIVOS

La evolución del LEL se ha aplicado desde el año 2010 en los casos de estudio de la asignatura Ingeniería de Requerimientos en UNLaM que realizan los alumnos del último año de la carrera Ingeniería en Informática.

En el análisis de las posibilidades para construir el LEL<sub>r</sub> se ha determinado que es diferente la evolución del LEL en el mismo documento que generar un documento independiente. La registración del vocabulario utilizado en el proceso de requisitos dentro del mismo LEL fue descartada debido a que estos glosarios tienen diferentes objetivos. Uno describe el vocabulario inicial del dominio, el otro el vocabulario necesario para planificar la solución. La inclusión en el LEL de los nuevos términos agrega complejidad dificultando no sólo su construcción sino

también su posterior lectura. Esto se debe a la necesidad de utilizar homónimos, a veces muchos, para diferenciar las descripciones correspondientes al vocabulario del UdeD actual de aquellas necesarias para el proceso. Por lo tanto, tener un único documento para ambos vocabularios en vez de mejorar la calidad del LEL y reducir la ambigüedad de los documentos de requisitos tiende a incorporar riesgos no deseados. Por tal motivo se ha llegado a la conclusión que es necesario construir un documento independiente.

Otro tema de profundo análisis es determinar el enfoque de construcción. Ya que puede realizarse en simultáneo con los EF, o sea interactivo, o una vez que estos se completan, en batch.

La construcción interactiva del LELr implica realizar el proceso de construcción del LELr en paralelo con el proceso de construcción de los EF. Durante la descripción de cada EF se analiza el vocabulario utilizado y se identifica el nuevo vocabulario o aquel que cambia de significado, ambos modelos se construyen en simultáneo. La ventaja de este enfoque es que en el momento de construir el glosario se cuenta con la información directa del UdeD. Pero este enfoque genera una sobrecarga de atención y concentración para el ingeniero de requisitos ya que debe construir dos modelos en paralelo que a pesar de ser complementarios tienen objetivos diferentes e inter-relaciones e intra-relaciones particulares para cada caso. Este exceso de objetivos puede disminuir la calidad de alguno de los dos modelos (EF o LELr), llegar a reducir la calidad de ambos o afectar la calidad de los modelos que los utilizan. Por otro lado, se dificulta cumplir con el principio de circularidad del LEL, donde se maximiza el uso de otros símbolos en las descripciones, ya que los EF que contendrán ese nuevo vocabulario aún no se han descrito.

La construcción en batch se refiere a construir el LELr una vez finalizado el proceso de construcción de los EF. Se debe analizar el vocabulario utilizado en cada escenario. En este momento toda la atención del ingeniero de requisitos se concentra en el vocabulario necesario para minimizar la ambigüedad de los EF y en construir un glosario de buena calidad. Este enfoque permite una visión en dos etapas con mecanismos cognitivos distintos. En la primera etapa la atención está puesta en describir y definir el nuevo sistema de software. Mientras que en la segunda etapa de construcción del LELr, se está pensando particularmente en el vocabulario y no en los procesos del negocio.

La construcción del LELr en batch también tiene algunos inconvenientes con la particularidad que son mitigados con la mera generación de una lista inicial de símbolos. Unos de estos problemas es que los EF se describen sin un vocabulario que los complementa, o con uno, el LEL, que lo hace parcialmente. Esto puede provocar que los EF requieran ser refinados luego de construir el nuevo glosario para eliminar detalle de los escenarios que ahora se encuentran en el LELr.

Se debe analizar y decidir el mejor enfoque para construir el LELr. Y se espera construir una heurística para la generación de dicho glosario.

## **FORMACION DE RECURSOS HUMANOS**

Esta línea de investigación es parte de la tesis doctoral de la Lic. Gladys Kaplan “Construcción de Storyboard semiautomático para validar escenarios futuros” que está desarrollando en la UNLP. También co-ayuda en el avance de la tesis de maestría que realiza en la UNLaM en simultáneo “Información Extemporánea en el Proceso de Requisitos”. Esta línea también co-ayuda en el avance de la tesis de maestría “Generación de Storyboard semiautomático” que está

desarrollando la Lic. Renata Guatelli en la UNLaM. Lo mismo ocurre con la tesis de maestría cuya denominación provisoria es “Detección de clusters semánticos en Glosarios de Requisitos” que está comenzando a desarrollar el Ing, Guillermo Hindi también en la UNLaM.

Parte de los resultados más avanzados en esta línea de investigación se han comunicado a alumnos de grado en el curso de Ingeniería de Requerimientos de la UNLaM y de posgrado a través del curso Tópicos de Ingeniería de Requisitos en la Maestría en Informática del Departamento de Postgrado de la UNLaM.

## REFERENCIAS

- [Rolland 98] Rolland C., Ben Achour C. (1998) Guiding the construction of textual use case specifications, *Data & Knowledge Engineering*.
- [Oberg 98] Oberg R., Probasco L, Ericsson M. (1998) Applying Requirements Management with Use Cases. Rational Software Corporation.
- [Wiedenhaupt 98] Weidenhaupt K., Pohl K., Jarke M., Haumer, P (1998) Scenarios in System Development: Current Practice., *IEEE Software*.
- [Alsplaugh 99] Alpaugh T.A., Antón A.I., Barnes T., Mott B.W. (1999) An Integrated Scenario Management Strategy, *International Symposium On Requirements Engineering (RE99)*, Limerick-Irlanda (IEEE Computer Society Press).
- [Robertson 06] Robertson S. and Robertson J. (2006) *Mastering the Requirements Process*, 2nd Ed, Addison-Wesley.
- [Leite 04] Leite J.C.S.P., Doorn J.H., Kaplan G.N., Hadad G.D.S., Ridaio M.N. (2004) Defining System Context using Scenarios, In: Leite J.C.S.P. and Doorn J.H (eds) *Perspectives on Software Requirements*, Kluwer Academic Publishers, ch. 8, pp.169-199.
- [Hadad 08] Hadad G.D.S., Doorn J.H., Kaplan G.N. (2008) *Creating Software System Context Glossaries*, In: Mehdi Khosrow-Pour (ed) *Encyclopedia of Information Science and Technology*. IGI Global, Information Science Reference, Hershey, PA, USA, ISBN: 978-1-60566-026-4, 2nd edn, Vol. II.
- [Leite 90] Leite J.C.S.P., Franco, A.P.M., (1990) “O Uso de Hipertexto na Elicitação de Linguagens da Aplicação”, *Anais de IV Simpósio Brasileiro de Engenharia de Software*, SBC.
- [Doorn 02] Doorn J., Hadad G., Kaplan G. (2002) *Comprendiendo el Universo de Discurso Futuro*, WER'02 - Workshop on Requirements Engineering, Valencia, Spain, pp.117-131.
- [Leite 00] Leite J.C.S.P., Hadad G.D.S., Doorn J.H., Kaplan G.N. (2000). A Scenario Construction Process, *Requirements Engineering Journal*, 5, (1).



## Proyectos de Evaluación de Productos de Software con un nuevo Framework de Calidad

Rolando Titiosky, Paula M. Angeleri, Amos Sorgen, Jaquelina Wuille Bille  
Facultad de Ingeniería y Tecnología Informática - Universidad de Belgrano  
Av. Federico Lacroze 1947 - Teléfonos: 4772-4010 interno 131/ 121  
1426 Ciudad Autónoma de Buenos Aires, Argentina  
{rolando.titiosky, amos.sorgen}@comunidad.ub.edu.ar  
{paula.angeleri}@ub.edu.ar

### Resumen

El objetivo de este artículo es presentar la situación actual del proyecto de investigación MyFEPS [1] *Metodologías y Framework para la Evaluación de Productos de Software, basados en normas internacionales*, en desarrollo en la Facultad de Ingeniería y Tecnología Informática de la Universidad de Belgrano, cuyo propósito es diseñar e implementar un framework para ayudar a técnicos, ingenieros y gerentes en todo el proceso de evaluación de software, desde la determinación de los objetivos de la evaluación, su planificación, ejecución y presentación de los resultados.

**Palabras clave:** Calidad de software, Calidad de productos, Evaluación de calidad, Evaluación de la calidad del producto *software*, modelo de calidad de software, *MyFEPS*.

### Contexto

Este proyecto se planificó en el año 2009 con el propósito de cubrir las falencias detectadas en modelos de calidad de productos de software, que habían quedado desactualizados e incompletos, con el avance de la tecnología. En especial el modelo de software propuesto por la norma ISO/IEC 9126-1:2001 [2] y el

proceso de evaluación cubierto por la serie de normas ISO/IEC 14598 [3], que tomaron un interés especial para Argentina por estar citadas en el Anexo VI de la Resolución 61 de la Ley 25.922 de Promoción de la Industria del Software [4]. Esta Ley, brinda la posibilidad de obtener beneficios fiscales a empresas desarrolladoras de software que certifiquen al menos un producto de software representativo para su negocio, entre otras posibles condiciones para ingresar al régimen de promoción. Para acompañar esta futura demanda de servicios el IRAM desarrolló el servicio de certificación de software[5], obteniendo una mención especial en la entrega de *premios Sadosky* [6] en 2008 [7] por ser un servicio innovador en América Latina. Este servicio toma de referencia las normas mencionadas, por ser las normas disponibles. Si bien estas normas están en revisión en la ISO (International Standardization Organization) bajo el número ISO/IEC 25000 [12], conocida como SquaRE<sup>1</sup> se trata de una amplia serie de normas cuya revisión lleva años, por el tipo de proceso de elaboración de

---

<sup>1</sup> sigla proveniente del nombre en Inglés "*Software Quality Requirements and Evaluation*".

normas de ISO, las cuales se aprueban por consenso entre sus países miembros <sup>2</sup>. A la fecha aún no se ha publicado la nueva revisión de las normas ISO/IEC 9126 partes 2 [8], 3 [9] y 4 [10] que establecen las métricas para poder llevar a cabo la evaluación, y sus versiones vigentes están incompletas. Ante esta situación, la estrategia del proyecto de investigación MyFEPS es abordar el desarrollo de una meta-metodología de evaluación, adaptable a diversos contextos, y un modelo de calidad que tenga trazabilidad con otros modelos, especialmente los planteados por la ISO.

## Introducción

Del buen funcionamiento de los sistemas informáticos puede depender la vida de las personas, el éxito o fracaso de los procesos de las organizaciones. El control de calidad de software no se satisface con la observación simple y directa de su funcionamiento. En realidad, es imposible tener la certeza de que un producto de software carezca de fallas por métodos analíticos. No basta con probar un producto de software para asegurarse de que cumple con las especificaciones para las cuales fue diseñado. La evaluación del software es una tarea compleja. Su complejidad está en relación directa con las características de calidad fijadas como objetivo de evaluación, su grado de calidad objetivo, la complejidad y el nivel de riesgo del producto a evaluar (riesgos económicos, riesgos de seguridad, riesgo de vida, etc), entre otros factores. Un desafío adicional es que la complejidad de este tipo de productos

aumenta con la cada vez más acelerada evolución de las TIC.

El proyecto de investigación MyFEPS tiene por objetivo principal desarrollar una meta metodología de evaluación de software, que permita establecer metodologías de evaluación ajustadas a las necesidades de cada proyecto. A su vez se espera definir un *framework* que incluya un modelo de calidad de productos *software*, guías, recomendaciones y plantillas para facilitar la implementación de metodologías de evaluación en las organizaciones que requieren evaluar software, ya sea un producto en desarrollo o terminado, propio o por adquirir.

## Plan de trabajo

Para alcanzar el objetivo principal del proyecto (explicado en el párrafo anterior) y los objetivos específicos, planteados en la sección **Resultados y Objetivos**, se propone un plan de trabajo conformado básicamente por las siguientes actividades:

### Etapa 1 “inicio del proyecto”:

- constitución del equipo de trabajo,
- establecimiento de la infraestructura,
- elaboración de los estándares de documentación
- planificación detallada del proyecto.

### Etapa 2 “relevamiento sobre evaluación”:

- análisis de lo publicado.
- búsqueda de diferentes enfoques (alternativas) a aplicar en la investigación,
- identificación el marco de las normas internacionales a seguir para los procesos de evaluación de productos de software e
- identificación de herramientas de evaluación.

### Etapa 3 “diseño del framework”:

- análisis de viabilidad,

<sup>2</sup> El Subcomité de ISO que elabora las normas de Ingeniería de Software es el *JTC1/SC7 Software and Systems Engineering* compuesto por un total de 59 países. *JTC1 Information Technology* es el Comité Técnico Conjunto 1 de ISO e IEC (International Electro Technical Commission) [9].

- establecimiento de la taxonomía de los tipos de productos de software y de objetivos de negocio,
- desarrollo de la meta-metodología de evaluación
- creación del framework (marco de trabajo), flexible y adaptable a distintos tipos de productos de software y a distintos objetivos de negocio que implemente la meta-metodología.

#### **Etapa 4 “aplicación del framework”:**

- aplicación del framework diseñado en proyectos pilotos.
- Análisis de los resultados obtenidos
- Validación del framework y de la meta-metodología en base a los hallazgos
- Ajustes a los modelos.

#### **Etapa 5 “transferencia del conocimiento”:**

- transferencia en industria y en academia.

## **Líneas de investigación y desarrollo.**

Las líneas de investigación sobre las que se ha trabajado y se está trabajando son las siguientes:

*Dominio de conceptos:* se ha definido un glosario de términos relacionados a la calidad del software.

*Modelo Básico de Calidad:* se ha definido un modelo de calidad sin ambigüedades, con trazabilidad a otros modelos de calidad existentes [15] como el que establecen las normas ISO/IEC 9126-1 [2] e ISO/IEC 25010 [13].

*Métricas de Calidad:* se han definido las métricas que permitan evaluar la calidad de un producto en el marco de un Modelo Básico de Calidad.

*Homogenización de los grados de calidad:* se ha definido un marco en el cual se pueda homogeneizar la disparidad conceptual entre los distintos ítems de calidad (características y atributos) de forma que se puedan calcular y comparar grados de calidad.

*Factores de la Evaluación:* se han investigado cuales son los factores que influyen en el proceso de evaluación y cómo afectan al proceso. Ejemplos de estos factores son:

- a. Objetivos de la evaluación
- b. Norma / Modelo de Calidad a usar
- c. Tipo de producto
- d. Tamaño del producto
- e. Complejidad del producto
- f. Objetivos de negocio del producto
- g. Características de los stakeholder

*La influencia de los stakeholder):* se ha investigado cómo introducir cuantitativamente la influencia de los diferentes stakeholders, con el objetivo de definir en cada proyecto de la forma más objetiva posible, la percepción de la calidad del producto objetivo.

*Algoritmos de evaluación:* dado un modelo de calidad en niveles (producto, características, subcaracterísticas, atributos) se ha investigado la manera de aglomerar evaluaciones a niveles más bajos para llegar a evaluaciones a niveles más altos.

*Esfuerzos y fidelidades de las mediciones:* se investigan algoritmos que puedan decidir, en el marco de la evaluación de un producto, con qué fidelidad medir cada métrica a ser medida, y cómo esa decisión afecta al esfuerzo (en horas/hombre) y al presupuesto necesarios para la evaluación.

## Resultados y Objetivos

Los objetivos se enuncian en términos de principal y específicos. El objetivo principal ya mencionado, consiste en la definición de una meta metodología de evaluación de software que permita definir metodologías apropiadas para diferentes proyectos, y generar un framework que facilite su implementación.

### Objetivos específicos:

1. Establecer una taxonomía de los tipos de productos de software y de los objetivos de negocio con respecto a su evaluación de calidad.
2. Relevar metodologías, técnicas, herramientas y normas de evaluación de productos de software.
3. Crear un repositorio de datos que incluya el cuerpo de conocimiento acerca de las metodologías de evaluación de software actuales: conceptos, modelos, principios, métodos, técnicas, mejores prácticas, estudios de casos, etc,
4. Seleccionar la base documental y desarrollar la base teórica que de sustento a un modelo de calidad de productos *software*.
5. Especificar el modelo de calidad, y la ponderación de características de calidad en base a objetivos de evaluación.
6. Especificar las métricas a utilizar para medir el grado de calidad de cada característica fundamental.
7. Elaborar una meta metodología de la cual se deriven metodologías de evaluación de software adecuadas para cada tipo de producto y objetivos de evaluación.
8. Tomando como referencia la meta metodología diseñar un framework (marco de trabajo) que incluya lineamientos, técnicas, recomendaciones de herramientas y plantillas para la estandarización de las actividades de evaluación.
9. Realizar proyectos piloto que permitan validar tanto la meta-metodología como el framework.
10. A partir de los hallazgos encontrados en los pilotos, realizar los ajustes necesarios.
11. Transferir los conocimientos adquiridos a la industria del software a través de conferencias, publicaciones y cursos de capacitación.
12. Capitalizar los conocimientos adquiridos incorporándolos en asignaturas de las carreras de grado y en cursos de postgrado y de actualización profesional.

Parte de los resultados obtenidos fueron ya publicados anteriormente[1] y difundidos en 3 Congresos: JAIIO 2011, CHIS 2011 TACNA Perú, CONEISC 2012 Chiclayo Perú, y en el Primer Seminario Internacional ISO de Calidad de Software, en UB, Buenos Aires, 2012.

Con respecto al resto de los objetivos planteados, el resultado a la fecha es haber cumplido los objetivos 1 al 8, y se han iniciado 3 proyectos piloto: *ConexionUB*, *eCom*, *SGH*.

Esperamos obtener de los resultados de los Pilotos una valoración en Métricas y Gradiente de Calidad tal, que nos permita mediante un mapeo, establecer su correlación en las métricas de la ISO, y de esta manera, validar que tan ajustados está el modelo de calidad MyFEPS.

Esperamos que MyFEPS sea más preciso que el modelo de ISO, pues cuenta con mayor información para describir las características y subcaracterísticas y sus Métricas.



## Recursos Humanos

**Directores:** Mg. Paula María Angeleri y Dr. Amos Sorgen.

### Integrantes:

Mg. Rolando Titiosky, Profesor Investigador.

Lic. Jaquelina Wuille Bille, Profesora Investigadora.

Esp. Jorge Ceballos, Profesor Investigador y Tesista de posgrado.

Martín Santi, Tesista de grado.

Agustín Ventura, Tesista de grado.

Los integrantes de este grupo de investigación dirigen a estudiantes y a consultores externos en estos temas de investigación, en diferentes contextos (becarios, tesinas, trabajos prácticos de alumnos, etc). Además participan en el dictado de asignaturas de grado y cursos de postgrado de la Facultad en los cuales este año se ha planificado enseñar el resultado del proyecto MyFEPS.

De esta manera, actualmente se está capacitando a:

- Tesistas
- Alumnos
- Docentes
- Administradores de Proyectos
- Ingenieros de software
- Testers

## Referencias bibliográficas

[1] Sorgen, A., Angeleri, P.: El Modelo de Evaluación del Proyecto MyFEPS, 40JAIIO - ASSE 2011 - ISSN: 1850-2792, 180-191, (2011), <http://www.40jaiio.org.ar/sites/default/files/T2011/ASSE/790.pdf>

[2] ISO/IEC 9126-1:2001 Software engineering -- Product quality -- Part 1: Quality model

[3] ISO/IEC 14598:1999 Information technology - Software product evaluation

[4] Resolución N° 61/2005

[http://biblioteca.afip.gob.ar/gateway.dll/Normas/ResolucionesComunes/res\\_62000061\\_2005\\_05\\_03.xml](http://biblioteca.afip.gob.ar/gateway.dll/Normas/ResolucionesComunes/res_62000061_2005_05_03.xml), consultada por última vez el 06/03/2013.

[5] Servicios de Certificación IRAM  
<http://www.iram.org.ar/seccion.php?ID=3&IDS=21#productprocservicios>, consultada por última vez el 06/03/2013.

[6] Premios Sadosky, El premio, Homenaje a Manuel Sadosky  
<http://www.cessi.org.ar/sadosky/ver-descripcion-303>, consultada por última vez el 06/03/2013.

[7] Premios Sadosky, Menciones 2008  
<http://www.cessi.org.ar/sadosky/mentions-2008>, consultada por última vez el 06/03/2013.

[8] ISO/IEC TR 9126-2:2003 Software engineering -- Product quality -- Part 2: External metrics.

[9] ISO/IEC TR 9126-3:2003 Software engineering -- Product quality -- Part 3: Internal metrics.

[10] ISO/IEC TR 9126-4:2004 Software engineering -- Product quality -- Part 4: Quality in use metrics.

[11] Uno de los autores, Paula M. Angeleri, es Delegada Argentina del Subcomité ISO/IEC JTC1/SC7.

[12] ISO/IEC 25000 series – Systems and software engineering – Software Product Quality Requirements and Evaluation (SQuARE).

[13] ISO/IEC 25010:2012 SQuARE – Software Product Quality Requirements and Evaluation – Quality Model.

[14] ISO/IEC FDIS 25040:2010(E)- Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuARE) — Evaluation process.

[15] Kan, S.H.: Metrics and Models in Software Quality Engineering, Addison-Wesley Longman Publishing Co., Inc (2002)

## Aplicación de Software colaborativos o groupware empleando metodología SCRUM en Sistemas y Organizaciones

Ing. Susana Eisbel Rey – Ing. Silvia Lanza Castelli –  
Ing. Sandra Allende

Departamento de Sistemas de Información - Facultad Regional Córdoba –  
Universidad Tecnológica Nacional.

Cruz Roja esq. Maestro López – Córdoba Capital

Te: 351-156807346

srey@sistemas.frc.utn.edu.ar

### RESUMEN

El proyecto investiga los métodos existentes en Entornos colaborativos y en Metodologías Ágiles (SCRUM) para la identificación, indagación y aplicación de sus conceptos y herramientas multimediales para establecer un entorno de trabajo en grupo para estudiantes mediante el uso de la plataforma Moodle.

Las preguntas que guían ésta investigación se refieren a: ¿En qué medida se aplican los entornos colaborativos en la selección y uso de información al momento de compartirla?. ¿Es posible aplicar metodologías ágiles en desarrollo de proyectos colaborativos?. ¿ Son análogos? ¿En qué medida se obtendrá ventaja de la aplicación del software libre?.

Se define una analogía como una comparación entre dos dominios o sistemas diferentes que poseen un conjunto de relaciones similares entre los elementos que los componen. Por ello creemos conveniente tomar y explicar temas nuevos y compararlos con temas que nos resultan familiares y que, a su vez, resultan difíciles de conceptualizar en sus propios términos.

Consideramos imprescindible crear ambientes y escenarios de aprendizaje facilitadores de intercambio entre los esquemas ya formados (nociones, habilidades, conceptos) con el medio que puedan resultar en la reformulación de esos esquemas y la construcción de nuevas estructuras de conocimiento.

*Palabras clave: Entornos Colaborativos –Scrum - Moodle*

### CONTEXTO

En la actualidad no existe información que permita conocer el uso, aprovechamiento,

impacto, aplicación y utilidad de un sistema de trabajo en grupo mediante el uso de un navegador web.

Disponer de nuevos recursos que puedan permitir nuevas formas de hacer las cosas no significa que necesariamente se produzcan cambios culturales.

Highsmith & Cockburn 2001– *“lo que es nuevo en los procesos ágiles no son las prácticas que usan, sino que reconozcan a las personas como primeros implicados en el éxito de un proyecto, además de un intenso foco en la efectividad y la manejabilidad. Esto genera una nueva combinación de valores y principios que definen una visión ágil del mundo.”*

Los espacios e instrumentos para el trabajo colaborativo como los servicios de correo electrónico, los foros, chats, servicios de videoconferencia, en algunos casos, éstas iniciativas podemos considerarlas aún experimentales, nos sirven para investigar sobre las mejoras que realmente pueden aportar a la virtualidad en la Toma de Decisiones, en las Organizaciones y en los estudios presenciales de los alumnos de primer nivel, para ir mejorando y ampliando sus funcionalidades.

La información que debe ser alimentada, preservada y empleada lo más ampliamente posible tanto por los individuos como por las organizaciones a las que éstos pertenecen y los procesos a los que es sometida como creación, recopilación, organización, transformación, aplicación y resguardo de la información existente en la organización, merecen estudios que aumenten su conocimiento como así también su prestabilidad al servicio de las personas.

Si bien se ha podido comprobar que a través de internet, y de las redes internas de la universidad, se pueden ofrecer servicios a los estudiantes que mejoran los sistemas de enseñanza y facilitan sus aprendizajes, nuestro propósito es guiar y encontrar soluciones en el manejo a la medida del ingresante, dada las

características de sus necesidades y su nivel de conocimiento.

Hemos observado que no todos estudian de la misma manera, algunos lo hacen por cátedra, otros estudian una determinada cantidad de horas por día, otros lo hacen cuando tienen tiempo. Cada uno administra según su capacidad y oportunidad, el tiempo de dedicación y modalidad de estudio.

Sondeos de opinión a estudiantes secundarios sobre la utilización de tecnología en la educación, han reconocido desconocimiento en saber usar la tecnología disponible en el momento justo, utilizando el medio más adecuado dependiendo de la actividad que se quiere realizar.

Esta situación y desconocimiento tiene como consecuencia que no se puedan establecer líneas específicas concretas de acción encaminadas a desarrollar aplicaciones informáticas específicas para los dictados de las cátedras del primer nivel, considerando la heterogeneidad de los ingresantes universitarios.

La interactividad permite que cualquier usuario puede en su momento plantear dudas y buscar ayuda, y también responder dudas de otros participantes y brindarle su ayuda, propiciando la configuración de comunidades de aprendizaje autogestivas y autosuficientes.

En el año 2001, Schwaber y Mike Beedle describieron la metodología ágil en el libro *Agile Software Development with Scrum*. Tomando de allí la definición de Scrum: "*es un modelo de referencia que definen un conjunto de prácticas y roles, y que pueden tomarse como punto de partida para definir el proceso de desarrollo que se ejecutará durante un proyecto*".

El uso de Scrum nos permitirá la creación de equipos autoorganizados impulsando la co-localización de todos los miembros del equipo, y la comunicación verbal entre todos los miembros y disciplinas involucrados en el proyecto.

Debido a lo anterior, la hipótesis que se plantea en ésta investigación es : ¿ Los involucrados, carecen de una metodología de trabajo, además de no aprovechar eficientemente la información que poseen?. En general los entornos colaborativos existentes ofrecen funcionalidades que

satisfacen necesidades de trabajo planteadas por los alumnos pertenecientes al primer nivel sin reconocerlo.

## 1. INTRODUCCIÓN

El Entorno Colaborativo de Trabajo (ECT) es una aplicación on line, amigable y flexible que se adapta de forma personalizada a cada usuario para dar soporte a sus tareas cotidianas facilitando el acceso a la información y reutilización del conocimiento; siendo el trabajo colaborativo, la comunicación y la participación activa, pilares en la ejecución de tareas de diversa índole.

El estudio tomará como unidad de análisis una cátedra de la carrera Ingeniería en Sistemas de Información, pretendiendo contribuir con una propuesta de aplicación de un proyecto de entorno colaborativo o groupware, enmarcados en conceptos de metodologías ágiles (SCRUM), e implementada en plataforma Moodle.

Si tenemos en cuenta que Scrum es un proceso en el que se aplican de manera regular un conjunto de mejores prácticas para trabajar colaborativamente, en equipo, y obtener el mejor resultado posible de un proyecto, y además, éstas prácticas se apoyan unas a otras, la elección de Scrum aportará al estudio la manera de trabajar en equipos colaborativos. Un principio clave de Scrum es el reconocimiento de que durante un proyecto los participantes pueden cambiar de idea sobre lo que quieren y necesitan (a menudo llamado requirements churn), y que los desafíos impredecibles no pueden ser fácilmente enfrentados de una forma predictiva y planificada. Por lo tanto, Scrum adopta una aproximación pragmática, aceptando que el problema no puede ser completamente entendido o definido.

En virtud de que la palabra Moodle era al principio un acrónimo de Modular Object-Oriented Dynamic Learning Environment (Entorno de Aprendizaje Dinámico Orientado a Objetos y Modular), lo que resulta fundamentalmente útil para la educación. También es un verbo que describe el proceso de deambular

perezosamente a través de algo, y hacer las cosas cuando se te ocurre hacerlas. Las dos acepciones se aplican a la manera en que se desarrolló Moodle y a la manera en que un estudiante o profesor podría aproximarse al estudio o enseñanza de un curso en línea.

Dada las características del proyecto consideramos que puede dar curso a propuestas de solución aplicables a otras unidades académicas como así también unidades organizacionales en pos de mejora en la Toma de Decisión, que asiduamente está frente a un universo incierto.

Se puede observar a modo de hipótesis de trabajo que las soluciones disponibles contemplan parte del proceso educativo. Esto significa que deberían aplicarse ciertas herramientas de consulta que comprendan tanto información administrativa como académica que sirva como base para el desempeño del alumno en su inicio académico.

El propósito es abrir un espacio de interacción para brindar asesoría en comunidad de aprendizaje, en dónde además de hacer preguntas los participantes pudieran responder a otros compañeros inquietudes o experiencias ya vividas.

La interactividad permite que cualquier alumno pueda en su momento plantear dudas y buscar ayuda, y también responder dudas de otros participantes y brindarle su ayuda, propiciando la configuración de comunidades de aprendizaje autogestivas y autosuficientes.

El mayor desafío es lograr la motivación y participación activa del recurso humano, teniendo en cuenta los aspectos tecnológico disponibles. Trabajo colaborativo o groupware son palabras para designar el entorno en el cual todos los participantes del proyecto trabajan, colaboran y se ayudan para la realización del proyecto.

Otro elemento del trabajo cooperativo es que "ayuda a potenciar aspectos afectivos actitudinales y motivacionales y genera aspectos muy positivos para el logro de los aprendizajes" Martí (1996, 54); el hecho de pertenecer a un grupo con un objetivo en común permite estrechar lazos en los participantes y les genera sentido de pertenencia. Claro está que el objetivo de un

trabajo colaborativo es producir algo, que puede ser un conocimiento o un objeto tangible; pero en ocasiones el hecho de pertenecer a una organización también puede ser el objetivo de los participantes; por lo tanto la motivación puede ser también intrínseca del propio proyecto.

Este estudio se basa en investigaciones previas relacionadas con Entornos Colaborativos y metodologías ágiles y se propone una aplicación concreta de éstos avances, propios de la disciplina informática, a la indagación, elaboración e implementación de un proyecto de aplicación, que permita su uso en una cátedra determinada, pero plenamente extensible a sectores organizacionales - empresariales como aporte a la comunidad.

## 2. LINEAS DE INVESTIGACIÓN y DESARROLLO

El estudio tomará como unidad de análisis la cátedra de Sistemas y Organizaciones de la carrera Ingeniería en Sistemas de Información de la Facultad Regional Córdoba, pretendiendo contribuir con ella, con una propuesta de aplicación de un proyecto de entorno colaborativo o groupware, enmarcados en conceptos de metodologías ágiles (SCRUM), e implementada en plataforma Moodle.

Las tecnologías de la información y la comunicación (TIC) no suponen por sí mismas una garantía de cambio positivo en la universidad, y aparecen nuevos retos que es necesario afrontar: nuevos programas, incorporación de docentes por concursos, el control de calidad de los materiales y servicios virtuales, buenas prácticas docentes en el uso de las tic.

La siguiente pregunta que se podría plantear es ¿Cuál es la importancia dentro de las cátedras que se le otorga a éstas actividades y a los sistemas informáticos en general? Otro interrogante a analizar es el siguiente: las actividades vinculadas al trabajo en grupo ¿son colaborativas entre los ingresantes o los procedimientos que se utilizan no son eficientes para el dominio en el cual se trabaja? Seguramente estos interrogantes se podrán responder en el transcurso de la investigación que se realiza.



### 3. RESULTADOS Y OBJETIVOS

Las nuevas metodologías se deben relacionar con algún aspecto existente y específicamente relevante de la estructura cognoscitiva. Si ello es así, nos planteamos ¿Por qué no tomar conceptos de las metodologías ágiles y aplicarlas a entornos colaborativos en un proyecto específico?.

#### Grado de Avance

La propuesta se encuentra en un grado de avance de inicio, realizando:

- Búsqueda Bibliográfica para la definición del proyecto.
- Identificación de los recursos necesarios para el desarrollo del proyecto.
- Análisis de factibilidad y viabilidad del estudio propuesto.
- Elaboración de un instrumento, como cuestionarios, para captar tanto de fuentes primarias como secundarias aspectos que se consideran relevantes para la investigación.

Es importante destacar que desde las actividades académicas y profesionales de los participantes se han abordado distintas instancias de estudio y aplicación de las herramientas de Entornos colaborativos y técnicas de Ingeniería de Requerimientos, necesarias para la obtención de información y análisis de la mismas, por lo cuál este estudio de aplicación reviste especial interés.

#### Objetivos Generales:

Investigar las características principales de los entornos colaborativos bajo una metodología de administración de proyectos (SCRUM), identificando ventajas y desventajas de su uso.

Promover la realización de un Proyecto en Entornos Colaborativos utilizando Plataforma Moodle bajo un marco de trabajo SCRUM.

#### Objetivos Específicos

- Lograr que el grupo de investigación adquiera conocimientos y experiencia en el desarrollo de un proyecto bajo una

metodología ágil poniendo énfasis en pautas de Scrum.

- Profundizar el conocimiento del uso de una metodología ágil (aplicada en la actualidad en proyectos de desarrollo de sistemas), y considerar su aplicación en los procesos críticos de las Organizaciones.

- Diseñar una propuesta teórica-práctica que permita la implementación de procesos de reflexión, análisis y promoción de las habilidades, en el uso de entornos colaborativos, en ingresantes a la carrera de Ingeniería en Sistemas de Información, Regional Córdoba - específicamente en la cátedra Sistemas y Organizaciones

- Definir los impactos, positivos y negativos, de la Plataforma Moodle como medio de comunicación de información ( amplitud, organización, interacción, aspectos técnicos) en el proceso de enseñanza y aprendizaje; e identificar las dificultades, necesidades y oportunidades que se presentaron, tanto con los estudiantes como con los profesores, de tal modo que se pueda corregir en un próximo abordaje.

- Identificar las apreciaciones particulares de los alumnos con respecto a las necesidades de información y conformar documentos que contemplen diversas visiones.

### 4. FORMACION DE RECURSOS HUMANOS

#### Transferencia de los resultados:

En el plano académico, los resultados que se generan a partir de ésta investigación se transforman en insumos para la concreción de nuevas instancias de aprendizaje en las aulas universitarias.

Es de notar que los informes finales de la investigación realizada se colocará a disposición del Departamento de Sistemas de la Universidad.

Participan en el Proyecto:

Alumnos:

A partir de la participación en las etapas de investigación, realizando las

actividades de recolección de datos e información inherente a la identificación de requerimientos de software.

La aplicabilidad será directa, ya que la carrera de Ingeniería en Sistemas, dentro de su currícula posee contenidos de conocimientos de las diferentes gestiones de las organizaciones y los tipos de entrevistas a llevar a cabo para la identificación de las necesidades de los alumnos y los Requerimientos de información para un desarrollo de sus actividades académicas.

#### Ayudantes Alumnos:

Suman su participación en el proyecto alumnos que se desempeñan en la actualidad como ayudantes Alumnos por lo que se iniciarán en las tareas formales de investigación.

La aplicabilidad de los conocimientos adquiridos en sus Trabajos Finales o Proyectos Finales de Tesis sumará valor a su labor.

#### Colaboradores directos:

Tomarán experiencia incorporándose a la Iniciativa Académica de Investigación y harán aporte de Conocimiento a la Sociedad.

## 5. REFERENCIAS

- León, Orfelio. Diseño de Investigación. Líneas de Investigación. México McGraw Hill 2000 8448110889.
- Taylor,S. Líneas de investigación . Introducción a los métodos cualitativos de investigación España Paidós 2002.
- Líneas de investigación Sierra Bravo, Restituto Técnicas de investigación social: teoría y ejercicios España Paraninfo.
- Senn, James. Analisis and Design of Information Systems, McGraw Hill, 1989.
- Sommerville Ian. Ingeniería de Software, Addison Wesley, 2002.Sexta Edición.
- Rumbaugh, Jacobson, Booch. The Unified Modeling Language Reference Manual, Addison Wesley 1998.
- Booch, Rumbaugh, Jacobson. Lenguaje de Modelado Unificado, Addison Wesley,Año 1999.
- Métodos Agiles: Una alternativa Real y competitiva a los procesos tradicionales de desarrollo. Prioto Sebastian editorial: Gradi

## Sistemas ubíquos: Desarrollo y Aplicaciones.

Hugo Ramón, Javier Charne, Carlos Di Cicco, Diego de la Riva, Ana Smail, Germán Osella Masa

Instituto de Investigación y Transferencia en Tecnología (ITT),  
Escuela de Tecnología, Universidad Nacional del Noroeste de la Provincia de Buenos Aires  
(UNNOBA)

{hugoramon, javier, carlosdicicco, delariva, anasmail}@unnoba.edu.ar  
{german.osella}@nexo.unnoba.edu.ar

### Resumen

Se presenta un proyecto de investigación aplicada al cómputo ubicuo (*ubicomp*), abarcando aspectos que van desde los fundamentos del desarrollo (técnicas de Ingeniería de requerimientos, metodologías de gestión, desarrollo de proyectos, técnicas de planificación, métricas, normas de calidad, redes de sensores y algoritmos inteligentes y procesamiento de señales) hasta la concepción de aplicaciones específicas orientadas a los procesos productivos de la región y el país, procesos de e-gobierno y e-ciudad, desarrollos específicos para televisión digital y la aplicación de tecnología informática para la educación en ambientes ubíquos.

**Palabras clave:** e-Learning, e-Government, TIC, educación superior, cambio cultural.

### Contexto

Este proyecto fue acreditado mediante evaluación externa y financiado por la Secretaría de Investigación, Desarrollo y Transferencia de la UNNOBA en el marco de la convocatoria a Subsidios de Investigación Bianuales (SIB2013).

Se desarrolla en el Instituto de Investigación en Tecnologías y Transferencia (ITT) dependiente de la mencionada Secretaría, y se trabaja en conjunto con la Escuela de Tecnología de la UNNOBA. Y s

Se coordina con otros dos proyectos presentados por el ITT, uno relacionado a la Generación y Gestión de Contenidos para Educación Ubicua y el otro relacionado a la Gestión de la Tecnología Aplicada en Contextos Educativos. Además se relaciona con el proyecto “*Certificación de Calidad y Digitalización de Procesos en Organizaciones presentado*” del Instituto de Políticas y Gobierno (IPG). Esto muestra la interdisciplinaridad y transversalidad de los proyectos.

### Introducción

Computación ubicua es el término que se le da a la tercera era de la computación. La primera corresponde a los mainframes (una gran computadora compartida de una compañía), la segunda a la computadora personal y la tercera corresponde a la computación ubicua caracterizada por productos de cómputos portables conectados.

Ejemplo de esto son los teléfonos inteligentes, las tablets y las computadoras embebidas en todos los dispositivos de uso cotidiano.

Desde hace diez años se vienen introduciendo constantes mejoras en los teléfonos llamados inteligentes, hasta superar el hecho de ser sólo celulares al integrar poder de cómputo y comunicación en un equipo que entra en un bolsillo. Otros ejemplos son las computadoras embebidas incorporadas en los SmartTV que permiten el desarrollo de la televisión interactiva; en los automóviles que proveen seguridad activa y pasiva; en los artefactos del hogar, y en los edificios inteligentes.

A fin de 2012, según estimaciones del Pronóstico Global de Tráfico Móvil de Datos de Cisco, habrá más dispositivos inteligentes que personas [1]. Morgan Stanley, en un informe de 2009 llamado “*The Mobile Internet Report*”, indica “... *La Internet móvil está aumentando más rápido que lo que lo hizo Internet de escritorio y creemos que en cinco años, más usuarios podrán conectarse a Internet a través de dispositivos móviles que desde computadoras de escritorio. Los puntos clave para esto son:*

- *Cinco Productos basados en IP que están en continuo crecimiento y convergiendo para el uso de internet móvil: 3G + redes sociales + video + VoIP + dispositivos móviles.(saqué “impresionantes”).*
- *Las plataformas de Apple y Facebook sirven para descubrir cómo los usuarios se conectan y comunican; la participación que tiene el desarrollador en esto no tiene precedentes.*

- *El Crecimiento masivo de datos móviles está impulsando el desarrollo de los carriers y proveedores de equipos.*
- *Los mercados emergentes tienen un potencial increíble dado el crecimiento de usuarios de Internet móvil. La baja penetración de telefonía fija y el valor agregado que se le está dando a los servicios móviles indican que Internet será móvil....” [2].*

Relacionando esta información con el desarrollo de aplicaciones y buscando lograr la integración (característica fundamental de los entornos de computación ubicua) las aplicaciones deben implementar mecanismos para descubrir las necesidades de los usuarios móviles con el fin de presentar la información pertinente en el lugar correcto y en el momento adecuado.

Esta y otras características intrínsecas de los sistemas de computación ubicua, hace necesario el uso de diferentes técnicas de ingeniería de software. El desarrollo orientado a servicios, desarrollo basado en componentes, plugins basados en arquitecturas, sistemas basados en eventos y la evolución del software dinámico son las principales técnicas que se pueden utilizar en la ingeniería de software para sistemas ubicuos.

El factor limitante de los dispositivos para sistemas ubicuos es la energía, dado que generalmente los dispositivos son portables o embebidos y utilizan baterías con lo cual el consumo de energía impacta directamente en su desempeño, es necesario investigar sobre protocolos de comunicaciones, oportunidad de carga/descarga y utilización de energía cinética en estos dispositivos.



La computación invisible es central y permite definir escenarios para aplicaciones ubicuas. Cuando hablamos de invisible nos referimos a psicológicamente invisible. Este tipo de aplicaciones monitorean el comportamiento humano, de una casa, de un hospital o del tránsito, proveyendo control inteligente y utilizando programación no tradicional.

Este proyecto trata sobre la computación ubiqua y dado que es un campo en donde se dan progresos rápidos y simultáneamente, es un campo de investigación joven y amplio.

La oportunidad de I+D+i se da en adaptar/diseñar aplicaciones o a la forma de construir esas aplicaciones las cuales gestionan información procedente de redes de sensores compuestas por diversos nodos distribuidos y localizadas en espacios cerrados (hospitales, fábricas, etc.) o abiertos (campos, bosque, etc.). Ambos espacios se encuentran en las diferentes actividades económicas de la región de aplicación del proyecto.

La investigación en computación ubiqua se puede dividir en los siguientes aspectos:

- *Sistemas*: hace foco en cómo construir frameworks de desarrollo para aplicaciones ubicuas, los aspectos a tener en cuenta en la infraestructura para soportar estas aplicaciones y la privacidad inherente a información personal.
- *Experiencia*: destaca los aspectos críticos del usuario relacionado a la interfaces gráficas y de acuerdo a la etnografía.
- *Sensores*: que permiten determinar y analizar el contexto del usuario al cual las aplicaciones deben

reaccionar. Además es de interés poder estudiar la posibilidad de involucrar todos estos requerimientos en una ambiente de desarrollo simulado, permitiendo con esta herramienta la evaluación de aplicaciones ubicuas sin el costo asociado al desarrollo de los mismos.

Las contribuciones al cómputo ubiquo se le atribuyen a Mark Weiser [3] (Weiser, 1993) cuando trabajaba en XeroxPARC. En ese momento pensó en una época en donde las tecnologías de la computación aparecerían en todos los artefactos de la vida moderna para dar soporte a las actividades diarios del trabajo, del hogar y del ocio y sobre todo que esas tecnologías deben pasar desapercibidas. Los dispositivos móviles y los sensores pueden comunicarse entre sí con la infraestructura circundante para proveer soporte a las actividades. Esto se puede hacer en estos momentos debido a las capacidades que nos dan conectividad *Bluetooth*, *RFID*, *Zigbee*, *WiFi* y muchas otras.

Esta visión comienza a hacerse realidad aproximadamente en 2005 en donde se lanza el *iPod*, la zapatilla *adidas\_1* y el robot *Roomba*, estos productos no se ven como una computadora tradicional. En 2005 también se dan factores en la industria que permiten diseñar productos de acuerdo a la visión de Weiser:

- El precio de las CPU han caído hasta el punto de considerarlo sin costo.
- Internet es claramente social y comercial.
- Se han desarrollado protocolos estándares de intercambio de datos y comunicaciones.

- Se establece la telefonía celular (o digital) definiendo el concepto de teléfonos inteligentes.
- Se estandarizó la comunicación wireless.
- De acuerdo al boom de las punto como se desarrollaron productos interactivos.

Con los factores mencionados, diversos investigadores han generado muchos términos relacionados al mismo concepto, *la satisfacción de las necesidades de los usuarios de una manera autónoma, omnipresente e imperceptible* (Kuniasvsky, 2010):

- *Computación Ubiqua* referida a embeber procesamiento de información y comunicaciones en todos los ambientes para poder proveer servicios continuos.
- *Computación Física* que describe como las personas interactúan con las computadoras a través de objetos físicos más que a través de un ambiente on-line.
- *Computación Pervasiva* que se refiere al nuevo modelo tecnológico. Inteligencia Ambiental que describe como los dispositivos se integran en forma inteligente en los espacios construidos por los humanos y pasan a ser parte del ambiente. La Internet de las cosas en donde cualquier objeto físico puede ser digitalmente identificado en forma análoga a como la información digital es organizada en la web.

Los primeros prototipos que se desarrollaron fueron *LiveBoard* (Scott Elrod, 1992), *ParcPad* y *ParcTab* (Want, 1995) desarrollados en XeroxPARC entre 1988-1994. Debido al éxito con el que resultaron fueron comercializados posteriormente.

Como se ha dicho la naturaleza de la computación ubicua de interactuar con diversos dispositivos utilizando el entorno, requieren de mecanismos capaces de localizar a las personas y otros dispositivos, necesarios para la realización de sus actividades, de una manera sencilla y con menos interacciones. La localización fue atribuido al trabajo definido en *Active Badge system* (Andy Harter, 1994).

También se han desarrollado aproximaciones al modelado de contexto para el entorno académico, de congresos, aeropuertos, estaciones de trenes, turismo, automóviles entre muchas otras.

La noción de contexto es importante, existen frameworks (Kaori Fujinami, 2005) para el prototipado rápido de aplicaciones dependientes del contexto (Dey, 2001) y existen esfuerzos para definir un estándar de desarrollo para aplicaciones ubicuas (Martin Modahl, 1994) y en el dominio de redes dinámicas, tecnologías inalámbricas, protocolos de malla o servicios ubicuos también son diseñados y evaluados utilizando simuladores de red como MANET (Laurent Ciarletta, 2008).

Existen numerosos congresos como *The Eighth International Conference on Pervasive Computing* [4], *IEEE International Conference on Pervasive Computing and Communications* [5] y revistas como *IEEE Pervasive Computing* [6] en donde se ve claramente que la computación ubicua representa un gran desafío, siendo esta un área atractiva para definir líneas de investigación estratégica para el desarrollo de una región y el país.

### Líneas de investigación y desarrollo

Se mencionan aquí las principales líneas de investigación y desarrollo abordadas en el marco del proyecto:

- Investigar/Analizar:
  - La Ingeniería de software para aplicaciones no tradicionales.
  - Las tecnologías de Redes de sensores aplicados a aplicaciones ubicuas.
  - Algoritmos inteligentes aplicados a cómputo ubiquesto.
  - Procesamiento de señales en cómputo ubiquesto.
  - Privacidad en computación Ubiqua.
  - Interfaces Ubiquas.
  - Ubicación de objetos y personas en cómputo ubiquesto.
  - Framework de desarrollos de aplicaciones ubicuas.
- Realizar Actividades de Inteligencia Competitiva.
- Proponer una Metodología para desarrollo de aplicaciones ubicuas.
- Analizar:
  - Oportunidades de publicación/transferencia.
  - Metodologías de integración de las tecnologías.
  - Oportunidades de mejoras/adaptación/desarrollo de frameworks.

### Resultados y Objetivos

El propósito de este proyecto es, entre otros, revisar los desafíos involucrados en los sistemas ubicuos, así como presentar una perspectiva de la ingeniería de software para hacer frente a estos respondiendo a preguntas como:

- ¿Cómo las fallas y mal funcionamiento del sistema podrían ser detectados en forma transparente y automática?
- ¿Cómo afecta la escala de los sistemas de computación ubicua esta detección transparente?
- ¿Qué pasa con las propiedades en tiempo real de un sistema ubiquesto?
- ¿Qué aspectos de estos sistemas deben proporcionar garantías de tiempo real?
- ¿Qué métodos son necesarios para verificar la corrección de estos sistemas?
- ¿Cuáles son los requerimientos para un framework de aplicaciones ubicuas?
- ¿Se puede implementar un ambiente de simulación genérico de estos sistemas ubiquesto?

Al mismo tiempo se plantea la posibilidad de implementar un entorno simulado para producir/evaluar/prototipar aplicaciones ubicuas y que éste sea un producto de transferencia al medio con licenciamiento (GPLv2, GPLv3 o Creative Commons de acuerdo al caso).

Para el abordaje del problema planteado se propone una investigación aplicada satisfaciendo los siguientes objetivos:

- Investigar sobre los frameworks de desarrollo de aplicaciones ubicuas.
- Investigar sobre protocolos de redes de sensores y dispositivos heterogéneos.
- Selección/Adaptación/Desarrollo de un framework para el desarrollo de aplicaciones ubicuas.
- Selección/Adaptación/Implementación de un ambiente de simulación de aplicaciones ubicuas.

Este proyecto de investigación tiene indirectamente otros objetivos:

- Avanzar en la capacitación de postgrado de los miembros de proyecto.
- Redacción, exposición y defensa de las Tesis de Magister, Tesis, Prácticas Profesionales Supervisadas además de la publicación y/o transferencia de resultados obtenidos.
- Interacción con otros grupos de I/D del país y del exterior en la temática del proyecto.

Así mismo, al finalizar el proyecto, se busca que todos los productos desarrollados por este proyecto sean transferidos a empresas/organismos de la región UNNOBA y/o del país.

### Formación de Recursos Humanos

En esta línea de I/D se esperan producir en los próximos dos años, al menos, la dirección de dos Tesinas de Licenciatura en Sistemas, tres Prácticas Profesionales Supervisadas de la Ingeniería en Informática, la dirección de un Trabajo de Especialista, y la dirección de dos Tesis de Magister.

### Referencias

[1] Pronóstico del Índice de Red Visual de Cisco (VNI) para el Tráfico de Datos Móviles Global entre los años 2012 y 2017, [http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white\\_paper\\_c11-520862.html](http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.html)

[2] Stanley M., "The Mobile Internet Report", 2009, <http://www.morganstanley.com/about/press/articles/4659e2f5-ea51-11de-aec2-33992aa82cc2.html>

[3] Weiser M. Ubiquitous Computing, IEEE Computer "Hot Topics", October 1993, <http://www.ubiq.com/hypertext/weiser/UbiCompHotTopics.html>

[4] The Eighth International Conference on Pervasive Computing, <http://pervasiveconference.org/2012/>

[5] IEEE International Conference on Pervasive Computing and Communications, <http://www.percom.org/>

[6] IEEE Pervasive Computing, <http://www.computer.org/portal/web/computingnow/pervasivecomputing>

### Bibliografía

- Weiser M. The Computer for the Twenty-First Century. Scientific American, Vol. 265, No. 3., pp. 94-104. 1991.
- Krumm J., Creating a New Book of Tutorials: Ubiquitous Computing Fundamentals, Microsoft Research Redmond, UbiComp 2009.



- Bravo C.; Redondo M.; Ortega M.; Bravo J. Evolución de un Entorno Colaborativo de Enseñanza Basado en Escritorio hacia la Computación Ubicua, (2002) COLINE'02 Investigación en Entornos de Interacción Colectiva, Workshop de Investigación sobre nuevos paradigmas de interacción en entornos colaborativos aplicados a la gestión y difusión del Patrimonio cultural. Granada, 11 y 12 de Noviembre del 2002.
- Hervás R., Nava S., Chavira G., Bravo J., Modelado de Contexto: Una Ontología Adaptativa al Usuario en Ambientes Inteligentes. Informe técnico. 2007.
- Carmona M.; González S.; Castro Ruiz, Innovación Tecnológica en Comunicaciones Móviles Desarrollada Con Software Libre: Campus Ubicuo Novática: Revista de la Asociación de Técnicos de Informática, ISSN 0211-2124, N°. 190, 2007.

## INGENIERIA DE SOFTWARE APLICADA A UN SISTEMA DE GESTION DE CALIDAD EN CENTROS EDUCATIVOS

Jorge A. Silvera, Daniel Arias Figueroa, Gustavo D. Gil, Ernesto Sánchez, Calos I. Orosco  
 Centro de Investigación y Desarrollo en Informática Aplicada (C.I.D.I.A.)  
 Facultad de Ciencias Exactas. / Universidad Nacional de Salta

Av. Bolivia 5150 - 0387-4255476

Mail: [daaf@cidia.unsa.edu.ar](mailto:daaf@cidia.unsa.edu.ar) ; [jsilvera@cidia.unsa.edu.ar](mailto:jsilvera@cidia.unsa.edu.ar)

### RESUMEN

Este trabajo presenta los resultados obtenidos producto de la línea de investigación sobre Sistemas de Gestión de Calidad en la Universidad Nacional de Salta, referido al estudio de la familia de Normas ISO aplicadas a Instituciones Educativas, en particular al Centro de Investigación y Desarrollo en Informática Aplicada (C.I.D.I.A.), dependiente de la Facultad de Ciencias Exactas.

Es sabido que cada vez son más las organizaciones que implementan y certifican un sistema de gestión de calidad (SGC) que enfoque los procesos de la misma y oriente los resultados hacia el cliente, con un adecuado nivel de eficacia y eficiencia. Concretamente la familia de Normas ISO 9001, es un modelo que plantea un conjunto de aspectos clave para una Gestión de la Calidad basada en Procesos, asegurando la calidad del producto o servicio, e incrementando la satisfacción del cliente.

Las entidades educativas, no son ajenas a esta tendencia, tanto es así que el Instituto Argentino de Normalización (IRAM), a través del Comité General de Aseguramiento de la Calidad, diseñó en marzo de 2001 “*La Guía de Interpretación de la Norma ISO 9001 para la Educación*” [9] con el objetivo de proveer lineamientos de cómo comprender e implementar la norma ISO 9001 en el ámbito de la educación.

Esta investigación se plantea desde dos perspectivas diferentes, por un lado se plantea estudiar las directrices que guían a los centros que presten servicio de capacitación y formación en todos los niveles, en cuanto a la

implementación de un sistema de gestión de calidad eficaz que cumpla los requisitos de la norma ISO 9001:2008) [8]. Una segunda perspectiva más tecnológica plantea el análisis, diseño e implementación de un prototipo funcional de un sistema de gestión de calidad (SGC) para el Centro de Investigación y Desarrollo en Informática Aplicada (C.I.D.I.A.), utilizando como metodología Web Modeling Language (WebML) [10] y el CASE WebRatio [14]. Esta investigación servirá de referencia para cualquier implementación de calidad, dentro del ámbito educativo, y en especial en los centros educativos.

### **Palabras claves:**

*Normas ISO 9000, Sistema de Gestión de Calidad (SGC), Herramientas informáticas de apoyo al SGC. ISO en Capacitación. Metodología WebML (Web Modeling Language).*

### CONTEXTO

El Centro de Investigación y Desarrollo en Informática Aplicada – C.I.D.I.A. funciona en la Facultad de Ciencias Exactas de la Universidad Nacional de Salta desde junio del año 2002, fue aprobado bajo RES. C.D. CS. Ex. N° 167/02 – Expediente N° 8545/01. El mismo nace como una evolución del L.I.D.T.I. – Laboratorio de Investigación y Desarrollo en Tecnologías Informáticas que funcionaba desde el año 1997 y que dependía del Departamento de Matemática de la misma Facultad.

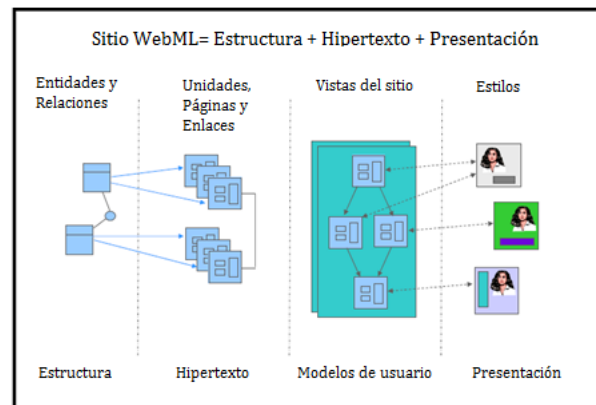
Entre los objetivos generales del C.I.D.I.A. podemos mencionar:

- Nuclear y apoyar a un importante grupo de docentes, profesionales e investigadores en áreas directamente vinculadas con las Nuevas Tecnologías Informáticas, facilitando las tareas de investigación y desarrollo.
- Desarrollar investigaciones científicas, tanto básica como aplicadas, individuales o interdisciplinarias, principalmente en los campos de las Redes de Comunicaciones de Datos, Internet, Informática aplicada en Educación y la Ingeniería de Software.
- Promover la capacitación en nuevas tecnologías informáticas a todos los niveles.

## 1. INTRODUCCION

La implementación exitosa de un sistema de calidad (SGC) aporta un gran número de beneficios a las organizaciones en general, y en este caso concreto a las educativas, que apuestan por esta estrategia. No sólo reducen sus costos de manera razonable, sino que además ayudan a lograr la tan preciada satisfacción de sus clientes, siendo una gran motivación para los integrantes de la misma. El enfoque de la metodología WebML combina componentes tradicionales bien conocidos por los desarrolladores, como el diseño conceptual de datos usando el modelo Entidad- Relación y la especificación de los casos de usos usando UML, con nuevos conceptos y métodos para el diseño de hipertextos, que son fundamentales para el desarrollo web. No obstante, el valor del enfoque de la propuesta no está en los componentes individuales, sino en la definición de un marco sistemático para que las actividades de desarrollo de aplicaciones web pueda ser organizado de acuerdo a los principios fundamentales de la ingeniería de software, de modo que todas las tareas, incluidas las más propias de la web, encuentren el soporte adecuado en los conceptos, las notaciones y técnicas adecuadas. La característica distintiva de este marco de desarrollo es el énfasis en el modelo conceptual. El modelo conceptual ha

demostrado éxito en muchos campos del software, y en diseño de base de datos, donde el modelo Entidad- Relación ofrece una notación de alto nivel e intuitiva para la comunicación de los requisitos de información entre los diseñadores y no técnicos, y es la base para la creación de esquemas de bases de datos de alta calidad. En esencia, WebML consiste en gráficos simples y conceptuales para expresar un hipertexto como un conjunto de páginas y operaciones, WebML representa una página como una estructura compuesta por unidades de contenido y links. WebML propone que un sitio web conste de tres grandes partes conceptuales, la estructura, el hipertexto y la presentación.



**Figura 1- Conceptos principales de WebML.**

La metodología está compuesta por la creación de los Modelos de Datos, Hipertexto, Presentación y Personalización. En este trabajo se desarrollaron todos los modelos de la metodología y para la implementación se utilizó la herramienta WebRatio [14] con licencia de uso gratuito para el modelado de procesos de negocios (BPM). Los motivos de la elección de la herramienta de desarrollo son los siguientes:

- Dispone de funcionalidades necesarias para diseñar modelos de datos y de navegación utilizando la metodología elegida WebML.
- WebRatio crea un único fichero XML con todos los modelos definidos. La sintaxis de estos ficheros no puede considerarse excesivamente complejo.

- Permite utilizar prácticamente cualquier gestor de base de datos (DBMS).
- Crea de forma automática el modelo de datos que incluye las tablas, atributos y relaciones indicados en el modelo conceptual.
- Genera automáticamente código Java de una aplicación web a partir del modelo de navegación.
- Distintas ediciones de la herramienta, gratuita, personal, estándar y Enterprise, de acuerdo a la necesidad del usuario.

## 2. LINEAS DE INVESTIGACION y DESARROLLO

Los principales ejes temáticos que se están investigando son los siguientes:

- Tecnología Informática aplicada en Educación.
- Gestión de Calidad aplicada a Centros o Instituciones de Capacitación.
- Herramientas informáticas y software libre para la implementación de un SGC ISO 9001.
- Aplicación de la metodología WebML (Web Modeling Language) para el diseño de una plataforma integral que soporte una implementación ISO 9001.

## 3. RESULTADOS OBTENIDOS/ESPERADOS

Desde el momento de su creación, los requisitos de la norma internacional ISO 9001 son genéricos y aplicables a todas las organizaciones sin importar su tipo, tamaño o producto o servicio suministrado. En particular el sistema de gestión de calidad (SGC) en un centro de capacitación está influenciado por los diferentes objetivos educativos, por las definiciones del proyecto educativo institucional, por los diversos métodos de enseñanza y por las prácticas

administrativas específicas de cada institución.

Como hemos podido observar, un sistema de gestión de la calidad se basa en muchos documentos y en un manejo muy estricto, y a su vez documentado, de esos documentos.

Esta es un área en la que claramente el apoyo de la tecnología informática es imprescindible. No alcanza con un procesador de textos y un espacio de almacenamiento compartido en un servidor de archivos.

Todo documento relacionado con el sistema de gestión de la calidad debe ser desarrollado a través de un proceso perfectamente documentado. Además, una vez aprobado, debe estar disponible para todos quienes participan en el SGC, y por otra parte no disponible para quienes no están involucrados en el. A su vez las sucesivas revisiones deben quedar claramente identificadas así como los cambios realizados. También, si los documentos obsoletos se mantienen en el SGC para poder ser consultados deben quedar claramente identificados como obsoletos para impedir que sean utilizados como actuales, por error. Por ejemplo, debe impedirse su modificación incluso a personal autorizado para generar documentos.

Esto es claramente el manejo de una base documental asociada a un proceso de decisión y de elaboración conjunta de los que típicamente se realizan con herramientas de trabajo en grupos.

En función a lo expresado, tomando como referencia las especificaciones y directrices investigadas por el C.I.D.I.A. para la aplicación de la familia de norma ISO en centros educativos [1], se realizó el análisis y diseño un SGC, utilizando la metodología WebML, logrando la implementación de un prototipo funcional de una herramienta integrada y genérica para acompañar una implementación de calidad bajo los requisitos de la norma ISO 9001:2008 en un centro educativo.

Por último, pero no por eso menos importante, también definimos como objetivos del proyecto, alcanzar la certificación ISO 9001:2008, de los procesos



más importantes, a través de alguno de los organismos oficiales.

En síntesis, los resultados obtenidos a la fecha, producto de los trabajos de tesis defendidos (detallados en el punto 4) son:

- El Análisis y diseño de un Sistema de Gestión de Calidad para el C.I.D.I.A. en etapa de implementación.
- Un prototipo funcional de un Sistema de Gestión de Calidad para centros educativos.

#### 4. FORMACION DE RECURSOS HUMANOS

La estructura del equipo de investigación es de 5 (cinco) miembros incluidos el Director y Co-director.

Uno de sus miembros obtuvo la Especialidad en Ingeniería de Software de la Universidad Nacional de La Plata, con el trabajo “Sistema de Gestión de Calidad bajo Normas ISO”.

Otro de sus miembros alcanzó el título de Licenciado en Análisis de Sistema, otorgado por la Universidad Nacional de Salta, con la tesis “Análisis y Diseño de un Sistema de Gestión de Calidad basado en la Norma ISO 9001”.

Además, otro de sus miembros se encuentra realizando el trabajo final del Magíster en Administración de Negocios de la Universidad Católica de Salta denominada “Calidad ISO en Centros de Capacitación y Formación”.

Continuamos con la dirección de tesis de grado de alumnos de la carrera Licenciatura en Análisis de Sistemas de la Universidad Nacional de Salta.

#### 5. BIBLIOGRAFIA

[1] Normas ISO y su Aplicación en Centros Educativos. Arias Figueroa, Daniel y otros. XIII Workshop de Investigadores en Ciencias de la Computación. Rosario. ISBN 978-950-673-892-1.

[2] Análisis y Diseño de Sistemas. Kendall E. Publicado por Prentice Hall Hispanoamericana. Edición año 1997.

[3] Análisis y Diseño de Sistemas. Pressman. Publicado por Prentice Hall Hispanoamericana. Edición año 1997.

[4] Ingeniería para los Sistemas Web: Una propuesta metodológica. Instituto politécnico nacional, centro de investigación en Computación. M. en I. A. Alejandro Peña Ayala, Dr. Agustín Gutiérrez Tornes. Año 2004.

[5] Process Modeling in Web Applications Marco Brambilla, Stefano Ceri, Piero Fraternali Dipartimento di Elettronica e Informazione, Politecnico di Milano, Italy Ioana Manolescu INRIA Futurs – LRI, PCRI, France. Año 2006.

[6] Ishikawa, Kaoru. ¿Qué es el control de la Calidad?: La modalidad Japonesa. Bogotá: Grupo Editorial Norma, 1997.

[7] ISO 9001:2000 - Documento Informativo. Paper elaborado por Christian Narbarte para el Instituto Valenciano de Educación y Calidad Educativa. Marzo 2007.

[8] Norma ISO 9001, 9001:2008 Elaborada por el Comité Técnico ISO/TC176 de ISO.

[9] Guía de Interpretación de la IRAM-ISO 9001 para la educación. IRAM 30000:2001.

[10] Web: <http://www.webml.org>. Sitio oficial de la Metodología Web Modeling Language.

[11] WebML para el diseño de la plataforma virtual de enseñanza GATMA. Seminario De sistemas, Universidad Nacional de Salta, 2006.

[12] Estudio e implementación de un sistema online de gestión de documentos usando la Metodología WebML. Seminario de sistemas, Universidad Nacional de Salta, 2009.

[13] Web: <http://www.qgestion.com> Página de una empresa gestora de sistemas ISO.

[14] Web: <http://www.webratio.com/>. Sitio oficial de la herramienta CASE WebRario.

# Transformación de modelos conceptuales mediante uso de XSLT

Juan Pablo Fernández Taurant, Claudia Castro, Marcelo Marciszack

Dpto. Ingeniería en Sist. de Información/ Facultad Regional Córdoba/ Universidad Tecnológica Nacional

{ jtaurant, ingclaudiacaastro, marciszack }@gmail.com

## Resumen

El objetivo de este trabajo es presentar el análisis de una alternativa que brinde soporte a la gestión de requerimientos de software, facilitando su trazabilidad. Concretamente, nos centraremos en el uso del lenguaje de transformaciones XSLT, que nos permitirá definir diferentes reglas de transformación, las cuales, utilizaremos para realizar transformaciones automáticas bidireccionales entre los modelos generados durante el proceso de captura de requerimientos, permitiendo mantener la trazabilidad de los mismos, independientemente de los cambios o modificaciones que puedan sufrir a lo largo de su ciclo de vida.

**Palabras clave:** *requerimientos, Transformación de modelos, conceptual, trazabilidad, XSLT, UML, BPMN.*

## Contexto

El presente trabajo se encuentra en el marco del proyecto “Validación de Requerimientos a través de Modelos Conceptuales”, que se encuentra consolidado dentro de la línea de investigación de Sistemas de Información en el Dpto. de Ingeniería en Sistemas de Información de la Universidad Tecnológica Nacional, Facultad Regional Córdoba.

El cual, busca dar solución a uno de los principales problemas de la ingeniería en

sistemas relacionado a la elicitación y especificación de requerimientos, que vincula las distintas etapas del proceso de desarrollo de software manteniendo la trazabilidad de los mismos hasta su validación e implementación.

## Objetivos de trabajo

Los objetivos del presente trabajo, se encuentran alineados dentro de los objetivos propuestos dentro del Proyecto de investigación, los cuales son los siguientes son los siguientes:

- Establecer marco teórico metodológico para la gestión de requerimientos de Software.
- Definir propuesta metodológica para la especificación de requerimientos.
- Validación de modelos conceptuales en la especificación de Requerimientos.
- Posibilitar en todo momento la trazabilidad de requerimientos desde su captura, modelado y seguimiento.
- Construir e implementar una herramienta de software para la gestión de requerimientos y transformación automática de modelos.

## Introducción

El proceso de captura de requerimientos es esencial para conocer las actividades de una organización y clave para el éxito del sistema de información a desarrollar, por lo que resulta de vital importancia mantener la trazabilidad de los mismos a lo largo del proceso de desarrollo. Por otra parte, están comprendidas en este proceso todas las actividades necesarias para crear y mantener los requerimientos de un sistema, estableciendo los lazos de comunicación entre el dominio del modelado del negocio y el dominio del sistema.

Sin embargo, esta comunicación no siempre es tan sencilla, ya que con frecuencia se ve dificultada por la existencia de una distancia semántica entre los dominios organizacional e informático[1], y por diferencias de formación y vocabulario entre las partes involucradas.

Para cerrar estas brechas de comunicación existen en la industria del modelado de negocio estándares ampliamente utilizados como Unified Modeling Language(UML)[2], y Business Process Modeling Notation (BPMN)[3]. Siendo BPMN más orientado al dominio organizacional, de fácil lectura, y más entendible para las partes.

Independientemente de la metodología utilizada, mantener la trazabilidad de requerimientos resulta una actividad compleja. Para esto, todos los modelos creados durante el proceso de desarrollo deben ser mantenidos y actualizados. La creación y mantención de modelos se realiza de forma manual, generando con gran frecuencia inconsistencias entre modelos, en detrimento de la trazabilidad de requerimientos.

En función de estas necesidades, sería de gran utilidad una herramienta que nos permita lograr la automatización total o parcial del

proceso de captura de requerimientos, en donde los requerimientos de software son obtenidos a partir del modelo de negocio realizando diferentes transformaciones entre modelos.

La aplicación de estas transformaciones permitirá que los modelos creados se mantengan actualizados, ya que los cambios realizados en uno de ellos serán propagados a los modelos restantes. De esta forma se contribuye a la trazabilidad de requerimientos, facilitando su gestión y validación.

- **Aplicación de transformaciones**

Las transformaciones entre modelos pueden llevarse a cabo de diferentes maneras. Una de ellas es mediante el estándar para transformación de modelos Query View Transformation Language (QVT) [4], creado y mantenido por la Object Management Group (OMG).

El principal problema que posee este estándar es la falta de disponibilidad en el mercado de herramientas de modelado con procesador QVT, más aún con soporte BPMN, ya que generalmente soportan sólo UML. Otro problema radica en la falta de interoperabilidad, ya que cada herramienta posee su propia representación interna de los modelos. Estos inconvenientes evidencian la necesidad de seleccionar una forma alternativa de transformación de modelos.

Otro tipo de herramientas disponibles para la representación y transformación de modelos son las basadas en Extensible Markup Language (XML) [5]. Para la representación de modelos se encuentran los estándares XML Metadata Interchange (XMI) [6] de la OMG, y XML Process Definition Language (XPDL) [7], creado por la Workflow Management Coalition (WfMC).

Por otro lado se encuentra el estándar Extensible Stylesheet Language

Transformations (XSLT) [8], también basado en XML, que permite realizar las transformaciones entre modelos.

- **Propuestas para la obtención de requerimientos**

Analizaremos dos propuestas para la obtención de requerimientos de software partiendo del modelo de negocio. Utilizaremos un dominio de ejemplo donde aplicaremos cada propuesta, indicando las ventajas y desventajas de cada una en función de los resultados obtenidos.

En la primera propuesta la obtención de requerimientos de software se realizará a partir del modelo de negocio en BPMN. En el que analista deberá seleccionar en forma manual las actividades de negocio que deben ser mapeadas al SI.

Para esto utilizaremos algún estereotipo de BPMN que servirá como marca para detectar si una actividad debe ser mapeada o no al SI. En principio, una de las opciones es utilizar el estereotipo “Script” de BPMN, que se utiliza para representar acciones que se realizar en forma automática.

Teniendo como objetivo liberar al analista de la tarea de seleccionar cada una de las

actividades, asumiremos que todas serán mapeadas al SI, exceptuando las indicadas por el analista, que representaremos con el estereotipo “Manual” de BPMN.

El siguiente paso consiste en generar mediante transformaciones XSLT, un diagrama de actividades UML por cada actividad de negocio que deba mapearse al SI. Cada diagrama de actividades generado deberá ser completado por el analista con las actividades de SI que correspondan.

El proceso inicial de transformación de BPMN a Diagrama de actividades UML (UML-AD) se realiza por dos razones. Una de ellas es debido a las similitudes existentes entre ambas notaciones. La otra se debe a la posibilidad de representar diferentes patrones de modelado con ambas notaciones, lo que nos permite realizar la transformación de un modelo a otro con relativamente poca pérdida de información.

Adicionalmente, obtendremos a partir de cada diagrama de actividades, diagramas de casos de uso (UML-UC) y Diagramas de clases de análisis mediante transformaciones XSLT.

Podemos observar esta propuesta en la figura 1, donde también se indica la etapa en la que se realizará la validación de requerimientos.

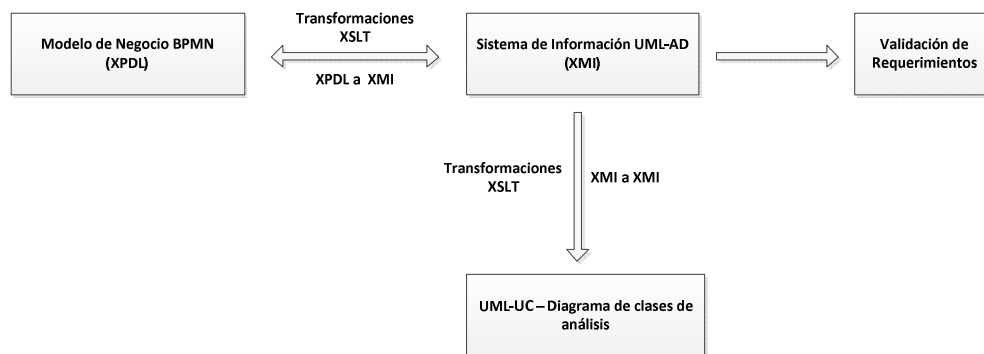


Figura 1 – Transformaciones BPMN a UML-AD

En la segunda propuesta se propone al igual que en la primera, la obtención de

requerimientos de software a partir del modelo de negocio en BPMN. En este caso



generaremos por medio de transformaciones un nuevo diagrama en BPMN para cada actividad del modelo de negocio que deba ser mapeada al SI. Cada diagrama se utilizará como plantilla para que el analista realice la descripción en forma manual de la actividad de negocio en términos de nuevas actividades BPMN que corresponderán cada una de ellas, a requerimientos del sistema de información a desarrollar. También obtendremos a partir de transformaciones XSLT diagramas de

actividades, diagramas de casos de uso y diagramas de clases de análisis que servirán como vistas compatibles con UML.

En la figura 2 se puede observar la estructura de la propuesta y sus transformaciones, indicando la etapa donde se validarán los requerimientos.

Posteriormente, evaluaremos los resultados obtenidos mediante la aplicación de ambas propuestas en un mismo dominio de ejemplo.

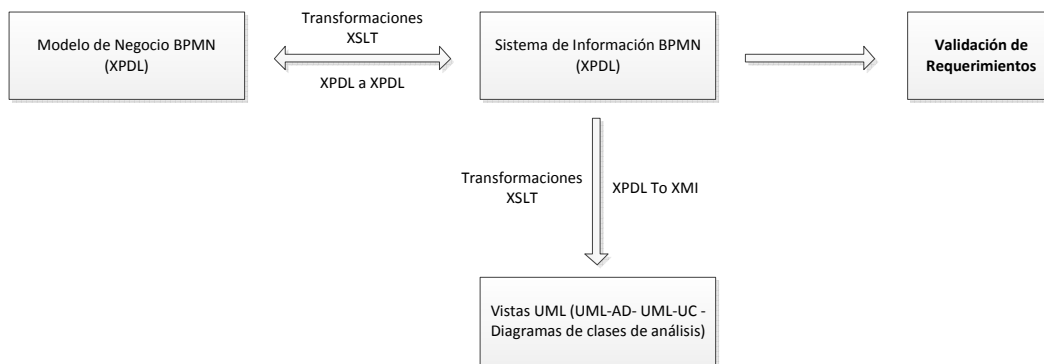


Figura 2 – Transformaciones BPMN a BPMN

## Resultados obtenidos

Analizando los resultados de la aplicación de cada propuesta en un caso de estudio de ejemplo, nos encontramos con que la obtención de requerimientos puede realizarse en forma parcial, ya que la pérdida de información producida en cada transformación se ve reflejada en los modelos resultantes. Dependiendo del tipo de modelo que se transforme podemos observar en el proceso la pérdida de relaciones, asociaciones, multiplicidades y hasta de elementos completos.

En la primera propuesta se plantea el uso de la notación BPMN solamente para el modelado de negocio. Mediante la transformación del modelo de negocio en

BPMN a UML-AD, el proceso de obtención de requerimientos se obtiene realizando transformaciones entre artefactos UML. La segunda propuesta en cambio, utiliza la notación BPMN para el modelado de negocio y para describir los requerimientos del SI, esto representa una desventaja si tenemos en cuenta que notación BPMN fue creada para modelar procesos de negocio.

Desde el punto de vista de la obtención de requerimientos a partir del modelo negocio, se hace evidente la necesidad de intervención del analista para la descripción de cada procesos de negocio.

Otro factor de gran importancia además de la obtención de requerimientos es su trazabilidad a lo largo de todo el proceso de transformación. En la primera propuesta, el

uso de diferentes notaciones hace que la pérdida de información asociada a la transformación entre artefactos sea mayor que en la segunda propuesta, que propone el uso de la misma notación tanto para modelar el negocio como los requerimientos del SI. Esta pérdida de información dificulta el proceso de generación de transformaciones inversas que nos permita reconstruir el modelo de nivel superior frente a los cambios realizados en los modelos de menor nivel, impidiendo la trazabilidad de requerimientos.

## Formación de Recursos Humanos

Este proyecto al estar circunscripto dentro de otro proyecto de investigación de I+D consolidado y al desarrollarse dentro del ámbito académico de la Universidad, prevé la transferencia de los conocimientos que se vayan adquiriendo, como así también la formación de los recursos humanos que participan del mismo. El mismo cuenta con tres becarios de investigación, dos tesis de la maestría en Ingeniería en Sistemas de Información y un doctorando de Ingeniería de Software que desarrollan sus trabajos en el ámbito del proyecto. Se prevé vinculaciones con otras redes de investigación, tratando de contribuir el conocimiento científico-técnico para su posterior transferencia a nivel social.

## Bibliografía

1. Taylor-Cummings, A., 1998. Bridging the user-IS gap: a study of major information systems projects. *Journal of Information Technology*, 13, 29-54.
2. OMG. Unified Modelling Language: Superstructure Version 2.0 (online), Julio 2005, <http://www.omg.org>
3. Object Management Group. Business Process Modeling Notation (BPMN). [http://www.omg.org/technology/documents/b\\_r\\_pm\\_spec\\_catalog.htm](http://www.omg.org/technology/documents/b_r_pm_spec_catalog.htm), version 1.2, 3 January 2009.
4. Object Management Group: MOF Query / Views / Transformations. Version 1.0, April 2008. [http://www.omg.org/technology/documents/modeling\\_spec\\_catalog.htm](http://www.omg.org/technology/documents/modeling_spec_catalog.htm).
5. World Wide Web Consortium: Extensible Markup Language (XML). Version 1.0(fifth edition), 26 November 2008. <http://www.w3.org/XML/>
6. Object Management Group: XML Metadata Interchange (XMI). version 2.1.1, 1 December 2007. [http://www.omg.org/technology/documents/modeling\\_spec\\_catalog.htm#XMI](http://www.omg.org/technology/documents/modeling_spec_catalog.htm#XMI).
7. <http://www.wfmc.org/xpdl.html>.
8. World Wide Web Consortium: XSL Transformations (XSLT). version 1.0, 16 November 1999. <http://www.w3.org/TR/xslt>

## Conceptos de dinamismo aplicados a servicios y workflows en BPMS basados en Cloud Computing

Lic. José Martínez Garro<sup>1</sup>, Mg. Patricia Bazán<sup>2</sup>

<sup>1</sup> Facultad de Informática UNLP <sup>2</sup> LINTI Facultad de Informática UNLP  
[josemartinezarro@gmail.com](mailto:josemartinezarro@gmail.com), [pbaz@ada.info.unlp.edu.ar](mailto:pbaz@ada.info.unlp.edu.ar)

### Resumen

Las organizaciones durante la última década han desarrollado BPM (*Business Process Management*) como metodología para la gestión de sus procesos de negocio. Diferentes cuestiones técnicas tales como la alta disponibilidad y la creciente conectividad, así como cuestiones económicas dadas por los costos de desarrollo y mantenimiento han llevado a gran número de organizaciones a optar por un modelo de *Cloud Computing*, donde muchos de estos riesgos tanto técnicos como económicos se ven disminuidos. En el caso de BPM, por tratarse de un paradigma naturalmente integrador, existen conceptos como dinamismo de *workflows* y servicios móviles y dinámicos que se resignifican al entrar en un contexto de *cloud*. En el presente artículo presentamos una serie de conceptos en la bibliografía actual que comienzan a reconsiderarse al insertar el paradigma de BPM dentro de la nube, así como propuestas de implementación para varios de ellos en un BPMS (*Business Process Management System*) particular presente en la industria.

**Palabras clave:** BPM, Cloud Computing, Mobile, SOA.

### Contexto

El presente artículo se enmarca dentro de un proyecto de investigación acerca de BPM, mejora continua de procesos y entornos de ejecución, encarado dentro del laboratorio LINTI en la Facultad de Informática de la Universidad Nacional de La Plata (UNLP).

### Introducción

Con el rápido desarrollo de IT en el contexto del lanzamiento y ejecución de arquitecturas basadas en *cloud*, las compañías se enfrentan con nuevos problemas. En particular, los procesos de negocio colaborativos ofrecen un potencial de optimización a través de la combinación de técnicas de *clouding* y BPM. Un factor común entre ambos es el enfoque flexible y ágil. El paradigma de computación en el *cloud* puede considerarse un facilitador de la

combinación mejorada de arquitecturas orientadas a servicios y un procedimiento ágil con respecto a la gestión de procesos de negocio. Pero este potencial depende de las condiciones de los distintos *frameworks*, las cuales pueden ser apreciadas desde un aspecto técnico tanto como económico.

*Vista técnica:* desde un punto de vista técnico se pueden identificar tres dimensiones para el diseño, implementación y operación exitosos de herramientas de BPM en un ambiente de *cloud*: programación, integración y seguridad.

- **Programación:** los sistemas complejos y distribuidos son altamente alcanzables en el campo de IT actual. En conexión con el objetivo de alcanzar una mayor usabilidad y flexibilidad, esta complejidad representa nuevos requerimientos para la Ingeniería de Software. Para resolver este problema es necesaria la adopción de nuevos lenguajes. De esta manera, recayendo sobre nuevos conceptos y técnicas innovadoras, los esfuerzos invertidos en desarrollo han sido reducidos a convertir la complejidad de estos nuevos aspectos en algo manejable.
- **Integración:** la integración puede ser dividida en integración de datos, integración de funciones e integración de procesos. A la luz de los desafíos involucrados, el tópico de la integración juega un rol fundamental en distintos escenarios. Por ejemplo, un *workflow* basado en *cloud* puede controlar actividades variables distribuidas más allá de las fronteras de las compañías. Para una ejecución no problemática de varias instancias de proceso se necesitan interfaces de integración y métodos estructurados.
- **Seguridad:** la seguridad puede ser dividida en tres categorías: seguridad funcional, de la información y de los datos. Todas estas categorías tienen una relevancia significativa para BPM, especialmente en cuanto a los *grids* de servidores de procesos de negocio. La seguridad funcional específica como el estado actual se corresponde con el estado deseado de funcionalidad. La

seguridad de información se enfoca en los cambios o extracciones de información no autorizados. La seguridad de los datos se encarga de los datos relacionados con el proceso.

Aún más, desde un punto de vista técnico, se debe responder a la pregunta de qué procesos de negocio son más apropiados para ejecutar en una arquitectura basada en *cloud*. Los riesgos posibles, como por ejemplo, opciones insuficientes de integración o interfaces de programación de aplicaciones deben ser tomados en consideración [1] [2] [3].

*Vista económica:* se pueden listar dos dimensiones desde el punto de vista económico:

- Disponibilidad: los servicios que son provistos por una infraestructura de *cloud* pueden ser accedidos en cualquier momento. Basados en un alto nivel de abstracción, la personalización e instalación se vuelven significativamente más fáciles. En adición a esta simplificación, el usuario final es capaz de trabajar con el servicio en forma inmediata.
- Riesgo de inversión: en el contexto de los distintos modelos de facturación variable como el de pago por transacción, el sistema orientado a *cloud* resulta más accesible que un sistema de licenciamiento tradicional.

#### a. El desafío de servicios dinámicos en BPM

Los avances en móviles y computación embebida han llevado a un ambiente computacional que considera personalización, movilidad y computación en tiempo real. Como resultado, SOA (*Service Oriented Architecture*) ha adoptado nuevos métodos de seguridad en el servicio, así como en la gestión y ejecución del mismo. Para los procesos de negocio estas adaptaciones representan desafíos en términos de modelado, desarrollo y ejecución. Para poder hacer que BPM se ejecute en estos nuevos ambientes debemos ver cuáles son los cambios requeridos para soportar servicios crecientes, dinámicos y descentralizados.

La investigación sobre arquitecturas emergentes de *web services* está colocando a BPM como un componente clave en el desarrollo de la siguiente generación de aplicaciones web. El uso de motores y modelos de procesos de negocio permite a los diseñadores de aplicaciones y analistas de negocio tener un foco clave en la funcionalidad del sistema sin la necesidad de enfocarse en arquitecturas técnicas más amplias. El uso de *web services* para crear procesos de negocio débilmente acoplados recae sobre la

orquestración para alcanzar la funcionalidad de la aplicación distribuida deseada. Típicamente esto se realiza mediante la combinación de servicios estables en un *workflow* estático. Sin embargo en los ambientes dinámicos los servicios son frecuentemente no estáticos y cambian de estado y características. O bien el ambiente puede ser inestable y como resultado el *workflow* debe adaptarse dinámicamente a cambios en las condiciones de ejecución.

Los ambientes de *web services* emergentes están adoptando móviles en forma creciente, así como datos personales y funciones de bajo nivel embebidas. Este tipo de servicios empujan los límites actuales de BPM. Las necesidades emergentes de estos ambientes computacionales generan demandas complejas en los procesos de negocio. En una aplicación típica de procesos de negocio el cambio puede ser manejado mediante el agregado de *loops* al proceso, o cambios entre *templates*. En ambientes más dinámicos los servicios se tienen que renegociar y los *workflows* se alteran durante la ejecución [1] [3] [4] [5].

Cambiar el modelo de proceso durante la ejecución (por ejemplo, insertando un *loop* o un *fork*) es un aspecto de investigación frecuente (sin mayores implicancias en los sistemas actuales de BPM). Mientras tanto, la adaptación de procesos de negocio dependiendo de los cambios en el ambiente y los recursos sigue siendo un tópico no explorado lo suficiente.

#### b. Servicios móviles

En los años recientes la adopción masiva de teléfonos móviles ha guiado al desarrollo de una nueva generación de aplicaciones web. Para las organizaciones estas aplicaciones representan una nueva posibilidad de mejorar los procesos de negocio.

Como vemos en la figura 1, se pueden observar desafíos al diseñador de procesos de negocio cuando el *workflow* está presente en un dispositivo móvil o ha hecho decisiones críticas basadas en servicios móviles en un ambiente emergente, como el *cloud*.

Aquí los *workflows* están presentes para manejar la respuesta a una emergencia desde una perspectiva centralizada. Sin embargo el cambio encarado en el *workflow* central es empeorado por factores locales. Por ejemplo el *workflow* puede tener que reaccionar a información contextual conectada a la localización de ambulancias y doctores en un sitio de emergencia. Aquí la localización de los dispositivos está constantemente cambiando y el contexto en que están presentes puede diferir.



Dicha no previsibilidad no puede ser modelada en un *workflow* estático. Aquí BPM debe adaptarse al estado de un número de servicios relacionados al contexto en que están presentes. Esto debe ser hecho en relación a la ejecución del *workflow* de respuesta más amplia [2] [6] [7] [8].

### c. Tecnología adaptativa de workflow

Las plataformas que al momento presentan características de *workflow* adaptativo presentan en común las siguientes características:

- Los lenguajes de modelado consideran la adaptabilidad mediante mecanismos de suspensión utilizando *breakpoints*, los cuales previenen al motor de *workflow* de correr tareas innecesarias que están a punto de ser modificadas. La notación gráfica suele estar basada en los modelos de actividad de UML. Sin embargo, esta notación parece ser demasiado técnica para los usuarios a los que la plataforma está destinada. En el caso particular de Cake [1], se ha rediseñado la misma mediante símbolos novel para la misma funcionalidad.
- La adaptabilidad de los *workflows* deriva en los requerimientos para la ejecución de funcionalidad que no puede ser alcanzada mediante servicios tradicionales. Se espera que las herramientas puedan lidiar con la adaptación de instancias de *workflows* en ejecución. Este componente puede mantenerse inalterado para la versión *cloud*.
- La adaptación de un *workflow* se puede realizar de una manera basada en casos, lo cual permite la posibilidad de reutilizar la experiencia en adaptación. Se recolecta la experiencia de episodios de adaptación previos en un caso base dedicado. Se poseerá entonces: un almacén de casos de adaptación, el propósito de la adaptación ejecutada, el *workflow* original, el *workflow* adaptado, así como la diferencia de ambos *workflows*. Cuando se presenta un nuevo requerimiento de adaptación (como parte de una actividad de modelado o como consecuencia de un evento que se encuentra en ejecución), esta es ejecutada en forma basada en casos mediante el reuso automático de un caso retornado desde su caso base.

- Las interfaces de usuario deben ser rediseñadas para satisfacer los requerimientos de los usuarios del *cloud* [2] [3] [9] [10] [11].

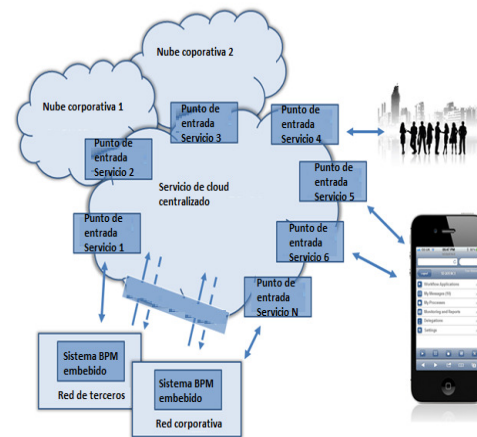


Figura 1: Esquema de *cloud* BPM conectado con móviles y sistemas BPM embebidos [2]

### d. Concepto de modelo de usuario

El concepto de modelo de usuario es un medio para mantener simplicidad y confiabilidad, los cuales son requerimientos esenciales en un escenario de *cloud*. Esto puede ser soportado mediante un mecanismo de gestión sofisticado de derechos de acceso para grupos y usuarios. La idea básica es que cada recurso en la plataforma de *cloud* (el *workflow*, sus tareas y recursos) posee un dueño al que se le permite manejar los derechos de acceso para el mismo.

Los usuarios pueden otorgar acceso abierto para el público a alguno de los recursos, en cuyo caso la autenticación de los usuarios se ejecuta mediante un *login* [3] [4] [12] [13].

### e. Conceptos de implementación

En la actualidad nos encontramos realizando pruebas de implementación sobre un BPMS desarrollado como parte de un proyecto interno. El mismo ha sido desarrollado desde su inicio con un enfoque mixto, tanto embebido como adaptado para el *cloud*. En cuanto al campo de investigación que nos ocupa actualmente, se han planteado las posibilidades expuestas en la Tabla 1.

Aspecto	Desarrollo
Dinamismo en la asignación de servicios	En un sistema embebido los servicios en general son asignados a las actividades en tiempo de diseño, con lo cual es previsible la estabilidad en el mismo. En un esquema de <i>cloud</i> encontramos una gran diversidad de locaciones accediendo a puntos disímiles del servicio de BPM. Incluso la mayoría de los servidores de <i>clouding</i> en la actualidad permiten implementar un esquema de <i>mirrors</i> (espejado) de servicios. Esta última característica crece por el uso de dispositivos móviles y los puntos de acceso. Para ello se hace necesario dentro de los nodos de ejecución del BPMS la existencia de tablas de ruteo que permitan, de acuerdo a ciertos criterios (como la localización del cliente, tiempo promedio de respuesta, velocidad de conexión, entre otros) decidir qué versión ejecutable del servicio asignar para mejorar la performance. Se dinamiza la ejecución del proceso y se necesita mayor robustez en la gestión de errores.
<i>Workflows</i> variables	Al igual que en el sistema Cake [1], se indica mediante <i>breakpoints</i> puntos del proceso en que la ejecución pudiera volverse no del todo predecible, o al menos variable. Esto permite optimizar la ejecución del proceso ya que el gestor del mismo no ejecuta innecesariamente actividades que podrían no repetirse a lo largo de su ejecución.
Dispositivos móviles	Se incrementa el dinamismo y la adaptabilidad del BPMS, creciendo el uso del mismo desde plataformas disímiles. Se observa aún la falta de estándares de seguridad realmente confiables que permitan el intercambio de datos sensibles entre los clientes y el BPMS. Esto se acentúa en un entorno de <i>cloud</i> , donde más aún los servidores de BPM pueden estar en un esquema de <i>mirrors</i> y sincronizados en forma distribuida

Tabla 1 Aspectos de implementación en el *cloud* [1] [14] [15].

El primero de los aspectos se encuentra en actual estudio y desarrollo. El interés de incorporar tablas de *ruteo* de servicios al BPMS permite incrementar la portabilidad y escalabilidad del mismo, ya que esto posibilita por un lado que los servicios web que se pueden invocar desde el mismo puedan estar diseminados por todo Internet. Además se favorece a la arquitectura distribuida del mismo al permitir *clustering* de servidores BPM ubicados incluso en nubes distintas.

El segundo de los aspectos es algo más complicado de implementar ya que requiere de modificaciones en el estándar BPMN, así como la consideración de un almacén de casuística que permita realizar las conversiones en tiempo de ejecución. La idea subyacente es poder indicar mediante elementos del diagrama la posibilidad de ejecutar caminos de ejecución variable dentro del proceso. Existe una versión implementada de carácter académico en la actualidad, aunque no se poseen en este aspecto versiones comerciales en el mercado [2] [16].

En cuanto al tercer punto referido a los estándares de seguridad en el uso de dispositivos móviles, se encuentran en el mercado actualmente varios BPMS que consideran sistemas de encriptación y firma de paquetes para su envío en un entorno de red inalámbrica, así como el establecimiento de conexiones seguras que permitan el envío de información sensible a través de las distintas actividades del proceso. Si bien estos BPMS lo hacen para sus versiones embebidas, aún es

necesario investigar con mayor profundidad las implicancias que conlleva aplicar este mismo concepto en un ambiente de *cloud* [3] [17] [18].

## LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

En la actualidad las líneas de investigación que nos ocupan se centran en la resignificación de conceptos de ejecución de los BPMS embebidos hacia un ambiente de *cloud*: acerca de cómo cobra más sentido en este nuevo contexto la consideración de *workflows* dinámicos, así como la asignación dinámica de servicios en la ejecución de flujos de proceso. Ambas características deben poder expresarse en el diagrama de proceso, con lo cual se hace necesario considerar modificaciones a la notación existente. Por otro lado toma gran significación el uso intensivo de dispositivos móviles, así como la consideración de estándares que garanticen la seguridad en su utilización dentro de un ambiente distribuido.

## FORMACIÓN DE RECURSOS HUMANOS

BPM ha cobrado importancia dentro del área de tecnología informática, la cual en los últimos años ha evolucionado desde el concepto producto hacia el paradigma de soluciones y servicios. El presente artículo ilustra una línea de trabajo iniciada en el año 2008, donde no sólo se cubre el desarrollo de procesos BPM sino también la mejora continua de los mismos

y la ampliación de sus ambientes de ejecución. En dicha línea se están formando alumnos para desarrollar su tesina e interactuar con docentes e investigadores formados, incorporando BPM y sus herramientas de soporte como línea de acción para la solución de problemas reales.

## REFERENCIAS

- [1] "The Challenge of Dynamic Services in Business Process Management", T. Kirkham, S. Winfield, T. Haberecht, J. Müller, G. De Angelis, University of Nottingham, United Kingdom, Springer, 2011
- [2] "Adaptive Workflow Management in the Cloud – Towards a Novel Platform as a Service", M. Minor, R. Bergmann, S. Görg, Business Information Systems II, University of Trier, Germany, 2012
- [3] "Innovative Approach for Agile BPM", M Mevius, R. Stephan, P. Wiedmann, eKNOW 2013: The Fifth International Conference on Information, Process, and Knowledge Management, 2013.
- [4] "Integration of Business Process Management and Complex Event Processing", Dr. Manuel Goetz, Germany, 2012.
- [5] "About the flexible Migration of Workflow Tasks to Clouds - Combining on and off premise Executions of Applications", M. Gerhards, V. Sander, A. Belloum, CLOUD COMPUTING 2012: The Third International Conference on Cloud Computing, GRIDs, and Virtualization, 2012
- [6] "Collaborative Business Process Modeling Approaches: A Review". S. Aleem, S. Molnar, and N. Mohamed, In Proc. of the 2012 IEEE 21st International workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises, pp. 274-279, June 2012.
- [7] "Simplifying Discovered Process Models in a Controlled Manner" Dirk Fahland, Wil M.P. van der Aalst Eindhoven University of Technology, The Netherlands, 2012.
- [8] "New methods for clinical pathways - Business Process Modeling Notation (BPMN) and Tangible Business Process Modeling (t.BPM)". Hubert Scheuerlein, Falk Rauchfuss, Yves Dittmar, Rüdiger Molle, Torsten Lehmann, Nicole Pienkos, Utz Settmacher. Springer-Verlag 2012
- [9] "Collaborative methods for Business Process Discovery". Marielba Zacarias, Paula Ventura Martins, Portugal, Springer-Verlag 2012
- [10] "Agile BPM in the age of Cloud technologies". JIRI KOLAR AND TOMAS PITNER Scalable Computing: Practice and Experience, 2012
- [11] "Information Flow Security for Business Process Models - just one click away". Andreas Lehmann and Dirk Fahland, University of Rostock, Germany, 2012
- [12] "On the Exploitation of Process Mining for Security Audits: The Process Discovery Case". Rafael Accorsi, Thomas Stocker, Günter Müller. Department of Telematics University of Freiburg, Germany. 2012
- [13] "BPM and iBPMS in the Cloud". Aleš Frece, Gregor Srdić, Matjaž B. Jurič, Proceedings of the 1st International Conference on Cloud Assisted ServiceS Bled, 25 October 2012
- [14] "Business Process Management in the cloud: Business Process as a Service (BPaaS)". Dr. Luis Ferreira Pires, April 1, 2012.
- [15] "A Framework for Reflective Business Process Management". S Balzert, P Fettke, P Loos, 2012 45th Hawaii International Conference on System Sciences.
- [16] "BPMN and Design Patterns for Engineering Social BPM Solutions". Marco Brambilla, Piero Fraternali, and Carmen Vaca, Politecnico di Milano, Piazza L. da Vinci 32, Milano, Italy, 2012
- [17] "Combining Social Web and BPM for Improving Enterprise Performances: the BPM4People Approach to Social BPM". Marco Brambilla, Piero Fraternali, Carmen Vaca, Stefano Butti, WWW 2012 – European Projects Track, April 16–20, 2012, Lyon, France
- [18] "Enhancement of traditional Business Process Management with reflection – a new perspective for Organizational Learning". S. Balzert, P. Fettke, P. Loos, Institute for Information Systems (IWi) at German Research Center for Artificial Intelligence (DFKI), Germany, 2012

# Validación de Requerimientos a través de Modelos Conceptuales - Modelos y Transformaciones

Marcelo Marciszack, Ramiro Pérez, Claudia Castro

Dpto. Ingeniería en Sist. de Información/ Facultad Regional Córdoba/ Universidad Tecnológica Nacional

{ marciszack, ramipez, ingclaudiacaastro }@gmail.com

## Resumen

El trabajo presentado en este artículo tiene como objetivo implementar una herramienta que permita gestionar y validar requerimientos de software, que ayudará a definir los límites del sistema al momento de formular los requerimientos, controlar y optimizar los procesos, y proveerá al grupo de desarrollo una base para la estimación del tiempo y costo del desarrollo de sistemas de software, permitiendo así conocer el estado del proyecto y el impacto de los cambios en caso de ser requeridos.

*Palabras clave: validación, requerimientos, casos de uso, modelos, conceptual, especificación, trazabilidad, software.*

## Contexto

El presente proyecto se encuentra consolidado dentro de la línea de investigación de Sistemas de Información en el Dpto. de Ingeniería en Sistemas de Información de la Universidad Tecnológica Nacional, Facultad Regional Córdoba.

El mismo busca dar solución a uno de los principales problemas de la ingeniería en sistemas relacionado a la elicitación y especificación de requerimientos, que vincula las distintas etapas del proceso de desarrollo de software manteniendo la trazabilidad de estos requerimientos hasta la validación e implementación de los mismos.

## Objetivos Generales

Los objetivos Generales del proyecto de investigación son los siguientes:

- Establecer marco teórico metodológico de técnicas para verificar y validar especificaciones de requerimientos de software / casos de uso.
- Construir e implementar una herramienta de software para la gestión de requerimientos/casos de uso, haciendo especial énfasis en la validación de los mismos.
- Definir propuesta metodológica para la especificación de casos de uso.

## Objetivos Específicos

- Validación de modelos conceptuales en la especificación de Requerimientos.
- Posibilitar en todo momento la trazabilidad de requerimientos desde su captura, modelado y seguimiento.
- Validación de Requerimientos Funcionales.

## Introducción

La actividad de análisis, diseño y construcción de sistemas de información involucra básicamente a tres tipos de actores: los desarrolladores, que codifican los programas en un lenguaje de programación determinado, los analistas, que especifican la funcionalidad que debe tener el sistema resultante y los usuarios, que poseen requerimientos acerca de lo que debería hacer el sistema para satisfacer sus



necesidades de información para la toma de decisiones.

El análisis de requisitos es la fase más importante en el desarrollo de un proyecto software, ya que de un correcto análisis dependerá la correcta implementación de la aplicación [1].

El documento de especificación de requisitos de software supone una especie de contrato entre usuario y desarrolladores en el que unos indican sus necesidades, mientras que los otros se limitan a implementar lo que se indica en el documento.

La tarea del análisis de requisitos es un proceso de descubrimiento, refinamiento, modelado y especificación y, por tanto, el desarrollador y el cliente tienen un papel activo en la obtención de estas necesidades [2]. Las últimas tecnologías utilizadas para la obtención de requisitos permiten una mejor comprensión de los documentos de especificaciones, que hasta ahora eran demasiado técnicos para la correcta comprensión por parte del usuario.

Estas técnicas modernas son los casos de uso, que forman parte del UML (Lenguaje Unificado de Modelado) [3]. Ésta es la principal herramienta utilizada para el diseño completo de proyectos software orientado a objetos. Los casos de uso modelan el sistema desde el punto de vista del usuario, permitiéndole así la comprensión completa del futuro sistema.

Finalmente, se debe indicar que esta fase es posiblemente la más costosa (temporalmente) en el desarrollo de un producto software. Esto se debe a que, en general, el cliente no sabe realmente lo que quiere y requiere la ayuda de los analistas para concretar las funciones que realmente se han de implementar. Por tanto, de la calidad del documento de ERS (Especificación de Requerimientos de Software) dependerá el desarrollo y calidad del producto final. La existencia de un estándar, como es el presentado en este proyecto, para la ERS [4] permite la coherencia en la especificación de requisitos y ayuda a no dejar cabos sueltos.

Los principales objetivos que se identifican en la especificación de requisitos software son:

1. Ayudar a los clientes a describir claramente lo que se desea obtener mediante un determinado software: El cliente debe participar activamente en la especificación de requisitos, ya que éste tiene una visión mucho más detallada de los procesos que se llevan a cabo.

2. Ayudar a los desarrolladores a entender qué quiere exactamente el cliente: En muchas ocasiones el cliente no sabe exactamente qué es lo que quiere. La ERS permite al cliente definir todos los requisitos que desea y al mismo tiempo los desarrolladores tienen una base fija en la que trabajar. Si no se realiza una buena especificación de requisitos, los costes de desarrollo pueden incrementarse considerablemente, ya que se deben hacer cambios durante la creación de la aplicación.

3. Servir de base para desarrollos de estándares de ERS particulares para cada organización: Cada entidad puede desarrollar sus propios estándares para definir sus necesidades. Una buena especificación de requisitos software ofrece una serie de ventajas entre las que destacan el contrato entre cliente y desarrolladores (como ya se ha indicado con anterioridad), la reducción del esfuerzo en el desarrollo, una buena base para la estimación de costes y planificación, un punto de referencia para procesos de verificación y validación, y una base para la identificación de posibles mejoras en los procesos analizados.

Las características deseables para una buena especificación de requisitos software que se indican en el IEEE son las siguientes:

- Correcta
- No ambigua
- Completa
- Verificable
- Consistente
- Clasificada
- Modificable
- Explorable

- Utilizable durante las tareas de mantenimiento y uso

Durante el proceso de validación de requerimientos, se deben llevar a cabo verificaciones sobre requerimientos en el documento de requerimientos [5].

Estas verificaciones comprenden:

1. Verificaciones de validez.
2. Verificaciones de consistencia.
3. Verificaciones de integridad.
4. Verificaciones de realismo.
5. Verificabilidad.

Pueden utilizarse varias técnicas de validación de requerimientos, sin embargo nuestro trabajo se centrará en la construcción de modelos conceptuales basados en los requerimientos para poder realizar su validación [6], [7], [8].

Entre algunas de las técnicas de validación de requerimientos existentes se pueden mencionar las siguientes:

1. Revisiones de requerimientos.
2. Construcción de prototipos
3. Generación de casos de prueba.
4. Análisis de consistencia automático

La validación de requerimientos es importante debido a que los errores en el documento de requerimientos pueden conducir a importantes costos al repetir el trabajo cuando son descubiertos durante el desarrollo o después de que el sistema esté en uso. El costo de arreglar un problema en los requerimientos haciendo un cambio en el sistema es mucho mayor que reparar los errores de diseño o los de codificación.

Principios generales a utilizar en la validación:

1. Especificación de los requisitos.

Una especificación documentada de los requisitos del software proporciona la base para la validación y la verificación. El proceso de validación del software no puede completarse sin un establecimiento de las especificaciones de los requisitos.

2. Prevención de defectos.

Las necesidades del aseguramiento de la calidad del software fijan su atención en la prevención de defectos en el proceso de desarrollo del software y no en probar la calidad del código del software después que se escribe. La prueba del software está muy limitada en su capacidad de detectar todos los defectos latentes en el código del software. La complejidad de la mayoría del software impide que sea probado exhaustivamente. La prueba del software es una actividad necesaria.

3. Tiempo y esfuerzo.

La validación del software requiere tiempo y esfuerzo. La preparación para la validación debe comenzar con anticipación; es decir, durante la planificación del diseño y desarrollo y el diseño de la entrada de datos. La conclusión final que muestra que el software se encuentra validado debe estar basada en la evidencia recolectada a partir de los esfuerzos planificados dirigidos a lo largo del ciclo de vida del software.

4. Ciclo de vida del software.

La validación del software tiene lugar dentro del ambiente del ciclo de vida establecido del software. El ciclo de vida del software contiene las tareas de ingeniería de software y la documentación necesaria para soportar la validación del software. Además, el ciclo de vida del software contiene las tareas específicas de verificación y validación que son apropiadas para el uso previsto del software. En la presente nota técnica no recomendamos un modelo particular de ciclo de vida (modelo lineal secuencial, modelo de construcción de prototipos, modelo de desarrollo rápido de aplicaciones, entre otros), sólo establecemos que se deben seleccionar los modelos más apropiados a utilizar en el proyecto de desarrollo del software. Varios modelos del ciclo de vida del software son definidos en la ingeniería del software.

El ciclo de vida puede ser seguido completamente o tener variaciones en su desarrollo, debido a las propias características y naturaleza del software que

se desea desarrollar y su dominio de implementación.

#### 5. Planificación.

El proceso de validación del software se define y controla a través de un plan. El plan de validación del software define lo que será logrado a través del proceso de validación del software.

#### 6. Procedimientos.

El proceso de validación del software se realiza a través del uso de procedimientos documentados. Estos procedimientos establecen "cómo", "quién" y "cuándo" se llevará a cabo la validación del software. Los procedimientos deben identificar las acciones específicas o la sucesión de acciones que deben tomarse para completar las actividades individuales de validación.

#### 7. Validación del software después de un cambio.

Debido a la complejidad del software, un cambio aparentemente pequeño puede tener un impacto significativo en todo el sistema. Cuando se hace cualquier cambio al software (incluyendo pequeños cambios), el estado de validación del software necesita ser restablecido. Siempre que el software sea cambiado, un análisis de validación debe dirigirse no solamente para la validación del cambio individual, sino también para determinar la magnitud e impacto de ese cambio en la totalidad del software.

#### 8. Alcance de la validación.

El alcance de la validación debe estar basado en la complejidad del software y en los riesgos de seguridad; no en el tamaño de la organización o la restricción de los recursos. La selección de las actividades y tareas a llevar a cabo durante la validación deben corresponderse con la complejidad del diseño del software y el riesgo asociado con la utilización del software para el uso previsto. La documentación de la validación debe ser suficiente para demostrar que todos los planes y procedimientos de validación del software se han completado con éxito.

#### 9. Independencia de la validación.

Las actividades de validación deben ser conducidas cumpliendo el precepto básico de gestión de la calidad sobre la "independencia de la revisión". La auto-validación es sumamente difícil.

Cuando sea posible, siempre es mejor llevar a cabo el proceso de validación de manera independiente, sobre todo para las aplicaciones que poseen un alto riesgo. Algunas organizaciones optan por la contratación de una tercera parte independiente, pero esta solución no siempre es factible. Otro enfoque es asignar la validación a miembros del personal de la propia organización que no están involucrados en el desarrollo del software o en su aplicación, pero que tienen suficiente conocimiento para evaluar el proyecto y conducir el proceso de validación. En estos casos las organizaciones más pequeñas necesitan ser creativas en la disposición y asignación de las tareas para mantener en todo momento la independencia.

#### 10. Flexibilidad y responsabilidad.

La aplicación específica de estos principios de validación del software puede ser muy diferente de una aplicación a otra. El fabricante o desarrollador del software tiene flexibilidad a la hora de escoger como aplicar estos principios de validación, pero es el responsable de demostrar que el software se ha validado. Las etapas del proceso de validación descritas aquí pueden ser simplificadas dependiendo de la naturaleza del software y de su uso previsto.

## **Líneas de investigación y desarrollo**

Para el modelado de requerimientos se selecciono la Notación para el Modelado de Procesos de Negocio (BPMN por sus siglas en ingles Business Process Modeling Notation) [9] es una notación grafica estandarizada que permite el modelado de procesos de negocio, en un formato de flujo de trabajo (workflow).

Haciendo uso de la capacidad de BPMN, de modelar procesos como flujos, extendemos su uso para describir los sub

procesos que el sistema debe realizar, podemos aprovechar las ventajas que nos brinda esta herramienta.

Se plantearon 2 niveles de modelado, el primero a nivel de procesos de negocio, como lo plantea originalmente BPMN. Y luego un nuevo nivel, a nivel de procesos de sistema de información, usando la misma notación, y con relación a los procesos de negocio que da soporte. Una capa adicional podría añadirse, a nivel de detalle de implementación, pero ya cae fuera de los alcances del proyecto de investigación.

Para alcanzar mapeo y trazabilidad entre la capa de negocio y la de sistema de información, evaluamos distintas herramientas, como QVT, MOF, y XSLT. Finalmente usamos transformaciones XSLT, ya que el formato estándar de BPMN es XPDL, el cual está basado en XML; asimismo no encontramos disponibles herramientas para QVT usando BPMN. Otro inconveniente que se presentó fue la curva de aprendizaje para usar QVT o MOF, a diferencia de XSLT que es bastante sencillo de aprender y usar.

De esta manera podemos tener requerimientos del sistema de información modelados de manera de workflows, que dan soporte a los procesos de negocio.

Aplicando transformaciones XSLT a los archivos XPDL podemos disponer de diagramas de casos de uso, de actividad y de clases, en formato XML.

Al mismo tiempo, se hizo un trabajo anexo usando casos de uso, principalmente usados en UML, dada la amplia aceptación como medio para describir las interacciones realizadas entre el sistema y su entorno. [10]

El Lenguaje Unificado de Modelado (UML) define un caso de uso como "la especificación de una secuencia de acciones, incluyendo variantes, que un sistema (o un subsistema) pueden llevar a cabo, en interacción con los actores del sistema" [11]

Actualmente UML es un estándar altamente adoptado que define los

conceptos centrales para el modelado de casos de uso. Al aplicar este estándar, los casos de uso se identifican y estructuran mediante los diagramas de casos de uso de UML y luego se especifican con descripciones textuales. [10]

Usando casos de uso, se encontró que no están expresamente especificados los atributos de un caso del uso. Además, una de las cosas más sorprendentes sobre el estándar UML es la falta de detalle sobre la estructura de los casos de uso. [12]

De esta manera, se observó la necesidad de contar con una herramienta que permita definir y administrar los casos de uso de un sistema software, considerando los atributos comunes y mínimos a distintos autores. Además, se investigó sobre la relación de los casos de uso, y las distintas versiones de un sistema software.

Distintas versiones del sistema, pueden tener más casos de usos agregados que la versión anterior, pueden tener casos de usos modificados o deprecados. Para ello es clave también contar con información de la relación de los casos de uso con cada una de las versiones del sistema, garantizando integridad de producto, trazabilidad y proporciona medios para agilizar la auditoría.

## Resultados y Objetivos

Para la línea de BPMN, fue posible realizar un modelo a nivel del sistema de información a partir del modelo de procesos de negocios, mediante transformaciones XSLT. También fue posible realizar transformaciones del modelo XPDL a archivos para simuladores de autómatas finitos.

En el corto plazo seguiremos modificando las transformaciones XSLT con el objetivo de usarlos con SMT Solver, redes de Petri o con autómatas finitos, para poder analizar el modelo, y su validez.

Así, para la primera versión de la herramienta de gestión de requerimientos usando UML, consideramos la administración de sistemas, sus versiones,



casos de uso de una versión del sistema, en conjunto con la administración de usuarios y el historial de modificaciones de cada componente por parte de los usuarios,

Las principales funcionalidades que provee la herramienta web de casos de uso se describen a continuación:

- Administración de los atributos de un sistema y versiones.
- Administración de los atributos de casos de uso, incluyendo precondiciones, pasos, alternativas, etc.
- Diseño del Modelo Conceptual.
- Gestión de cambios en los requerimientos.
- Clasificación los requerimientos.
- Priorización de los requerimientos.
- Trazabilidad de los requerimientos.
- Validación del Modelo Conceptual.
- Visualización de requerimientos.
- Generación de reportes y exportación de los modelos a distintos formatos como XML, PDF, XMI, etc.
- Gestión de la configuración de los requerimientos.

Cabe destacar que haremos especial énfasis en la validación de los requerimientos, pero también abarcaremos los demás aspectos de la gestión de los mismos para lograr una herramienta integral que permita hacer el seguimiento continuo de su evolución.

## Formación de Recursos Humanos

Este proyecto prevé la formación de recursos humanos que formen parte del mismo, y que al participar del proyecto contribuyan en su formación.

El mismo cuenta con dos becarios de investigación, que forman parte del presente proyecto, y que serán de ayuda en la recolección, manipulación y tabulación de la información de entrenamiento del sistema.

Actualmente participan 3 tesis de la maestría de Ingeniería en Sistemas de

Información y un doctorando de Ingeniería de Software que desarrollan sus trabajos de en el ámbito del proyecto, lo cual permitirá una contribución a su formación académica.

Al mismo tiempo se prevé vinculaciones con otras redes de investigación; al igual que la realización de un acuerdo de transferencia de resultados con algún organismo que desee considerar los resultados producidos por el sistema que se desarrollará.

## Referencias

- [1] I. Sommerville. Software Engineering, Computing Department, Lancaster University, John Wiley & Sons Ltd., 2005.
- [2] A. Davis Software requirements Object, functions and states; Prentice Hall international Inc. 1993.
- [3] J. Rumbaugh, I. Jacobson, G.Booch "The Unified Modelling Language Reference" Addison Wesley 1999.
- [4] ANSI/IEEE Standard 830-1984: Standard for software Requirements Specifications, The institute of Electrical and Electronic Engineers, New York, 1984.
- [5] M. Jackson "Software Requirements & Specifications". Addison Wesley.1995.
- [6] G. Kotonya, I. Sommerville. Requirements Engineering, Process and Techniques. John Wiley & sons. 1998.
- [7] Emilio Insfrán, Isabel Díaz y Burbano Margarita. Modelado de Requisitos para la Obtención de esquemas conceptuales. Disponible en: <http://www.dsic.upv.es/~einsfran/papers/39-ideas2002.pdf>
- [8] C. Leonardi, J.C.S. Leite, Gustavo Rossi. "Una estrategia de Modelado Conceptual de Objetos , basada en Modelos de requisitos en lenguaje natural ".Tesis de Maestría Universidad Nacional de la Plata.
- [9] J. D. Pérez. "Notaciones y lenguajes de procesos. Una vision global". Tesis de Doctorado Universidad de Sevilla.
- [10] V.E. Saldaño. "Descubrimiento de Servicios Geográficos a partir de casos de

uso textuales”. XV Congreso Argentino de Ciencias de la Computación.

[11] S. Somé. “A Meta-Model for Textual Use Case Description”. J. of Object Technology, 8(7).

[12] D. Coleman. “A Use Case Template: draft for discussion”. Fusion Newsletter.

## REDES BAYESIANAS APLICADAS A LA INGENIERÍA DE SOFTWARE

Carlos Said, Rosana Piergallini, Nelson Di Grazia, Sabrina Pompei.

Escuela de Tecnología, Universidad Nacional del Noroeste de la Provincia de Buenos Aires (UNNOBA), Buenos Aires, Argentina

[carlossaid@gmail.com](mailto:carlossaid@gmail.com), [mrosanapierg@yahoo.com.ar](mailto:mrosanapierg@yahoo.com.ar), [nelson.digrazia@telefonica.com](mailto:nelson.digrazia@telefonica.com),  
[sabrinapompei@unnoba.edu.ar](mailto:sabrinapompei@unnoba.edu.ar),

### Resumen

En ausencia de una medida acordada de la calidad del software, la densidad de defectos ha sido una medida comúnmente usada. En consecuencia ha habido numerosos intentos para construir modelos para la predicción del número de defectos residuales en el software. Las variables clave en dichos modelos son en general métricas de tamaño y complejidad o medidas derivadas del testing de la información y la codificación. Sin embargo estos enfoques presentan dificultades estadísticas y teóricas. El uso de redes de creencias bayesianas puede superar algunos de estos problemas, teniendo en cuenta los diversos factores implícitos en la prevención de defectos, detección y complejidad.

La línea de investigación propuesta se centra en el uso de redes de creencias bayesianas aplicado al modelo de densidad de defectos proporcionando un nuevo enfoque para los procesos de modelización y de ingeniería de artefactos de software. La naturaleza dinámica de este modelo proporcionará una manera de simular diferentes hechos e identificar cursos óptimos de acción basados en el conocimiento incierto.

**Palabras clave:** *bayes, ingeniería de software, calidad, métricas, costos, redes bayesianas de conocimiento.*

### Contexto

Esta línea de investigación fue aprobada por la Secretaría de Investigación, Desarrollo y Transferencia de la UNNOBA, en el marco de la convocatoria para la acreditación de proyectos y solicitud de subsidios de investigación bianuales (SIB) 2013.

El equipo está constituido por docentes e investigadores del departamento de Informática y Tecnología, quienes dictan clases en las asignaturas Métricas de Software, Ciencias de la Computación, Análisis y Diseño de Sistemas y Programación Orientada a Objetos pertenecientes a las carreras Licenciatura en Sistemas e Ingeniería en Informática, en la Escuela de Tecnología de la UNNOBA.

### Introducción

La forma más común de insertar defectos en un producto de software es mediante la aplicación de cambios al mismo y/o mediante la ausencia de aplicación de metodologías de testing de manera exhaustiva. En general, esto último ocurre ya que la llamada ‘crisis de producción de software’ explica que al llegar a las etapas de testing en el proceso de producción de software, los tiempos y presupuestos se han ‘agotado’. Otras

intenciones de explicar esta carencia de productos confiables se sustentan en la ausencia de metodologías aplicadas de control de producción orientadas a la ‘confiabilidad’ del producto. En cada versión de software, decenas o cientos de cambios se aplican al producto de software. Todo cambio tiene su característica única basada en el proceso y los factores del producto. Esta densidad de defectos ha sido una medida comúnmente usada en la calidad del software.

Cada cambio tiene su propia probabilidad de inyectar un defecto. En consecuencia han habido numerosos intentos para construir modelos para la predicción del número de defectos residuales en el software. Por lo general las variables clave en los modelos son o métricas de tamaño y complejidad o medidas derivadas del ‘testing’ de la información y la codificación. Hay sin embargo dificultades estadísticas y teóricas con estos enfoques.

El uso de redes de creencias bayesianas puede superar algunos de los problemas presentes, teniendo en cuenta los diversos factores implícitos en la prevención de defectos, detección y complejidad. A nivel general podemos ver cómo el uso de estas redes aplicadas al modelo de densidad de defectos proporciona un nuevo enfoque importante para los procesos de modelización y de ingeniería de artefactos de software.

La naturaleza dinámica de este modelo proporciona una manera de simular diferentes hechos e identificar cursos óptimos de acción basados en el conocimiento incierto. Estos beneficios se ven reforzados cuando examinamos cómo el modelo explica resultados conocidos, en particular el dilema ‘es más grande mejor?’.

Este nuevo enfoque muestra cómo podemos construir redes complejas de interconexión entre los procesos, productos y recursos de una manera hasta ahora poco viable. También podemos integrar la incertidumbre y los criterios subjetivos en el modelo sin sacrificar el rigor, e ilustrar cómo la toma de decisiones durante el proceso de desarrollo de software tiene influencia en la calidad alcanzada.

Los beneficios esperados de este enfoque son:

- mayor utilidad para la gestión de proyectos que el análisis de valores atípicos y los estadísticos clásicos;
- Incorporación de ideas actuales de la investigación y de la experiencia;
- uso para formación de los gerentes de proyectos de software ;
- permitir la comparación de diferentes decisiones por medio de simulación y los análisis ‘qué sucede si?’;
- Integración de una forma de predicción de costo y calidad.

Se busca desarrollar un modelo de diseño orientado por objetivo, lo que suele llamarse ‘goal oriented’, una descripción detallada del marco de medición del proceso que incluya los conceptos correspondientes de minería de datos. El enfoque se basa en los datos del proceso y los del producto. El modelo resultante es llamado el Modelo de Producción del Software.

Este Modelo de Producción del Software estará construido sobre la base de una versión ampliada del proceso de construcción de software modelo-V que a su vez se deriva del modelo de cascada. Los marcos de procesos y productos de medición combinados con el conocimiento de los expertos describen todas las fases del modelo-V, y de esta manera permite la



representación como un modelo coherente del proceso completo de ingeniería.

El Modelo de Producción del Software estará diseñado principalmente para optimizar la calidad del producto y la distribución del esfuerzo de desarrollo en el tiempo. Permitirá evaluar las medidas del proceso que influyen en la calidad del producto. Debido a las capacidades dinámicas del mismo permitirá además analizar el impacto de los posibles cambios en los procesos y productos, que ocurren en el tiempo.

Estas capacidades ofrecerán visiones diferentes a los administradores de proyectos vinculados a productos de software, para poner de relieve la utilidad práctica del modelo.

El Modelo de Producción del Software estará basado en una Red Bayesiana Dinámica, con capacidad para representar múltiples características de cambio de una versión de software específico. Permitirá a los administradores de proyectos supervisar el impacto de cada cambio realizado en esta versión del software tomando estas características de cambio individuales en cuenta.

Las Redes Bayesianas, elemento de base para cada Red Bayesiana Dinámica, son parte de las técnicas modernas de inteligencia artificial (AI). Aunque un teorema fundamental de la probabilidad condicional fue presentado por Thomas Bayes en el siglo XVIII, el término "Red Bayesiana" y sus conceptos fueron introducidos en la década de 1980 en el trabajo de Judea Pearl. Investigadores han presentado declaraciones referidas a las Redes Bayesianas, por ejemplo Darwiche en 2009: "Las Redes Bayesianas son tan importantes para la Inteligencia Artificial (AI) y el Aprendizaje Automático como los circuitos booleanos son a la informática", "

las Redes Bayesianas han revolucionado la AI".

Una Red Bayesiana Dinámica es un modelo de formación de una secuencia de Redes Bayesianas individuales donde cada Red individual refleja el estado del sistema en un punto de tiempo específico. Una Red Bayesiana puede ser percibida desde dos perspectivas: (1) como una gráfica donde los nodos representan variables aleatorias y los arcos representan las relaciones entre las variables, (2) como una definición de modelo matemático formal donde las variables se expresan como distribuciones de probabilidad condicional.

Las Redes Bayesianas se pueden construir mediante la aplicación de un algoritmo de aprendizaje para el conjunto de datos que permite el aprendizaje de una estructura. El mayor potencial de las mismas reside en la capacidad de combinar el conocimiento experto con los datos empíricos.

Otros beneficios de estas redes incluyen:

- capacidad de reflejar las relaciones causales,
- ejecutar el modelo con datos incompletos y supuestos,
- razonamiento hacia adelante y hacia atrás,
- modelización de la incertidumbre,
- combinación de datos cuantitativos y cualitativos.

#### **Líneas de investigación y desarrollo**

- Redes bayesianas en la ingeniería de software: redes bayesianas, toma de decisión frente a la incertidumbre, inferencias aplicando redes bayesianas, limitaciones de las redes bayesianas;

- Calidad del software: modelos de análisis de la calidad del software, modelos de predicción de calidad en el software, métricas de calidad, paradigmas para el diseño confiable.
- Herramientas: herramientas para la construcción de Redes Bayesianas de conocimiento (BBN), comparación y selección.

### Resultados y Objetivos

El objetivo general del presente proyecto es construir un modelo de estimación de defectos en la construcción de artefactos de software de mayor precisión que el causal clásico.

Objetivos específicos:

- Mejorar la precisión de estimación de defectos en la construcción de software respecto al modelo causal clásico.
- Definir un Modelo que responda ante características como datos incompletos, variables estocásticas e incertidumbre.

### Formación de Recursos Humanos

Integran el equipo de trabajo un docente investigador, quien dirige el proyecto, y tres docentes investigadores en formación.

El proyecto fue aprobado en Diciembre 2012 por lo que dentro de la línea presentada no se cuenta aún con tesis y/o tesinas presentadas.

### Bibliografía

[1] Angelis, L., & Stamelos, I. (2000). A Simulation Tool for Efficient Analogy Based Cost Estimation. *Empirical Software Engineering*, 5, 35–68.

doi:10.1023/A:1009897800559

[2] Baresi, L., Morasca, S., & Paolini, P. (2003). Estimating the design effort of Web applications, In *Proceedings Ninth International Software Measures Symposium*, September 3-5, (pp. 62-72).

[3] Bibi, S., Stamelos, L., & Angelis, L. (2003) Bayesian Belief Networks as a Software Productivity Estimation Tool. In *Proceedings 1st Balkan Conference in Informatics*, Thessaloniki.

[4] Briand, L. C., El-Emam, K., Surmann, D., Wiczorek, I., & Maxwell, K. D. (1999). An Assessment and Comparison of Common Cost Estimation Modeling Techniques. In *Proceedings of ICSE 1999*, Los Angeles, CA, (pp. 313-322).

[5] Costagliola, G., Di Martino, S., Ferrucci, F., Gravino, C., Tortora, G., & Vitiello, G. (2006). Effort estimation modeling techniques: a case study for web applications, In *Proceedings of the Intl. Conference on Web Engineering (ICWE'06)*, (pp. 9-16).

[6] Druzdzal, M. J., & van der Gaag, L. C. (2000). Building Probabilistic Networks: Where Do the Numbers Come From? *IEEE Transactions on Knowledge and Data Engineering*, 12(4), 481–486. doi:10.1109/TKDE.2000.868901

[7] Kuo, W. (2007) Compatibility and simplicity: the fundamentals of reliability, *IEEE Trans. Reliab.*, vol. 56.

[8] Raheja, D. G., and Allocco M. (2006) *Assurance Technologies Principles and Practices: A Product, Process, and System Safety Perspective*, 2nd ed., Wiley, Hoboken, NJ.

[9] Kan, S. H., *Metrics and Models in Software Quality Engineering*, Addison-

Wesley, Reading, MA, 1995, p. 192  
(Rayleigh model discussion).

[10] IEEE Guide for Selecting and Using  
Reliability Predictions, IEEE 1413.1–2002  
(based on IEEE 1413), IEEE, Piscataway, NJ.

[11] Analysis Techniques for System  
Reliability: Procedure for Failure Mode and  
Effects Analysis (FMEA), IEC 60812,  
International Electrotechnical Commission,  
Geneva, 2006.

## Aplicación de un Método para el estudio de la Calidad de los Procesos de Negocio en la comparación de Metodologías Ágiles de Desarrollo

C. Salgado, M. Peralta, M. Berón, D. Riesco, G. Montejano  
Departamento de Informática – Facultad de Ciencias Físico Matemáticas y Naturales  
Universidad Nacional de San Luis  
Ejército de los Andes 950, San Luis, San Luis, Argentina  
C.P. 5700 – Tel.: 54-026644520300 – Int. 2101  
[csalgado, mperalta, mberon, driesco, gmonte]@unsl.edu.ar

web: <http://www.sel.unsl.edu.ar>

### Resumen

La complejidad de los procesos de negocio (PN) ha promovido la investigación sobre distintos aspectos de estos procesos, como la utilidad [1], evaluación de la calidad [2] o la medición [3]. En este contexto, son frecuentes los estudios referentes a la utilización de herramientas y lenguajes para modelar PN [4, 5]. La motivación principal para investigar en esta área, es la variedad de notaciones y lenguajes de modelado, definición y ejecución de PN, y la necesidad de tener modelos de alta calidad.

El desarrollo de modelos conceptuales es una tarea clave en las primeras etapas del ciclo de vida de los PN. Por ello, es fundamental que dichos modelos sean de calidad en cuanto a su entendibilidad y mantenibilidad.

En función de esto, hemos definido un método para evaluar la calidad de modelos de procesos de negocio. Así, como una validación práctica de dicho método, se aplicó en el análisis de Modelos de PN a los metamodelos de metodologías ágiles de desarrollo, partiendo del concepto de que dichas metodologías pueden verse como PN debido a las características que poseen.

**Palabras clave:** Procesos de Negocio, Modelos Conceptuales de Procesos de Negocio, BPMN.

### Contexto

El presente trabajo se enmarca en el Proyecto de Investigación: Ingeniería de Software: Aspectos de alta sensibilidad en el ejercicio de la profesión de Ingeniero de Software – Facultad de Ciencias Físico-Matemáticas y Naturales, Universidad Nacional de San Luis. Proyecto N° 22/F222. Dicho proyecto es la continuación de diferentes proyectos de investigación a través de los cuales se ha logrado un importante vínculo con distintas universidades a nivel nacional e internacional. Además, se encuentra reconocido por el programa de Incentivos.

### Introducción

Al hablar de calidad en el modelado conceptual, se debe distinguir entre la calidad del producto (relacionada con las características del modelo conceptual) y la calidad del proceso de modelado (relacionada a cómo se desarrollan los modelos) [6]. Al respecto, la complejidad de un modelo conceptual puede estar altamente influenciada por los diferentes elementos que lo componen, tales como tareas, subprocessos, participantes, eventos, etc. Por lo tanto, no es aconsejable definir una medida general para su complejidad [7].

Si bien existen muchas definiciones de calidad en los distintos campos de investigación, no se ha encontrado una definición consensuada respecto a qué es la



calidad de los modelos conceptuales. Al respecto Moody, en [8], propone que la calidad de los modelos conceptuales podría definirse en base a la definición de calidad de ISO 9000 [9]. Así, Moody dice que la calidad de los modelos conceptuales se podría definir como: *La totalidad de los rasgos y características de un modelo conceptual que influyen en su habilidad de satisfacer las necesidades implícitas o declaradas.*

Es fundamental que toda propuesta para la evaluación adhiera a estándares aceptados y aplicados prácticamente [8]. En particular, Moody propone que deberían ser consistentes con las normas de calidad ISO 9000 [9], e ISO/IEC 9126 [10], ya que un modelo conceptual es un tipo particular de producto (ISO 9000) y, dentro de ISO/IEC 9126 los modelos conceptuales existen como modelos de sistemas de información.

Como se mencionó previamente, la complejidad de un modelo conceptual puede estar altamente influenciada por los diferentes elementos que lo componen. Por lo tanto, no es aconsejable definir una medida general para su complejidad [7]. Así, Rolón en [11] propone un conjunto de medidas para la calidad de modelos conceptuales de procesos de negocio desarrollados en BPMN (**B**usiness **P**rocess **M**odeling **N**otation). Estas medidas se basan en la propuesta de García Rubio [12] de medidas para la calidad del proceso del software.

Es importante destacar que, en la literatura, no se encontraron trabajos que propongan la definición de un método para la evaluación de modelos conceptuales de procesos de negocio. Por ello, se definió un método para evaluar la calidad de estos modelos [13]. Dicha evaluación se realiza en función de su mantenibilidad. En especial, se hace énfasis en el análisis de su entendibilidad y de su adaptabilidad, independientemente de su representación.

La motivación del método surgió de la necesidad de las organizaciones de tener un medio que permita representar sus procesos de negocio de una manera eficiente y que,

además, posibilite comunicarse e interactuar con otros procesos, ya sea de la misma organización o de organizaciones externas a ella. En función de ello, el objetivo del método es proveer a los diseñadores, analistas y/o desarrolladores (que intervienen en las distintas fases de la definición y el modelado de los procesos de negocio de una organización) un medio que ayude a obtener modelos de calidad del proceso analizado o tratado. Para lograr su objetivo, la primera etapa del método propuesto consiste en la determinación, agrupamiento y análisis de las mejores características, sub-características y componentes más relevantes y frecuentes que deberían satisfacer los modelos conceptuales de los PN. De esta manera, se propone plasmar sobre una estructura dichas características y componentes. Esta estructura permitirá, en las siguientes etapas del método, estudiar el grado en que los modelos satisfacen las características mencionadas. Dichas características, se toman de distintos estándares (como por ejemplo de BPDM), la experiencia de expertos en el modelado de Procesos de Negocio y el estudio/análisis de modelos de algunos casos de estudio particulares.

Acorde a la propuesta de Moody en [8], se utilizan conceptos de estándares como ISO 9000, ISO/IEC 9126, ISO/IEC 14598 [14], en su definición. Ello permite el uso de un vocabulario conocido y provee una forma para realizar la evaluación y organizar los resultados. Principalmente, el método se basa en el estudio de la mantenibilidad de los modelos conceptuales de procesos de negocio. Para ello, se hace hincapié en las características de entendibilidad y adaptabilidad de la norma ISO/IEC 9126 para analizar la mantenibilidad de los modelos evaluados.

Además, el método servirá a los encargados del proceso como apoyo en la toma de decisiones. Por lo que es de mayor utilidad en las primeras fases del modelado de los PN. Esto reducirá los costos que implica detectar y solucionar fallas o errores en etapas posteriores.

En todo proceso de evaluación, los resultados de dicho proceso deben ser documentados. De esta manera, la documentación que se obtiene debe ser organizada y almacenada para poder aspirar a una mejora continua en la organización. Por este motivo, en el método se propondrá una manera de documentar los resultados obtenidos en la evaluación. Esto servirá, además, como punto de partida para analizar la evolución de los modelos en futuras evaluaciones. Esto es fundamental, ya que los evaluadores pueden ser personas de la misma empresa, evaluadores internos, o personas externas a la organización que actuarán como evaluadores externos.

Desde otro punto de vista, el desarrollo de software puede ser visto como un PN. Prueba de ello es que existen numerosas propuestas metodológicas que inciden en distintas dimensiones del proceso de desarrollo. Por una parte, existen propuestas más tradicionales que se centran especialmente en el control del proceso, estableciendo rigurosamente las actividades involucradas, los artefactos a producir, y las herramientas y notaciones a utilizar. Estas propuestas demostraron ser efectivas y necesarias en un gran número de proyectos, pero han presentado problemas en otros. Una posible mejora es incluir en los procesos de desarrollo más actividades, más artefactos y más restricciones, basándose en los puntos débiles detectados. Sin embargo, el resultado final sería un proceso de desarrollo más complejo que puede incluso limitar la propia habilidad del equipo para llevar a cabo el proyecto. Otra alternativa es centrarse en otras dimensiones, como por ejemplo el factor humano o el producto software. Esta es la filosofía de las metodologías ágiles, las cuales dan mayor valor al individuo, a la colaboración con el cliente y al desarrollo incremental del software con iteraciones muy cortas.

Desde esta perspectiva, no existe una metodología universal para hacer frente con éxito a cualquier proyecto de desarrollo de software. Toda metodología debe ser adaptada al contexto del proyecto (recursos

técnicos y humanos, tiempo de desarrollo, tipo de sistema, etc.). Históricamente, las metodologías tradicionales han intentado abordar la mayor cantidad de situaciones de contexto del proyecto, exigiendo un esfuerzo considerable para ser adaptadas, sobre todo en proyectos pequeños y con requisitos muy cambiantes. Las metodologías ágiles ofrecen una solución casi a medida para una gran cantidad de proyectos que tienen estas características. Una de las cualidades más destacables en una metodología ágil es su sencillez, tanto en su aprendizaje como en su aplicación, reduciéndose así los costos de implantación en un equipo de desarrollo. Esto ha llevado hacia un interés creciente en las metodologías ágiles.

No obstante las fortalezas de dichas metodologías, y como se mencionó previamente, no existe una metodología que se adapte a todas las situaciones posibles. Por ello, será de gran utilidad tener una manera de evaluarlas y compararlas para determinar cuál de ellas se adapta más a un problema determinado.

## **Líneas de Investigación y Desarrollo**

Bajo estas consideraciones previamente descritas, hemos definido un método de evaluación y comparación de modelos conceptuales de procesos de negocio [13]. Para su definición, en una primera etapa dentro de esta línea de investigación, se realizaron algunos estudios en la comparación de lenguajes de modelado [15, 16]. Dichas propuestas se realizaron con el objetivo de brindar un medio que ayude en la toma de decisión a la hora de seleccionar y adoptar un lenguaje de modelado que se adecue más a las necesidades de las empresas.

En esta misma línea de investigación, el método propuesto se extendió y adaptó de manera que permita el análisis y evaluación de la calidad de los modelos de procesos de negocio, independientemente del lenguaje de modelado utilizado.

El objetivo final del método propuesto es unificar la expresión de conceptos básicos de procesos de negocio, de manera de facilitar la evolución y mejora continua de los procesos y las políticas de negocio de la empresa.

Se debe destacar que el método propuesto se centra en la calidad del producto. Además, se lo definió de manera que adhiera a los estándares de calidad y se independice de la notación utilizada para la definición de los modelos evaluados.

## Resultados Obtenidos y Objetivos

En un primer paso para la evaluación y validación del método propuesto, se lo utilizó ante el requerimiento de una empresa del medio que, si bien tenía en cierto grado modelado sus procesos de negocio, las exigencias competitivas le llevaron a tomar la decisión de realizar una reestructuración de su organización. Su objetivo era mejorar el rendimiento de aquellas áreas en las que se presentaban mayores conflictos o deficiencias.

En función de ello, se aplicó el método para la evaluación de los modelos de procesos preexistentes de la organización, y de los nuevos modelos definidos.

Además, como parte de la validación empírica del método, y bajo el concepto de que una metodología de desarrollo puede verse como un PN, se construyeron modelos que representan los metamodelos de las metodologías ágiles OpenUp [17], eXtreme Programming (XP) [18] y Scrum [19] y se aplicará el método para analizar y comparar dichas metodologías. Dicho análisis se realizará en función del número de tareas y recursos que cada metodología necesita para poder ser aplicada. La aplicación del método, no sólo en el análisis de modelos de PN reales, sino en los metamodelos de metodologías de desarrollo, nos dará una importante visión del alcance del mismo como también permitirá vislumbrar posibles extensiones en el futuro. Esto nos permitirá obtener una valoración de la aplicabilidad del método

en distintos ámbitos del modelado de procesos de negocio.

Cabe destacar que, previamente, el método se aplicó en un ámbito empresarial particular y a modelos de procesos de negocio de trabajos de estudiantes de posgrado con el fin de mostrar su aplicabilidad, tanto en el ámbito profesional como en el ámbito de la investigación.

## Formación de Recursos Humanos

Bajo esta línea, en el grupo de investigación se presentó un trabajo final de Especialización en Ingeniería de Software ([20]). En la actualidad se están realizando tesis de Maestría en ingeniería de Software referente al Modelado de Procesos de Negocio, analizando los modelos conceptuales a partir de las componentes existente en los distintos lenguajes de modelado del mercado, como así también se están llevando a cabo algunas tesinas de grado para la Licenciatura en Ciencias de la Computación.

## Referencias

- [1] M. A. Rappa, "The utility business model and the future of computing services," *IBM Systems Journal*, vol. 43, pp. 32-42, 2004.
- [2] J. Becker, M. Rosemann, and C. von Uthmann, "Guidelines of Business Process Modeling," *Business Process Management, Models, Techniques and Empirical Studies (BPM'00)*. Springer, pp. 30-49, 2000.
- [3] V. Vitolins, "Business Process Measures," presented at Int. Conference on BALTIC DB&IS. Riga, Latvia., 2004.
- [4] C. Dewalt, "Business Process Modeling with UML," *Johns Hopkins University*, 1999.
- [5] S. A. White, "Process Modeling Notations and Workflow Patterns," in *Workflow Handbook 2004*, L. Fischer, Ed.: Published in association

- with the Workflow Management Coalition (WfMC), 2004.
- [6] M. Piattini, F. Ó. Garcia Rubio, and I. Caballero, *Calidad de Sistemas Informáticos: Alfaomega-RA-MA*, 2007.
- [7] Fenton, "Software Measurement: A Necessary Scientific Basis," *IEEE Transactions on Software Engineering*. 20(3), pp. 199-206, 1994.
- [8] D. Moody, "Theoretical and practical issues in evaluating the quality of conceptual models: current state and future directions," *Data & Knowledge Engineering. Elsevier B.V.*, pp. 243–276, 2005.
- [9] ISO, "ISO Standard 9000-2000: Quality Management Systems: Fundamentals and Vocabulary, International Standards Organisation (ISO)." 2000.
- [10] ISO/IEC, "ISO/IEC Standard 9126: Software Product Quality, International Standards Organisation (ISO), International Electrotechnical Commission (IEC)," 2001.
- [11] E. Rolon, F. Ruiz, F. Ó. Garcia Rubio, and M. Piattini, "Aplicación de Métricas Software en la Evaluación de Modelos de Procesos de Negocio," *Revista Electrónica de la Sociedad Chilena de Ciencia de la Computación*, 2005.
- [12] F. Ó. García Rubio, "FMESP: Marco de Trabajo Integrado para el Modelado y la Medición de los Procesos Software," in *Departamento de Informática*. Ciudad Real. España: U.C.L.M. Universidad de Castilla La Mancha. España, 2004, pp. 491.
- [13] C. Salgado, M. Peralta, M. Berón, and G. Montejano, "Un Método para la Evaluación de Modelos Conceptuales de Procesos de Negocio. Un Caso de Estudio," presented at Jornadas Argentinas de Informática - JAIIO 2012, La Plata, Bs. As, Argentina, 2012.
- [14] ISO/IEC, "ISO/IEC Standard 14598: Software Product Evaluation, International Standards Organisation (ISO), International Electrotechnical Commission (IEC)," 1999.
- [15] C. Salgado, M. Peralta, M. Berón, D. Riesco, and G. Montejano, "SLMPN: un Modelo para la Evaluación y Comparación de Lenguajes de Modelado de Procesos de Negocio," *Proceedings of ASSE 2010 - 39 JAIIO 2010 - UADE, Buenos Aires*, 2010.
- [16] N. Debnath, C. Salgado, M. Peralta, D. Riesco, M. Berón, and G. Montejano, "A Strategy Based on Lsp for the Evaluation of Specific Languages for Business Process Modeling," presented at 20th International Conference on Software Engineering and Data Engineering (SEDE 2011), Las Vegas - USA, 2011.
- [17] M. Yang, "Introduction to OpenUP," <http://epf.eclipse.org/wikis/openup/>, 2012.
- [18] XP, "Extreme Programming: A gentle introduction," <http://www.extremeprogramming.org/>, 2012.
- [19] J. Eaglesham, "Scrum Overview," <http://epf.eclipse.org/wikis/scrum/>, 2012.
- [20] C. Salgado, "Tesis de Especialización en Ingeniería de Software: El Modelado de Procesos de Negocio: Aplicando LSP para la Evaluación de Lenguajes de Modelado de Procesos de Negocio," in *Facultad de Ciencias Físico-Matemáticas y Naturales. UNSL*, 2010.



## GESTIÓN DEL MANTENIMIENTO APLICANDO UNA METODOLOGÍA BASADA EN SISTEMAS Y TECNOLOGÍAS DE INFORMACIÓN - FASE II

**Abet Jorge Eduardo (1), Carrizo Blanca Rosa (2), Nacuse Oscar (3)**

GICAPP “Grupo de Investigación en Control Avanzado de Procesos y Producción”. / Dpto. Ingeniería Industrial / Dpto. Ingeniería Mecánica / Facultad Regional Córdoba / Universidad Tecnológica Nacional  
Maestro Marcelo López s/n. Ciudad Universitaria.

Teléfono: 5986011

[jabet@industrial.frc.utn.edu.ar](mailto:jabet@industrial.frc.utn.edu.ar) / [bcarrizo@tecnicatura.frc.utn.edu.ar](mailto:bcarrizo@tecnicatura.frc.utn.edu.ar)

### RESUMEN

Esta propuesta de trabajo es la continuación del PID VAPRC0778 denominado “Estudio probabilístico de Fallos, uso del Dataminig (DM) y Datawarehouse (DW) para su aplicación al Mantenimiento” Fase I; cuyo objetivo era cómo aplicar herramientas informáticas basadas en Tecnologías y Sistemas de Información (TI/SI) de Base de Datos, como lo son el Almacén de Datos (DW) y la Minería de Datos (DM) en el estudio de longevidad y fallos de los equipos, a través de la predicción de tendencias y comportamientos que permitan tomar decisiones proactivas basadas en repositorios de datos históricos y en la criticidad de los sistemas en funcionamiento.

Dado que, no se han popularizado modelos de análisis de información especialmente orientados a la gestión del mantenimiento; la actual propuesta plantea optimizar un modelo planteado en la Fase I para la gestión del mantenimiento, que se realiza a partir de los atributos de fallas presentados en equipos y sistemas, que automatice esta función; basado en un historial de fallas presentadas en los equipos y sistemas durante determinado periodo de análisis.

En resumen, el resultado tangible será desarrollar una capa de presentación estándar para un modelo optimizado que será configurable para distintos tipos de industria a través del cual el usuario podrá diseñar sus salidas de información a medida de sus requerimientos y potenciarla mediante servicios de mensajería, alertas, notificaciones, etc.

**Palabras Clave:** Gestión. Fiabilidad. Mantenimiento. Sistemas de Información (SI). Tecnologías de Información (TI). Servicios de Transferencia de Datos (DTS).

### 2. CONTEXTO

Este trabajo denominado “Gestión del Mantenimiento aplicando una metodología basada

en SI y TI - Fase II”, Código UTN1652, forma parte del pool de proyectos del GICAPP, “Grupo de Investigación en Control Avanzado de Procesos y Producción”, creado por Res. N° 522/99 de Rectorado, encubado en el seno del Dpto. Ingeniería Industrial.

### 1. INTRODUCCIÓN

Objetivo de esta propuesta es optimizar el modelo de almacenamiento desarrollado en la Fase I, estandarizando sus entradas de datos mediante herramientas automáticas de integración de datos (DTS) a la estructura de la Base de Datos, con la finalidad de mejorar la gestión de mantenimiento de una Industria pertinente al objeto de estudio, que brinde una interfaz gráfica y amigable al usuario para el diseño de sus salidas de información específicas.

En lo referido a metodología a aplicar, la estrategia para la obtención de datos, está basada en un enfoque mixto que combina estrategias cualitativas (análisis del contenido de documentación, observación y entrevistas) con estrategias cuantitativas (pruebas o testing y métricas), desde la tipología del proyecto.

Es decir que, el desarrollo de un SI (Sistema de Información) integra una solución estándar a través del desarrollo de los servicios de transferencia de datos (DTS); la cual será por fases o etapas, dado que el mismo está encuadrado dentro de un ciclo de vida útil de un SI; en el cual cada etapa no es excluyente; ya que a veces hay actividades que se comparten o realizan en forma paralela.

Dado que, el ciclo de desarrollo de un SI comprende distintas etapas, los instrumentos que se utilizan se detallan a continuación:

- Etapa I – Requerimientos.
  - Herramientas: Entrevistas, observaciones, análisis de documentación existente.
- Etapa II - Análisis / Diseño.

- Herramienta: Lenguaje Unificado de Modelado (UML), y algunos de sus diagramas: de estructura, de comportamiento e Interacción.
- Etapa III – Construcción.
  - Herramienta: Lenguaje de Programación basado en la filosofía de Software Libre o Propietario.
- Etapa IV – Pruebas.
  - Herramientas: Prueba de Caja Negra, Prueba de Caja Blanca, Pruebas de versión alfa y beta, Pruebas de Volumen y Performance, Pruebas de Regresión, Pruebas de Estrés, Prueba Unitaria y Prueba de Integración
- Etapa V - Producción / Mantenimiento.
  - Herramientas: métricas referidas a simplicidad (Nº ciclomático, Nº Secuencias), concisión (longitud del programa, nivel del módulo), autodescriptivo (frecuencia de comentarios), legibilidad (niveles anidados), entre otros.

## 2. LINEAS DE INVESTIGACIÓN Y DESARROLLO

- Mantenimiento y predicción de fallos. [1]
- Bases de Datos y Almacén de Datos. [2]
- Fiabilidad. [3]
- Ciclo de Vida de un Sistema de Información.
- Lenguajes de programación orientados al desarrollo web. [4]
- Métricas de calidad para el modelo de conocimiento obtenido.

## 3. RESULTADOS Y OBJETIVOS.

Entre los resultados del avance de este proyecto se pueden mencionar los siguientes:

- Se ha optimizado el diseño del Datawarehouse, basado en el Sistema de Administración de Bases de Datos relacionales, ratificando su estructura parametrizable a distintos requerimientos estándares de registro de fallos.
- Se ha iniciado el desarrollo de un SI que integra una solución estándar a través del desarrollo de los servicios de transferencia de datos e implica el desarrollo de herramientas automáticas de integración de datos, que estandarice los mismos.
- Se ha continuado transfiriendo al seno de la cátedra “Mantenimiento”, pertenecientes tanto a las carreras Ing. Industrial como Mecánica, los avances / resultados obtenidos; así como al seno de la Tecnicatura en Mantenimiento Industrial;

integrando estas actividades académicas en eventos comunes, charlas y firma de convenios.

- Se ha continuado trabajando con la firma Cba Solutions, en lo referido a explorar otras aplicaciones orientadas a la gestión del mantenimiento, que nos permitan mejorar el estudio probabilístico de fallos en forma automática.
  - Se ha participado activamente en la creación del IAM (Instituto Argentino de Mantenimiento), Modelo Universitario de Organización Empresa - Universidad que pretende constituirse en un punto de encuentro entre los diversos elementos tecnológicos y pedagógicos que posee la Universidad sumados al aporte indiscutible de los profesionales y técnicos de las empresas de la Región de Córdoba en particular y el país en lo General, en pos de un mismo objetivo: el Mantenimiento. La Facultad Regional Córdoba de la Universidad Tecnológica Nacional avala esta iniciativa a través de la Res. Nº 215/11.
  - Se ha concretado un intercambio interregional, en el marco de un Acta de reciprocidad suscripta entre las Regionales Córdoba y San Nicolás, en la carrera Ingeniería Industrial, entre las cátedras de 1º Nivel “Informática I” y “Pensamiento Sistémico”, donde se analizó desde punto de vista práctico un caso real de Mantenimiento relevado de una empresa metalúrgica de San Nicolás; desde una visión automatizable del mismo (primero en planilla de cálculo, luego en Sistema de Bases de Datos y finalmente en un desarrollo a medida). Esta actividad académica estuvo dirigida a toda la comunidad educativa y se desarrolló para toda la comunidad universitaria en el Salón de Usos Múltiples (SUM) de la Regional.
  - Se ha participado activamente en Congresos Nacionales (I, II, III y IV CAIM / COINI 2012, JEIN 2012, WICC 2012), Latinoamericanos (CLICAP 2012) como Internacionales (CCO 2012 en Venezuela y 1º Congreso Latinoamericano de Educación a Distancia).
- Dado que, el desarrollo de este Sistema será por fases o etapas, es necesario destacar que cada etapa no es excluyente; ya que a veces hay actividades que se comparten o realizan en forma paralela.
- De las etapas arriba detalladas, cabe aclarar que las Etapas I y II se han cumplimentado casi en su totalidad; lo cual representa un 40% del proyecto aproximadamente.

El objetivo de la “Etapa I – Requerimientos” fue desarrollar de un modelo del área Mantenimiento, donde se representan los procesos que se llevan a cabo, la información utilizada por ellos y las reglas políticas y prácticas de la Empresa relacionada con estos procesos.

Se proyectaron estrategias, procesos y flujos de datos al igual que las interrelaciones entre procesos y datos, con el fin de desarrollar un plan de sistema de información.

Algunas herramientas utilizadas en esta instancia fueron: entrevistas a usuarios, observaciones en lugar de trabajo, análisis de documentación física existente, análisis de datos registrados en soportes lógicos (planillas de cálculo).

El objetivo de la “Etapa II - Análisis / Diseño” fue desarrollar el diseño arquitectónico de los sistemas, utilizando los requerimientos obtenidos en la primera etapa; en el cual se engloben dos componentes: los datos y los procesos; los cuales serán analizados y diseñados desde una perspectiva conceptual a una física.

En esta etapa, se incluyó la sub-etapa evaluación de paquetes; es decir la factibilidad técnica, económica y operativa de implementar determinado lenguaje de programación, en función del motor de Base de Datos ya implementado en nuestro caso.

Este Lenguaje de Programación a seleccionar, basado en la filosofía de Software Libre o Propietario; interactuará con el motor de Base de Datos ya diseñando bajo la filosofía de software propietario.

Algunas herramientas utilizadas fueron: Lenguaje Unificado de Modelado y los diagramas de estructura, comportamiento e interacción.

Actualmente, se ha iniciado la Etapa III – Construcción, donde se ha planificado los tiempos y recursos involucrados en el desarrollo de la Infraestructura de la DTS, verificar la adaptación del paquete así como el desarrollo de Unidades de diseño Interactivas y de Unidades de Diseño Batch; para concluir con el diseño de las Unidades de diseño Manuales.

Se ha planificado avanzar, en este ciclo lectivo en las etapas siguientes:

- Etapa IV – Pruebas, que consiste en verificar si el sistema desarrollado no falle y funcione de acuerdo a sus especificaciones y a la manera que los usuarios esperan que lo haga, y de esta forma poder detectar cualquier anomalía, antes de que el sistema sea puesto en marcha y se dependa de él.

- Etapa V - Producción / Mantenimiento, se inicia una vez que el sistema pasa a formar parte de la vida diaria de la empresa, cada programa, cada procedimiento y cada estructura de datos se convierte en una pieza del negocio que, como tal, deberá funcionar en forma constante, exacta y confiable.

En nuestro caso en particular, y teniendo en cuenta que el resultado de este proyecto será transferido a Pymes de nuestra provincia, se analizará la relación costo - beneficio en lo referido a selección no sólo del lenguaje de programación sino a su interacción con el motor de Base de Datos.

#### 4. FORMACIÓN DE RECURSOS HUMANOS

Todos los integrantes de este PID son Especialistas en Docencia Universitaria, todos poseen categoría de Investigador tanto en la UTN como en Programa de Incentivos del MECyT.

Cabe aclarar que, todos son Profesores Ordinarios con cargos de Titular (dos), Asociado (dos) y Adjunto (dos).

Algunos integrantes forman parte del GICAPP, del FODAMI (Foro de Consulta y Perfeccionamiento Docente del Área Mecánica de las Ingenierías) y del INAM (Instituto Nacional de Mantenimiento).

En el seno de la cátedra de “Mantenimiento”, tanto de Ing. Industrial como Mecánica, se desarrolla un trabajo de campo extraído de un caso real de una Empresa, en la cual el alumno aplica significativamente todos los contenidos aprendidos en el aula y la transfiere al ámbito laboral en el cual está inserto (en el 80% de los casos). Algunos de estos trabajos, con ampliados a nivel complejidad de requerimientos por el Profesor de Trabajo Final o Práctica Supervisada y se transforman en Tesinas de Grado.

#### 5. BIBLIOGRAFÍA

- [1] “Fiabilidad, mantenibilidad, efectividad un enfoque sistémico”. Alberto Sols (2002).
- [2] “Extracción Automática de Conocimiento en Base de Datos e Ingeniería del Software”. Quintana Ramírez María José, Orallo José Hernández, España (2003).
- [3] “Teoría y Práctica del Mantenimiento”. Francois Monchy. (1995).
- [4] “Sistemas de Información para la gestión empresarial”. Alberto Lardent. (2001).

# Definición de métricas en un metamodelo usando OCL para diseño Web

Jesús Francisco Aguirre y Daniel Riesco

Departamento de Informática, Facultad de Ciencias Físico Matemáticas y Naturales - UNSL

Ejército de los Andes 950 - San Luis - Argentina

{jaguirre,driesco}@unsl.edu.ar

## Resumen

La relevancia que han tenido en el último tiempo las aplicaciones Web ha producido una evolución en los métodos, herramientas y lenguajes que dan soporte al *Proceso de Desarrollo de Software* (PDS). De esta manera, el software obtenido es de alta calidad, permitiendo lograr productos competitivos [1]. A pesar de esto, existen carencias en la utilización de las recomendaciones de los expertos y de las organizaciones que dan soporte para mejorar la usabilidad, la accesabilidad, la navegación, etc.

El *Object Management Group* (OMG) propone un PDS denominado *Arquitectura de Desarrollo basado en Modelos* (MDA) que está basado en modelos y transformaciones entre ellos y permite facilitar la especificación y el desarrollo de las aplicaciones Web. De esta forma se logra, a partir de los modelos, la generación automática de código en forma transparente a las diferentes plataformas tecnológicas.

El objetivo de esta línea de investigación es definir una propuesta para definir métricas formales usando OCL sobre la metodología *Object-Oriented Web Solutions* (OOWS) para mejorar el desarrollo del software Web a nivel de diseño en un entorno MDA.

**Palabras clave:** MDA, OOWS, metamodelos, modelos, métricas, diseño web, OCL

## Contexto

Si bien existen algunas investigaciones relacionadas a la temática que se aborda en esta línea

de investigación, las mismas están orientadas a la evaluación de la usabilidad utilizando diversas metodologías para aplicaciones en general [1, 2]. Esta línea de investigación está enfocada únicamente en el PDS de las aplicaciones Web, debido a que éstas presentan características diferentes al PDS en comparación con otras aplicaciones en el contexto de MDA.

Esta propuesta de investigación se desarrolla como tesis de la Maestría en “Ingeniería de Software” de la Universidad Nacional de San Luis. La misma posee un convenio para la doble titulación en el marco del *Programa Centros Asociados para el Fortalecimiento de Posgrados Brasil/Argentina* (CAFP-BA).

Este Proyecto es denominado “Contribución a la homogeneidad académica regional considerando estándares de alta calidad y a la competitividad de la Industria del Software, mediante la formación de Recursos Humanos de Cuarto Nivel en Ciencias Aplicadas a la Informática y Computación, aplicado en las zonas de influencia de la Universidad Nacional de San Luis y de la Universidad Federal de Minas Gerais”, vigente desde el año 2008.

## Introducción

Los procesos software constituyen la base fundamental de una Organización de Software. Dichos procesos se aplican en forma de proyectos, y como concreción de éstos se obtienen los procesos. Por lo tanto, surge la necesidad de la *Ingeniería de Software* (IS) por mejo-



rar continuamente el PDS, dando origen nuevas metodologías, lenguajes, patrones y herramientas CASE. Cada aproximación propone métodos y estrategias diferentes que permiten gestionar la complejidad y la diversidad del desarrollo de las aplicaciones para la definición, ejecución, medición, control y mejora de sus procesos [3].

Entre estas etapas de desarrollo se encuentra la medición del software [4], creada para obtener alta calidad en las aplicaciones. A partir de ello, surge una disciplina de medición y análisis de datos para que se realice de una manera efectiva y consistente [5].

La IS se ha concentrado históricamente en problemas de funcionalidad y de persistencia, dejando a un segundo plano aspectos de la interacción con el usuario. Con la aparición de la *Ingeniería Web* (IW), se promueve el desarrollo de aplicaciones que guíen el PDS teniendo en cuenta los atributos de calidad más relevantes en este tipo de aplicaciones, como son la usabilidad, la accesibilidad, la navegabilidad, el contenido, etc., de esta manera se minimizan los riesgos del desarrollo que pueden tener serias implicaciones. Existen estándares [6], guías [7, 8, 9, 10] o recomendaciones de expertos [11, 12] que muestran las características que los Sistemas de Información Web (SIW) deben cumplir para mejorar la usabilidad.

El estándar ISO/IEC 9126-1 [6] define la usabilidad como “la capacidad de un producto de software de ser entendido, aprendido, usado y atractivo para el usuario, cuando éste es usado bajo condiciones específicas”. A su vez, estas características se descomponen en atributos sobre los que luego se pueden aplicar métricas para obtener su valor de usabilidad.

Generalmente las herramientas no evalúan al SIW durante el desarrollo del mismo, lo realizan al final del PDS. Estas herramientas validan el código HTML y CSS del SIW [13, 10], sin poder considerar adelantar algunas de estas mediciones sobre los modelos de navegación, presentación, etc.

Ante esta situación, este trabajo considera realizar la definición de algunas de las métricas más representativas durante el modelado del SIW. Existen numerosos trabajos sobre la evaluación y

mejora de la usabilidad [14]. Debido a que la calidad de los modelos navegacionales es una característica importante en los SIW, se ha seleccionado un subconjunto de métricas para evaluar su calidad en tiempo de modelado. Por ejemplo, establecer el número máximo de links que un usuario necesita para realizar una tarea, número de links a otras secciones de la página, número de links a otras páginas del sitio Web, etc. El resto de las métricas no fueron tenidas en cuenta porque no pueden ser medidas y formalizadas sobre los modelos.

OMG [18] propone un abordaje conocido MDA que facilita el PDS, utilizando un conjunto de especificaciones abiertas “independientes del vendedor”. Estas especificaciones permiten resolver la integración y la interoperabilidad e incluyen al *Lenguaje de Modelado Unificado* (UML) [19], a la *Facilidad de Meta Objetos* (MOF) [20], al *Metamodelo Común de Almacenes de Datos* (CWM) [21] y al *Intercambio de Metadatos XML* (XMI) [22].

Este enfoque utiliza como idea principal iniciar el PDS a partir de modelos de alto nivel de abstracción, para posteriormente ir transformando estas abstracciones en primitivas de más bajo nivel (M2M), hasta que en última instancia se podrá corresponder con líneas de código en un lenguaje de programación (M2T) [16, 17]. De esta manera un compilador genera aplicaciones funcionales a partir de modelos conceptuales sin que el analista necesite escribir una línea de código. Estos modelos utilizan una notación basada en UML.

El propósito de este trabajo es mostrar como el “modelo de medición” puede integrarse con los “modelos navegacionales” de una metodología en concreta como OOWS [23]. Este proceso de desarrollo se divide en 3 etapas: 1) Especificación de los requisitos de usuarios, 2) Modelado conceptual y 3) Desarrollo de la Solución. OOWS propone para la etapa del modelado conceptual, describir la aplicación Web con primitivas de alto nivel de abstracción a través de 6 modelos: *Modelo de Objetos* (describe la estructura estática del sistema), *Modelos Dinámicos y Funcional* (describen el comportamiento del sistema), *Modelo de Usuarios* (define los potenciales usuarios del sistema), *Modelo de Navegación* (describe la nave-

gación permitida para dichos usuarios) y *Modelo de Presentación* (define las propiedades de la presentación de la información). El desarrollo de la solución es generado automáticamente por la aplicación Integranova Model Execution System [24].

Por ello, se está desarrollando una tesis para definir métricas formales sobre el metamodelo OOWS [25] para mejorar el PDS Web. El objetivo principal es lograr identificar los atributos de usabilidad más representativos que puedan ser considerados en etapas tempranas del PDS para definirlos en OCL[26]. La ventaja de considerarlos al comienzo incorpora mayor complejidad al desarrollo debido a tener que contemplar mayor cantidad de información.

## Líneas de Investigación y desarrollo

La presente línea de Investigación esta orientada a profundizar en:

- Analizar las diferentes herramientas para la evaluación de usabilidad en métodos de desarrollo de software Web basado en transformaciones de modelos,
- Definir el metamodelo navegacional de la metodología OOWS,
- Estudiar la semántica navegacional de los modelos conceptuales proporcionados por la metodología OOWS para la identificación de las primitivas y propiedades conceptuales sobre los mapas navegacionales,
- Identificar las diferentes métricas que tiene en cuenta la literatura para evaluar la usabilidad en sistemas Web a través de entornos MDA de una manera temprana y automática,
- Definir las métricas que expresan la navegabilidad sobre los mapas navegacionales usando la metodología OOWS,
- Expresar las anteriores métricas mediante expresiones OCL.

## Resultados y Objetivos

La mayor parte de las propuestas de evaluación de la usabilidad estan orientadas a métodos de desarrollo de software tradicionales. Pocas son las propuestas que definen la usabilidad de una manera abstracta.

El principal objetivo de esta tesis es la adaptación de mecanismos de usabilidad para los PDS utilizando el entorno MDA y sus estándares e integrándola con la metodología OOWS. La idea principal reside en modelar la usabilidad a nivel conceptual utilizando las primitivas de abstracción que permiten representar la navegación. Esto permite que al utilizarla en etapas tempranas del PDS los beneficios obtenidos sean mayores, porque la generación de código a partir de los modelos conceptuales es automática y por lo tanto, si una característica de usabilidad no satisface los requisitos del usuario, se podrán hacer los cambios necesarios en el modelo conceptual y volver a generar el código de la aplicación Web.

## Formación de Recursos Humanos

Actualmente en esta línea de investigación se está realizando una tesis de maestría en “Ingeniería de Software” perteneciente a la Universidad Nacional de San Luis, en el marco del Proyecto 004/08 del programa CAFB-BA (Centros Asociados para el Fortalecimiento de Posgrados Brasil/Argentina). Además se aprobaron durante el 2010, la tesis de maestría denominada “Definición de Métricas en OCL según el Metamodelo de la OMG aplicadas al Diseño Orientado a Aspectos”, utilizando la arquitectura OMG, cuyo director es Daniel Riesco (Diciembre de 2010), y otra tesis denominada “Transformación en QVT de Procesos de Desarrollo de Software Basados en SPEM a Workflows, usando arquitectura MDA”, tesis defendida en la Universidad de la República, Uruguay, cuyo Director fue Daniel Riesco (Octubre de 2010) y codirectora Nora Szazs.

## Referencias

- [1] Abrahão, S., Pastor, O. Olsina, L.; "A Quality Model for Early Usability Evaluation", INTERACT05 International COST 294 Workshop on User Interface Quality Models (UIQM), Rome, Italy. pp: 68 a 77, 2005.
- [2] Abrahão, S., Insfrán, E.; "Early Usability Evaluation in Model Driven Architecture Environments", Sixth International Conference on Quality Software (QSIC'06), Beijing, China. pp. 287-294, 2006.
- [3] Florac, W., Carleton, A., Barnard, J.; "Statistical Process Control: Analyzing a Space Shuttle Onboard Software Process", IEEE Software, 2000.
- [4] Fenton N., Pfleeger S.; "Software Metrics: A Rigorous Approach", Second Edition, PWS 1998.
- [5] Brown, M., Dennis, G.; "Measurement and Analysis: What Can and Does Go Wrong?", 10th IEEE International Symposium on Software Metrics (METRICS'04), pp. 131-138, 2004.
- [6] ISO/IEC 9126-1, Software engineering - Product quality - 1: Quality model, 2001.
- [7] Guías de Estilo de Java, "http://www.java.com", última visita 1 de abril de 2012.
- [8] Guías de estilo de Microsoft, "http://www.microsoft.com", última visita 1 de abril de 2011.
- [9] Usability Guides, U.S. Department of Health and Human Services, "http://usability.gov", última visita 1 de abril de 2012.
- [10] World Wide Web Consortium (W3C), "http://www.w3c.es/", última visita 1 de abril de 2012.
- [11] Nielsen J., Designing Web Usability: The practice of Simplicity, New Riders Publishing, 2000.
- [12] Krug S.; No me hagas pensar. Una aproximación a la usabilidad en la Web, segunda edición, Prentice Hall, 2005.
- [13] T. A. W.: Test de Accesibilidad Web, URL: <http://www.tawdis.net> (Consulta: noviembre de 2012).
- [14] Ivory, M., Hearst, M.: Towards Quality Checkers for Web Site Designs.
- [15] Frankel, D.; "Model Driven Architecture", Wiley Publishing Inc., 2003.
- [16] Kleppe, A., Warmer, J., Bast W.; "MDA Explained: The Model Driven Architecture: Practice and Promise", Addison Wesley, 2003.
- [17] Mellor, S., Scott, K., Uhl, A., Weise, D.; "MDA Distilled: Principles of Model-Driven Architecture", Addison Wesley, 2004.
- [18] OMG, Object Management Group, "http://www.omg.org", (Consulta: noviembre de 2010).
- [19] UML Unified Model Language, 2011, "http://www.omg.org/spec/UML", Version 2.4.1 (Consulta: noviembre 2011).
- [20] MOF, Meta Object Facility core Specification, 2011, "http://www.omg.org/spec/MOF", Version 2.4.1 (Consulta: noviembre 2011).
- [21] CWM, Common Warehouse Metamodel Specification, 2003, "http://www.omg.org/spec/CWM", Version 1.1, Volume 1 (Consulta: noviembre 2011).
- [22] XMI, MOF 2 XMI Mapping Specification, 2011, "http://www.omg.org/spec/XMI", Version 2.4.1 (Consulta: noviembre 2011).
- [23] Conceptual modelling of Web applications: the OOWS approach, capítulo del libro "Web Engineering - Theory and Practice of Metrics and Measurement for Web Development, Mendes E. (Eds.) Springer 2005 (2005)

- [24] Integranova Model Execution System (IMES). Care Technologies, “<http://www.care-t.com>” (Consulta: Octubre 2012)
- [25] Generación Automática de Servicios Web a partir de Modelos Conceptuales, Tesis de Doctorado de Marta Ruiz Server (2010)
- [26] OCL, Object Constraint Language, 2012, “<http://www.omg.org/spec/ocl>”, Version 2.3.1 (Consulta: Noviembre 2011).



## Aplicación de Técnicas de Comprensión de Programas para la Obtención de Información de Modelos de Procesos Workflow

M. Peralta, C. Salgado, M. Berón, D. Riesco, G. Montejano, L. Baigorria  
Departamento de Informática – Facultad de Ciencias Físico Matemáticas y Naturales  
Universidad Nacional de San Luis  
Ejército de los Andes 950, San Luis, San Luis, Argentina  
C.P. 5700 – Tel.: 54-026644520300 – Int. 2101  
[mperalta, csalgado, mberon, driesco, gmonte, flbaigor]@unsl.edu.ar  
web: <http://www.sel.unsl.edu.ar>

### Resumen

Una de las tecnologías más significativas para soportar la automatización de los procesos de negocio son los Sistemas de Gestión Workflow. Para favorecer y dar flexibilidad a dichos sistemas, de manera que puedan adaptarse a los cambios constantes del negocio, es fundamental tener un lenguaje de modelado que permita una fácil definición y adaptación de los modelos. Igualmente, es fundamental tener herramientas que permitan medir la calidad de dichos modelos. Acorde a esto, en esta línea de investigación, nuestra propuesta se centra en la definición de un marco de trabajo para el modelado y medición de procesos workflow que ayude a la mejora y mantenimiento de los modelos y de los procesos que ellos representan.

Dicho marco pretende proveer un medio que permita incluir todos los aspectos más relevantes del modelado workflow. Para ello consideramos que es necesario definir un conjunto más amplio de medidas que permitan la medición individual de cada uno de los aspectos relevantes para el modelado, como así también la complejidad estructural global del proceso completo y de los modelos que lo representan.

**Palabras clave:** Workflow – Sistema de Gestión Workflow – Proceso de Negocio – Lenguaje de Modelado Workflow – Métricas – Especificación de Métricas – XPD L – XQUERY.

### Contexto

El presente trabajo se enmarca en el Proyecto de Investigación: Ingeniería de Software: Aspectos de alta sensibilidad en el ejercicio de la profesión de Ingeniero de Software – Facultad de Ciencias Físico-Matemáticas y Naturales, Universidad Nacional de San Luis. Proyecto N° 22/F222. Dicho proyecto es la continuación de diferentes proyectos de investigación a través de los cuales se ha logrado un importante vínculo con distintas universidades a nivel nacional e internacional. Además, se encuentra reconocido por el programa de Incentivos.

### Introducción

Analizando el ciclo de vida de los procesos de negocio (Georgakopoulos and Tsalgati dou 1998), es de gran importancia llevar a cabo una mejora continua de los mismos. Ello ha llevado a las organizaciones a buscar herramientas que proporcionen el soporte necesario para poder realizar dichas mejoras. Hoy en día, la Gestión de Procesos de Negocio (BPM), definida como “*la capacidad de descubrir, diseñar, desarrollar, ejecutar e interactuar con la operación, optimización y análisis de procesos de negocio a nivel de diseño*” (Smith 2002), proporciona este soporte mediante los Sistemas de Gestión de Procesos de Negocio (BPMS). En este sentido, una de las tecnologías más significativas para dar soporte a la gestión de procesos de negocio son los Sistemas de Gestión de Flujos de Trabajo (Workflow Management Systems, WFMS) que soportan la automatización de los procesos de negocio y que la Workflow

Management Coalition define como: “*un sistema que define, crea y dirige la ejecución de flujos de trabajo a través del uso de software que funciona en uno o más motores de flujo de trabajo, siendo capaz de interpretar la definición de proceso, de interactuar con los participantes del flujo de trabajo e invocar el uso de herramientas y aplicaciones de las IT*” (Lawrence 1997).

En el campo de los procesos workflow, se puede observar un importante trabajo en cuanto al modelado de dichos procesos, surgiendo así diversas líneas de investigación. Entre estos trabajos podemos destacar propuestas en las que se utilizan los Patrones Workflows para realizar la evaluación y/o comparación de distintos lenguajes de modelado. Por ejemplo, en (Wohed, van der Aalst et al. 2002) se utilizan los patrones workflow, junto con un conjunto de patrones de comunicación para analizar BPEL4WS. En el trabajo se presenta una representación posible de cada patrón, respecto del flujo de control, en BPEL4WS. Además, se lo compara con otros lenguajes de modelado workflow como XLANG y WSFL, y Staffware PLC's Staffware e IBM's MQSeries Workflow, cuya evaluación es presentada en (van der Aalst, ter Hofstede et al. 2002). En (White 2004) se examina cómo dos lenguajes de modelado: *Diagramas de proceso de Negocio* (BPMN) de BPMI y los *Diagramas de Actividad* de UML de la OMG pueden representar gráficamente los patrones workflow respecto al flujo de control. En (Wohed, van der Aalst et al. 2004) los autores presentan una evaluación de los diagramas de Actividad contrastados con los patrones workflow desde las tres perspectivas: del flujo de control, datos y recursos. En (Wohed, van der Aalst et al. 2006) se presenta una evaluación de BPMN en función de su expresividad respecto de los Patrones Workflow.

Desde otro punto de vista, en el ámbito de la medición, se pueden observar diversos trabajos en cuanto a la medición de los

procesos de negocio como en (Rolon, Ruiz et al.) y (Rolon, Garcia Rubio et al.). Sin embargo, y a pesar de la importancia de medir la calidad de los procesos workflow, es muy poco el trabajo detectado en cuanto a la medición de calidad de estos procesos y de los modelos que los representan. Entre los trabajos en este campo se puede destacar las propuestas realizadas en (Cardoso 2005)(Cardoso 2005)[13] (Cardoso 2005), y (Cardoso 2006), donde se propone una métrica para la medición de la complejidad del flujo de control basada en la complejidad ciclomática de McCabe. En (Cardoso 2006) se introduce el concepto de complejidad workflow y se propone una definición de esta complejidad y una clasificación de los procesos workflows de acuerdo a su complejidad en *Ordenados, Estructurados y Aleatorios*. Otros trabajos que se pueden mencionar en este campo es la propuesta introducida en (Reijers 2003; Reijers and Vanderfeesten 2004). En estos trabajos los autores introducen métricas para evaluar la cohesión interna de las actividades en un proceso workflow y el acoplamiento entre sus actividades.

Estas propuestas presentan alternativas para la medición de algunos aspectos del modelado de procesos workflows. Pero no cubren todos los aspectos relevantes a los mismos.

Bajo estas consideraciones, en el ámbito de nuestro trabajo de investigación, consistente en la definición de un marco de trabajo para el modelado y la medición de procesos workflow que ayude a la mejora y mantenimiento de los modelos y de los procesos que ellos representan, hemos definido un conjunto de medidas para la medición de la calidad de modelos de procesos workflow desde el punto de vista de su mantenibilidad (Peralta, Garcia Rubio et al. 2008). Las medidas propuestas ayudarán a la evaluación, comparación y mejora de los modelos workflow y, en consecuencia, de los procesos que ellos representan. En este contexto, hemos realizado experimentos aplicando dichas

medidas para la evaluación y comparación de lenguajes de modelado (Peralta, García et al. 2007; Debnath, Peralta et al. 2011), como parte del proceso de validación práctica de las medidas propuestas.

Dentro del proyecto se está trabajando en la Comprensión de Programa (CP) en este punto se intenta utilizar técnicas, métodos y herramientas de esta área de estudio aplicadas al estudio y análisis de los procesos Workflow.

La *Comprensión de Programas* (CP) es un área de la Ingeniería de Software que tiene como objetivo facilitar el entendimiento de los sistemas, mediante el desarrollo de Métodos, Técnicas, Estrategias y Herramientas que permiten comprender las funcionalidades del sistema de estudio.

Uno de los principales desafíos en CP es establecer una relación entre el *Dominio del Problema* y el *Dominio del Programa*. Es decir, poder relacionar el comportamiento del sistema de estudio con las componentes del mismo que producen dicho comportamiento. Una forma de construir esta relación consiste en elaborar una representación para cada dominio y luego establecer un procedimiento de vinculación entre ambas representaciones. Para lograr esto, es necesario poder extraer información de ambos dominios (para poder crear las representaciones), para lo cual existen múltiples técnicas.

### **Líneas de Investigación y Desarrollo**

Todo sistema de gestión workflow debe garantizar que la organización realiza las tareas correctas, en el momento y de la forma adecuada. Por ello se considera fundamental tener una buena representación del proceso en la que se incluyan todos los aspectos de interés para dicho proceso. Para poder lograr esta representación es necesario disponer de un lenguaje de modelado de procesos adecuado que, además, permita adaptar esa representación a los cambios continuos que los negocios de hoy experimentan.

Otro de los aspectos fundamentales en todo proceso workflow, es tener herramientas que permitan medir sus elementos más relevantes con el fin de detectar las áreas y aspectos a mejorar del proceso y, de esta manera, promover su mejora continua. Para ello creemos que es necesario proveer un marco que permita realizar dicha medición y tener criterios que ayuden a tomar esta decisión.

El grupo de investigación centra su trabajo en desarrollar nuevos métodos, técnicas, herramientas y procesos, que ayuden a los ingenieros de software e informáticos a construir sistemas de mayor calidad.

De acuerdo a esta necesidad, siguiendo la metodología propuesta en (Serrano, Piattini et al. 2002), hemos definido un conjunto de medidas iniciales y elementales que creemos servirán como indicadores de la complejidad estructural de los modelos de procesos workflow.

Apoyándonos en la comprensión de programa, algunas de las temáticas que actualmente se están abordando. Son:

**Instrumentación de Código:** esta técnica consiste básicamente en la inserción de sentencias de extracción de información dentro del código fuente de los programas, de manera que al ejecutarse se obtenga información respecto del comportamiento del sistema. Para más información leer (Denker, Greevy et al. 2007; Berón, Henriques et al. 2009).

Todas estas técnicas dinámicas proveen grandes cantidades de información, las cuales deben ser analizadas y estudiadas para poder lograr una comprensión concreta de los programas. Por esta razón, el Análisis Dinámico se encarga de crear técnicas de manejo y estudio de la información para poder aprovecharla de la mejor manera. Entre las técnicas existentes se pueden mencionar:

**Técnicas de Interconexión de los Dominios del Problema y del Programa :** utilizan la información provista por las trazas de ejecución (Dominio del Programa) y la vinculan con información del Dominio del Problema para lograr

establecer conexiones entre ambos dominios que permitan llegar a una comprensión más concreta de los sistemas (Lieberman and Fry 1994; Berón, Henriques et al. 2009; Berón, Riesco et al. 2010).

Además, debido a que no se han detectado trabajos que propongan la definición de un marco de medición y comparación de la mantenibilidad de los modelos workflow, nuestra investigación se centra en la evaluación de lenguajes de modelado workflow desde el punto de vista de la mantenibilidad de los modelos definidos con dichos lenguajes. El objetivo es proporcionar a los diseñadores de procesos workflow los criterios necesarios para la selección del lenguaje de modelado más adecuado que facilite la evolución de los modelos desarrollados.

### Resultados Obtenidos/Esperados

Como mencionamos en el apartado anterior, y siguiendo los objetivos de nuestra investigación, hemos definido un conjunto de medidas para la evaluación de modelos workflow. Dichas medidas son una adaptación de las propuestas en (García Rubio 2004) para la medición de procesos software.

Con el fin de validar empíricamente estas medidas, se utilizarán herramientas de comprensión de programas que permitan extraer información de documentos XML, generados desde los modelos workflow, y aplicar las medidas en un análisis de la información obtenida. Dicho análisis se puede llevar a cabo mediante las diferentes herramientas existentes para XML, tales como creadores automáticos de árboles, grafos, entre otras.

Esto nos permitirá realizar un análisis y comparación de modelos, los cuales pueden ser creados en distintas herramientas. Para ello, tomamos un conjunto de modelos provenientes de diversas fuentes. Algunas de dichas fuentes son trabajos de alumnos de posgrado, modelos tomados de prestaciones o

servicios que brinda el proyecto a la comunidad, etc.

Dentro de la misma línea, el grupo de investigación se encuentra abocado a: (a) Definir técnicas para el análisis de trazas de ejecución para optimizar su estudio y comprensión de sistemas de software. (b) Definir estrategias de visualización de trazas de ejecución de manera de facilitar el entendimiento de la información extraída,

### Formación De Recursos Humanos

Basados en la temática planteada, se están desarrollando tesis de Maestría y Doctorado por parte de algunos integrantes del Proyecto. Se ha finalizado una tesis de Especialización en Ingeniería de Software (Peralta 2010). En el marco de la Maestría en Ingeniería de Software que se dicta en la UNSL, dentro del contexto del Proyecto, se han dictado charlas destinadas a los maestrandos acerca de la temática de Modelado Workflow y sobre el trabajo que se está desarrollando al respecto. Además se están desarrollando tesinas de grado para la Licenciatura en Ciencias de la Computación.

### Referencias

Berón, M., P. Henriques, et al. (2009). Inspección de Programas para Interconectar las Vistas Comportamental y Operacional para la Comprensión de Programas. San Luis, Universidad Nacional de San Luis (Argentina) - Universidade do Minh (Portugal).

Berón, M., D. Riesco, et al. (2010). Estrategias para Facilitar la Comprensión de Programas. XII Workshop de Investigadores en Cs de la Computación., El Calafate, Argentina.

Cardoso, J. (2005). "Control-flow Complexity Measurement of Processes and Weyuker's Properties."

Cardoso, J. (2005). How to Measure the Control-flow Complexity of Web Processes and Workflows. Workflow Handbook 2005.

Cardoso, J. (2006). Approaches to Compute Workflows Complexity. Dagstuhl Seminar,



The Role of Business Processes in Service Oriented Architectures, Dagstuhl, Germany.

Debnath, N., M. Peralta, et al. (2011). Metrics for Evaluation of Workflow Models: An Experiment for Validation. 20th International Conference on Software Engineering and Data Engineering (SEDE 2011), Las Vegas - USA.

Denker, M., O. Greevy, et al. (2007). Supporting Feature Analysis with Runtime Annotations. 3<sup>o</sup> International Workshop on Program Comprehension through Dynamic Analysis, Holanda.

García Rubio, F. Ó. (2004). FMESP: Marco de Trabajo Integrado para el Modelado y la Medición de los Procesos Software, U.C.L.M. Universidad de Castilla La Mancha. España: 491.

Georgakopoulos, D. and A. Tsalgatidou (1998). "Technology and Tools for Comprehensive Business Process Lifecycle Management,." Workflow Management Systems and Interoperability. Springer V. p.: 324-365.

Lawrence, P. (1997). Workflow Handbook 1997. New York, Workflow Management Coalition.

Lieberman, H. and C. Fry (1994). Bridging the Gulf Between Code and Behavior in Programming. ACM Conference on Computers and Human Interface, Denver, USA.

Peralta, M. (2010). Los Procesos Workflow y su Modelado. Un Estudio de los Patrones Workflow en distintos Lenguajes de Modelado. Departamento de Informática - Facultad de Ciencias Físico, Matemáticas y Naturales. San Luis - San Luis - Argentina, Universidad Nacional de San Luis.

Peralta, M., F. García, et al. (2007). "Un experimento Comparativo de dos Lenguajes de Modelado Workflow: YAWL vs Diagramas de Actividad." 8th Argentinean Symposium on Software Engineering (ASSE 2007): 145-154.

Peralta, M., F. Ó. Garcia Rubio, et al. (2008). "Un Conjunto de Medidas para la Evaluación

de Modelos Workflow." Congreso Argentino de Ciencias de la Computación (CACIC'08).

Reijers, H. A. (2003). "A Cohesion Metric for the Definition of Activities in a Workflow Process." Eighth CAiSE/IFIP8.1 International Workshop on Evaluation of Modeling Methods in Systems Analysis and Design 2003: 116-125.

Reijers, H. A. and I. T. P. Vanderfeesten (2004). "Cohesion and Coupling Metrics for Workflow Process Design." BPM 2004, LNCS 3080: 290-305.

Rolon, E., F. Ó. Garcia Rubio, et al. (2006). "Validating a Set of Measures for Business Process Models Usability and Maintainability."

Rolon, E., F. Ruiz, et al. (2005). "Aplicación de Métricas Software en la Evaluación de Modelos de Procesos de Negocio." Revista Electrónica de la Sociedad Chilena de Ciencia de la Computación.

Serrano, M., M. Piattini, et al. (2002). Un método para la definición de métricas de software. 1er Workshop en Métodos de Investigación y Fundamentos filosóficos en Ingeniería del Software y Sistemas de Información (MIFISIS'2002),.

Smith, H. (2002). "The Emergence of Business Process Management." CSC's Research Services.

van der Aalst, W. M. P., A. H. M. ter Hofstede, et al. (2002). Workflow Patterns, Queensland University of Technology, Brisbane.

White, S. A. (2004). Process Modeling Notations and Workflow Patterns. Workflow Handbook 2004. L. Fischer, Published in association with the Workflow Management Coalition (WfMC).

Wohed, P., W. M. P. van der Aalst, et al. (2004). "Pattern-based Analysis of UML Activity Diagrams."

Wohed, P., W. M. P. van der Aalst, et al. (2002). Pattern Based Analysis of BPEL4WS, Queensland University of Technology, Brisbane.

Wohed, P., W. M. P. van der Aalst, et al. (2006). On the Suitability of BPMN for Business Process Modelling. 4th International Conference on Business Process Management (BPM 2006), LNCS., Vienna, Austria.

# Comprensión de Especificaciones de Procesos de Negocios Escritas en BPMN

Mario Berón, Carlos Salgado, Mario Peralta, Fernando Saez  
Universidad Nacional de San Luis  
Ejército de los Andes 950  
{mberon,csalgado,mperalta,bfsaez}@unsl.edu.ar

## Resumen

La Comprensión de Programas es una disciplina de la Ingeniería de Software cuyo objetivo es proveer Modelos, Métodos, Técnicas y Herramientas con el propósito de facilitar el estudio y entendimiento de los sistemas de software. Uno de los desafíos más importantes en el contexto de Comprensión de Programas consiste en Relacionar el Dominio del Problema con el Dominio del Programa. El primero hace referencia a la salida del sistema. El segundo a las componentes de software utilizadas para producir dicha salida. Una manera de construir esta relación consiste en definir una representación para cada dominio y luego establecer un procedimiento de vinculación entre ambas representaciones. En este artículo se presenta una línea de investigación que tiene como principal objetivo la elaboración de estrategias para interconectar los dominios del problema y del programa de especificaciones de procesos de negocios escritas en BPMN (Business Process Modeling Notation).

**Palabras Clave:** Comprensión de Programas, Procesos de Negocio, Métricas, BPMN, BPD, XML.

## Contexto

La línea de investigación descrita en este Artículo se encuentra enmarcada en el contexto del proyecto: Ingeniería de Software: Aspectos de alta sensibilidad en el ejercicio de la profesión del Ingeniero de Software de la Universidad Nacional

de San Luis. Dicho proyecto, es reconocido por el programa de incentivos y es la continuación de diferentes proyectos de investigación de gran éxito a nivel nacional e internacional.

## Introducción

La Comprensión de Programas (CP) es una disciplina de la Ingeniería de Software cuyo objetivo es ayudar al programador a entender programas. La CP es útil para Mantenimiento y Evolución de Software (MES), Ingeniería Reversa (IR), Re-Ingeniería (ReI) y Educación. En las tres primeras disciplinas, ayuda a reducir costos y esfuerzos. En la última, asiste al novicio en el proceso de comprensión de algoritmos.

MES es una tarea crítica porque implica tres importantes tareas: perfectiva, correctiva y adaptativa. La primera está relacionada con la incorporación de nuevas funcionalidades. La segunda describe el proceso de eliminar errores de programación. La tercera aborda el problema de adaptar el sistema a nuevos contextos, por ejemplo, usar un nuevo formato de archivo, cambiar la organización del código, etc. Esas tres principales actividades consumen mucho tiempo y recursos. Xi y Hassan [9,10,11] muestran que el 39% de las actividades son perfectivas, 56.7% son correctivas y 2.2% se corresponden con tareas adaptativas. Finalmente, el 2.1% restante representa otras tareas relacionadas con MES.

Por otra parte, la historia del MES muestra un incremento lineal en los costos de proyectos dedicados a MES. Esta información justifica el interés de la comunidad científica por reducir los

esfuerzos dedicados a esas tres principales actividades.

Otra área muy importante es IR. Esta disciplina estudia la creación de estrategias de extracción de la información desde diferentes fuentes (como por ejemplo código fuente, archivos makefiles, documentación, etc.), con el propósito de: Elaborar documentación, Extraer la arquitectura del sistema, etc. IR implica grandes costos y esfuerzos porque generalmente incluye actividades similares a MES. ReI usa técnicas de IR para extraer información y proporciona estrategias para modificar el sistema de acuerdo a los requerimientos del programador. Por esta razón, ReI, de la misma forma que MES e IR, es un área donde se necesitan propuestas para minimizar costos y esfuerzos.

En todas las disciplinas (MES, IR y ReI) mencionadas en los párrafos precedentes, el programador debe entender grandes documentos con diferentes formalismos y metodologías. Además, debe vincular la información recuperada con el código fuente y comprender cómo el sistema lleva a cabo sus funcionalidades en un alto nivel de abstracción. La Comprensión de Programas asiste al programador en esas tres principales áreas. Esta actividad se lleva a cabo por medio de métodos, estrategias y herramientas que disminuyen las tareas operativas y cognitivas del programador [9,10,11,12]. Como un efecto colateral de esta última característica, los costos también se reducen.

Actualmente, existen muchas Herramientas de Comprensión de Programas (HCP) con sofisticadas técnicas de exploración de código. Las mismas funcionan adecuadamente, sin embargo ciertas tareas de comprensión son todavía muy complejas.

Muchas HCP usan solamente análisis estático para la extracción de información del sistema. Después de eso, esa información se representa visualmente con el objetivo de proveer diferentes vistas. De esta forma, el proceso de comprensión de programas se simplifica porque se puede analizar el sistema desde varios puntos de vista. No obstante, el programador está forzado a encontrar el código fuente usado en ciertas

ejecuciones del sistema. Esta última tarea consume tiempo y esfuerzo. Otra clase de herramientas analizan el flujo de ejecución y proveen otro tipo de visualización del sistema. Esas herramientas, conocidas con el nombre de depuradores, frecuentemente proporcionan información de bajo nivel que es útil en estados avanzados de la CP. Por ejemplo, en la comprensión de rutinas específicas del sistema de estudio.

Finalmente, otras herramientas combinan técnicas de recuperación de información estática y dinámica, pero su salida consiste de interesantes y complejos análisis de cada clase de información. Generalmente, las HCP no contemplan algunos aspectos importantes de la CP. Un ejemplo de tales aspectos es la relación entre la salida del sistema y las componentes, a nivel programa, usadas para producirla. Esta relación se conoce como: La Interconexión entre los Dominios del Problema y Programa.

La característica mencionada previamente simplifica la exploración porque el programador solamente inspecciona las partes del sistema relacionadas con la funcionalidad de estudio.

En las últimas décadas los procesos de negocio y su modelado [1, 2] han adquirido una gran popularidad. Esto se debe a que posibilitan que las empresas puedan especificar sus tareas y analizar cómo ellas se llevan a cabo. Esta peculiaridad simplifica el proceso de mejora continua debido a que hace más fácil la evaluación y comprensión de cómo las actividades de la empresa se están llevando a cabo.

Los modelos de procesos de negocio y sus lenguajes de ejecución, al igual que el software convencional, se encuentran sujetos a tareas de mantenimiento. Como fue mencionado en los párrafos precedentes dicha tarea consume mucho tiempo y esfuerzo. Es precisamente en este aspecto en que los métodos, técnicas y estrategias proporcionados por la Comprensión de Programas, pueden ayudar a paliar las dificultades proporcionadas por las actividades de mantenimiento.

El proceso para comprender especificaciones de procesos de negocio es similar al de comprender software convencional. Es decir se deben crear



presentaciones de los dominios del problema y programa y luego intentar relacionar ambas representaciones con el propósito de identificar qué partes de la especificación/implementación llevan a cabo tareas específicas del dominio del problema.

Es claro que proporcionar información objetiva acerca de la mantenibilidad de los modelos, facilita la evolución de los procesos de negocio de aquellas empresas que evalúan constantemente sus procesos al estar involucradas en un proceso de mejora continua. Al mismo tiempo se proporciona soporte a la gestión de procesos de negocio al facilitar la evaluación temprana de ciertas propiedades de calidad de sus modelos, con lo cual las organizaciones se ven beneficiadas de dos maneras: 1) Garantizando el entendimiento y la difusión de los procesos, y su evolución, sin afectar su ejecución; 2) Reduciendo el esfuerzo necesario para cambiar los modelos con la consecuente reducción de los esfuerzos de mantenimiento y/o mejora.

## Línea de Investigación

El grupo de investigación centra su trabajo en desarrollar nuevos métodos, técnicas, herramientas y procesos, que ayuden a los ingenieros de software e informáticos a construir/modificar/reconstruir sistemas de mayor calidad. En este sentido son tres las temáticas que actualmente se están abordando. A continuación se describen cada una de ellas:

- **Construcción de Representaciones del Dominio del Programa:** Una forma de llevar a cabo esta tarea consiste en aplicar técnicas de análisis estático y dinámico a las especificaciones/implementaciones de procesos de negocios. El análisis estático posibilita la construcción de una representación fidedigna del proceso de negocio. Esta representación permite, la definición de métricas y su evaluación concreta sobre la representación antes mencionada. Por estas razones, uno de los intereses de la línea de investigación se centra en la definición/utilización de métricas para la evaluación a nivel conceptual y de implementación de los modelos de procesos de negocio. Esto puede ser un aspecto clave para obtener modelos de calidad que sirvan como soporte para mejorar el mantenimiento y la

adaptabilidad de los procesos de negocio a los cambios constantes a los que dichos procesos están sometidos. En función de esto, se propone implementar un conjunto de métricas definidas para modelos conceptuales de procesos de negocio especificados con BPMN [3]. Dichas métricas se basan en el metamodelo del OMG para definir de forma uniforme cada una de ellas en los diferentes modelos del paquete BPDM (BPDM, por su sigla en inglés de Business Process Definition Metamodel [4]).

El análisis dinámico es tan importante como el análisis estático. Esto se debe a que es posible extraer información relacionada con la ejecución de los procesos de negocio. El análisis estático usa algunas de las técnicas de análisis estático pero con un fin diferente. Dicho fin consiste en incorporar la información necesaria a las especificaciones/implementaciones que permitan transformarlas en otra equivalente pero que realice ciertas tareas adicionales, como por ejemplo: Informar qué partes de ellas fueron utilizadas para llevar a cabo ciertas tareas (las cuales son de interés para la gerencia como para los ingenieros involucrados en el proceso de mejora continua) del dominio de la solución. Este tipo de información, conocida con el nombre de información dinámica, es muy importante porque, además de las características previamente mencionadas, posibilita el cómputo de métricas centradas en el comportamiento en tiempo de ejecución del proceso de negocio. Por otra parte, complementa a la información estática y además puede ser utilizada para decorar la representación estructural del proceso de negocio.

- **Construcción de Representaciones del Dominio del Problema:** El lector puede observar que, hasta este momento, sólo se ha hablado de representaciones del dominio del programa (dominio de especificaciones/implementaciones). No obstante, para facilitar la comprensión de los procesos es necesario construir representaciones del dominio del problema. Es claro que esta tarea no es sencilla, la diversidad de contextos de aplicación

dificulta la elaboración de una representación válida para todos los dominios. No obstante, el grupo de investigación está considerando utilizar formalismos para formalizar dominios específicos. Uno de tales formalismos son las ontologías. Una ontología es una especificación formal de un dominio. Dicha especificación está compuesta por conceptos y relaciones entre ellos. Asociado a cada uno de los componentes antes mencionados se encuentran un conjunto de axiomas que permiten colocar restricciones sobre los conceptos, las relaciones y la ontología en sí. Otras aproximaciones se están analizando (modelos de dominio de UML, especificaciones construidas con métodos formales, etc.) con el objetivo de analizar su aplicabilidad y utilidad para comprender procesos de negocio.

- **Elaboración de Estrategias de Interconexión de Dominios:** Las construcciones de las representaciones fidedignas y útiles de los dos dominios antes mencionados no son suficientes para facilitar la comprensión de los procesos; es necesaria la elaboración de estrategias que permitan vincular ambos dominios. Con respecto a esta temática, el grupo de investigación se encuentra analizando la posibilidad de trasladar estrategias definidas para los lenguajes de programación convencionales al contexto de los procesos de negocio. Las estrategias analizadas hasta el momento se pueden clasificar según el momento en el que se lleve a cabo la vinculación en: Estrategias en Vida y Estrategias Post-mortem. Las primeras muestran la relación entre ambos dominios a medida que se llevan a cabo las tareas del proceso de negocio. Las segundas realizan la misma tarea pero una vez finalizadas las tareas del proceso bajo análisis. También, este tipo de estrategias, se pueden clasificar de acuerdo al tipo de información utilizada para llevar a cabo la relación en: Estáticas, Dinámicas e Híbridas. Básicamente, el equipo de investigación se encuentra dedicado a la elaboración de estrategias que se ubiquen en las clasificaciones antes mencionadas pero que posean las siguientes características: No sean

demasiado complejas de aplicar y el costo computacional y de implementación no sea demasiado alto.

## Resultados y Objetivos

Las tareas realizadas en la línea de investigación objeto de este artículo son las siguientes:

- **Estudio de herramientas de modelado BPMN:** Se ha realizado el estudio de distintas interfaces gráficas de código abierto para el modelado de procesos de negocio en el lenguaje BPMN. Más puntualmente se estudió, aplicando técnicas de Comprensión de Programas, la representación interna de los distintos componentes de los modelos de procesos de negocio, como por ejemplo Tareas, Subprocesos, entre otros. Esto permite tener una trazabilidad de cada uno de los objetos del modelo con su representación en código.
- **Especificación formal de métricas para modelos conceptuales de proceso de negocio:** Se llevó a cabo la especificación utilizando OCL, basados en el estándar de BPD [5], de métricas para modelos de procesos de negocio tomadas de la literatura, como por ejemplo las propuestas en [6, 7]. Para realizar la especificación de dichas métricas se consideraron distintos aspectos que abarca el metamodelo BPD, como por ejemplo: Objetos de Conexión, Carriles y Artefactos; Control de Decisiones de los Objetos; Evento de los Objetos de Flujo, etc.
- **Elaboración e Implementación de Herramientas:** el grupo de investigación se encuentra abocado a la especificación de una herramienta que permita la aplicación de manera automatizada de las métricas al código XPD de los modelos. Esto permitirá un estudio y análisis más confiable de los mismos. Desde esta perspectiva, es necesario tener un medio que permita analizar los códigos generados para los modelos de procesos de negocio y aplicar sobre dichos códigos las métricas propuestas. Esta tarea tiene por finalidad obtener un indicador de la calidad de los modelos estudiados. Desde este punto de vista, y debido a que todo proceso de

desarrollo de Software puede verse como un proceso de negocio, es factible aplicar técnicas de Comprensión de Programas en el análisis del código generado a partir de los modelos antes mencionados. Esto tiene como finalidad determinar la aplicabilidad de las métricas propuestas y así lograr los indicadores de calidad buscados

En la continuidad de este trabajo se prevé: i) Definir nuevas métricas para los modelos conceptuales de procesos de negocio; ii) Definir representaciones de los dominios del problema y programa; iii) Elaborar estrategias para interconectar los dominios antes mencionados y iv) Proponer formas de visualizar la información de cada dominio, como así también la provista por las estrategias de interconexión.

## Formación de Recursos Humanos

Las tareas relacionadas a la presente línea de investigación están siendo realizadas con el objetivo de elaborar tesis de doctorado, maestría, especialización y trabajos de fin de carrera. Se espera a corto plazo poder contar con una o más propuestas de tesis de cada una de las categorías mencionadas previamente. Esta tarea tiene como principal objetivo fomentar el estudio de las temáticas descriptas en esta línea de investigación a través de la captura de alumnos de postgrado y grado que lleven a cabo las tesis antes mencionadas.

## Referencias

- [1] M. A. Rappa, "The utility business model and the future of computing services," IBM Systems Journal, vol. 43, pp. 32-42, 2004.
- [2] J. Becker, M. Rosemann, and C. von Uthmann, "Guidelines of Business Process Modeling," Business Process Management, Models, Techniques and Empirical Studies (BPM'00). Springer, pp. 30-49, 2000.
- [3] OMG, "Business Process Modeling Notation (BPMN)," BPMI - OMG. <http://www.omg.org/spec/BPMN/1.2> 2009.
- [4] OMG, "Business Process Definition MetaModel (BPDM), Beta 1 OMG Adopted Specification," <http://www.omg.org/docs/dtc/07-07-01.pdf>, 2007.
- [5] N. Debnath, C. Salgado, M. Peralta, D. Riesco, and G. Montejano, "Optimization of the Business Process metrics definition according to the BPDM standard and its formal definition in OCL," presented at ACS/IEEE International Conference on Computer Systems and Applications - AICCSA 2010, Tunes, 2010.
- [6] E. Rolon, F. Ruiz, F. Ó. Garcia Rubio, and M. Piattini, "Aplicación de Métricas Software en la Evaluación de Modelos de Procesos de Negocio," Revista Electrónica de la Sociedad Chilena de Ciencia de la Computación, 2005.
- [7] V. Vitolins, "Business Process Measures," presented at Int. Conference on BALTIC DB&IS. Riga, Latvia., 2004.
- [8] OMG, "OMG Object Constraint Language (OCL)," 2012.
- [9] M. Berón, P. Henriques, R. Uzal. Inspección de Programas para Interconectar las Vistas Comportamental y Operacional para la Comprensión de Programas. Tesis Doctoral. Universidad Nacional de San Luis. 2009.
- [10] <http://www3.di.uminho.pt/~gepl/QUIXOTE/>
- [11] E. Oliveira, P. Henriques, M. Pereira. Proposta de um Sistema de Compreensao de Aplicações Web. Tesis de Maestría. Universidade do Minho. 2007.
- [12] S. Rifkin and L. Deimel. Applying Program Comprehension Techniques to Improve Software Inspections. Proceedings of the 19th Annual NASA Software Engineering Laboratory Workshop, Greenbelt, MD, Nov, 1994.

## Análisis y comparación de plataformas BRMS a través de una prueba de concepto

Gonzalo Martin Ibarra<sup>1</sup>, Mg. Patricia Bazán<sup>2</sup>

<sup>1</sup> Facultad de Informática UNLP <sup>2</sup> LINTI Facultad de Informática UNLP

[ibarragm@gmail.com](mailto:ibarragm@gmail.com), [pbaz@ada.info.unlp.edu.ar](mailto:pbaz@ada.info.unlp.edu.ar)

### Resumen

En la última década la implementación de los sistemas BRMS (*Business Rules Management System*) viene creciendo a medida que el negocio requiere más flexibilidad y velocidad a la hora de un cambio de reglas de negocio.

BRMS permite la administración y ejecución de reglas de negocio de distintas formas, de modo que no sea necesario la implementación con “código duro” y que las mismas sean escritas por los usuarios de negocio sin la participación del equipo técnico. Es importante mencionar el tema de costos ya que las implementaciones tradicionales requieren de esfuerzo en diferentes áreas y como consecuencia un mayor costo en tiempo y dinero.

Si bien desde el concepto mencionado en el párrafo anterior se puede inferir como un simple cambio o mejora en la creación e implementación de las reglas de negocio, podemos afirmar que en la realidad no es así ya que hay varios factores o características diferentes en ejecutar un proyecto basado en reglas de negocio que un proyecto de desarrollo de aplicaciones tradicionales.

Pensar en implementar un BRMS implica conocer que la metodología de desarrollo de los proyectos es diferente a la tradicional, que existen diferentes actores y roles para la ejecución de los proyectos, que es necesario un fuerte compromiso desde el área técnica y de negocio de las empresas, también en la adopción de una herramienta de implementación BRMS entre otros temas a destacar.

El área técnica deberá comprender que el ciclo de vida de un proyecto difiere de un proyecto de implementación de una aplicación tradicional y también el área de negocio deberá comprender que cumple un papel fundamental en el éxito del proyecto, siglas como BRE (*Business Rules Engine*), SBVR (*Semantics of Business Vocabulary and Business Rules*), actores como *rule architect*, *rule analyst*, documentos como “*Decision Point Table*”, “*Rule Template*” serán moneda corriente en una metodología de desarrollo BRMS.

Existen diferentes escenarios y también plataformas de aplicación para la ejecución de

reglas de negocio. Inicialmente podemos mencionar escenarios de aplicación en plataforma Java, servidores centrales (mainframes, donde se ejecutan programas COBOL), entornos SOA (*Services Oriented Architecture*) también sobre plataforma .Net.

Entendiendo que BRMS comprende la utilización de plataformas, herramientas y metodologías diferentes, no tan exploradas como las tradicionales y con fuerte indicio de crecimiento sostenido se decidió obtener más conocimiento investigando las diferentes opciones del mercado y además realizando una prueba de concepto con alguna de las plataformas más conocidas del mercado [1] [2] [3][8].

**Palabras clave:** BRMS (*Business Rules Management System*), BRE (*Business Rules Engine*), SBVR (*Semantics of Business Vocabulary and Business Rules*).

### Contexto

El presente es un trabajo de fin de carrera de Licenciatura en Sistemas de la Facultad de Informática de la UNLP, del alumno Gonzalo Martin Ibarra, dirigida por la Mg. Patricia Bazán.

### Introducción

La investigación realizada hasta el momento nos permite abordar y mencionar las características y los aspectos generales de la implementación general de un BRMS, definiendo los ítems a comparar luego de la prueba de concepto.

#### a. Clasificación de las reglas de negocio

Existen diferentes tipos de declaraciones que clasifican a las reglas de negocio, sin embargo no existe una declaración, definición o tipificación estándar de reglas.

Una clasificación general es categorizar las reglas en Términos, Hechos y Reglas. Un **término** es un sustantivo o una proposición con una definición acordada, Un **hecho** es una declaración que conecta términos, a través de preposiciones y frases verbales las cuales forman una regla de negocio y una **regla** es una



declaración que aplica lógica o cálculo de valores de la información.

Una regla se utiliza para el descubrimiento de nueva información o de una decisión acerca de tomar acción [3].

**b. Metodologías de implementación**

Los sistemas tradicionales son abordados por diferentes metodologías como RUP, XP, SCRUM, solo por nombrar algunas de las más conocidas, los cuales presentan diferentes etapas y documentos a elaborar, sin embargo no son parte de este estudio y entendemos que los lectores de este artículo poseen conocimientos sobre las mismas.

Los sistemas BRMS pueden ser aplicados en diferentes escenarios de implementación, es decir, en un sistema ya implementado donde es necesario realizar una reingeniería con el fin de extraer las reglas (*bottom-up*) o bien un sistema implementado obteniendo las reglas de negocio realizando relevamientos con los usuarios de negocio (*top-down*).

A continuación presentamos las diferentes actividades y fases de la metodología ABRD (*Agile Business Rule Development*) las cuales se detallan en la Figura 1.

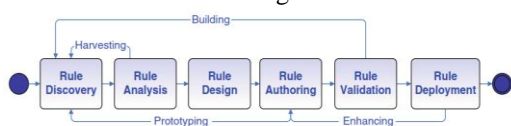


Figura 1. Metodología ABRD [1]

**Rule Discovery:** también llamado *Business Modeling* en la industria, tiene como objetivo desarrollar artefactos simples de modelado como descripciones de reglas de negocio, diagramas de entidad, y los mapas de procesos de negocio.

**Rule Analysis:** el objetivo de la actividad es entender el significado de la regla expresada por la persona de negocios y expertos en la materia y de quitar cualquier ambigüedad y la cuestión semántica o de interpretación.

**Rule Design:** en esta actividad se elaboran diferentes documentos para plasmar las reglas de negocios y comenzar a estrechar la brecha entre el análisis y la implementación de las reglas. Existen herramientas que facilitan el diseño de las reglas.

**Rule Authoring:** es la actividad de creación de las reglas de modo que puedan ser interpretadas por el motor de reglas. Si realizamos una analogía con las metodologías tradicionales sería la actividad de programación de la aplicación.

**Rule Validation:** esta etapa corresponde al test de las reglas en diferentes ambientes y facilidades que presentan las plataformas de BRMS.

**Rule Deployment:** es la actividad de "desplegar" las reglas en los diferentes ambientes definidos por el proyecto [4][10].

**c. Arquitectura**

Existen diferentes plataformas y arquitecturas, se presenta la arquitectura de IBM JRules como muestra la figura 2.

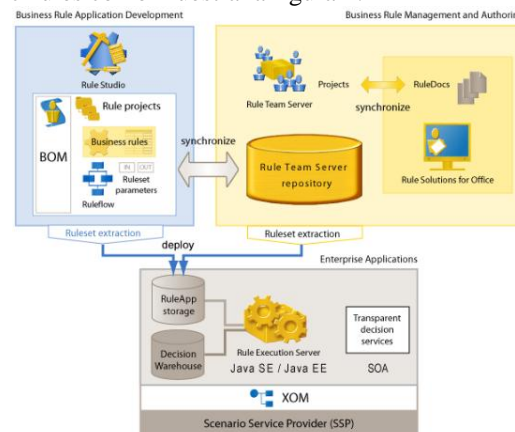


Figura 2. Arquitectura IBM JRules [1]

Es necesario un ambiente de trabajo para crear las reglas, en este caso sería el *Rule Studio* y también necesitamos un componente que interprete o ejecute las reglas, entonces el *Rule Execution Server* nos presta el servicio. Por último, el *Rule Team Server* permite conservar y manipular las reglas ya definidas.

La arquitectura JRules nos brinda dos componentes adicionales, *Decision Validation Services*, que básicamente se utiliza para realizar test y también *Rules Solution for Office* que es una herramienta pensada para facilitar la integración de reglas con componentes Microsoft.

**Rule Studio:** es el entorno de desarrollo para aplicaciones de reglas de negocio. Permite diseñar, crear, revisar y depurar reglas de negocio. Se integra en Eclipse. Los desarrolladores pueden tomar ventaja de esta

integración para desarrollar sus proyectos Java con los proyectos de reglas. *Rule Studio* también dispone de herramientas para mantener las reglas sincronizados con el repositorio de *Rule Team Server*

**Rule Execution Server:** el motor de reglas es un componente utilizado para aislar y aplicar la lógica de negocio (las reglas) de una aplicación

**Rule Team Server:** es un servidor y repositorio que dispone de una herramienta web colaborativa para la creación, gestión, validación e implementación de reglas de negocio. Proporciona un sistema de almacenamiento centralizado de reglas de negocio y sus metadatos.

**Decision Validation Services:** es un módulo que los desarrolladores pueden utilizar para probar y simular las reglas de negocio con respecto a los escenarios, que representan casos reales o ficticios.

**Rule Solutions for Office:** es un entorno de creación de reglas en Microsoft Office. Las reglas pueden ser escritas utilizando Microsoft Word, las tablas de decisión se editan utilizando Microsoft Excel.[1][9]

#### d. Roles

La metodología comentada anteriormente implica la participación de roles especiales que no son parte de una metodología tradicional de desarrollo de aplicaciones. Dichos roles son:

**Arquitecto de Reglas:** es responsable de diseñar la infraestructura necesaria para editar reglas, definir el repositorio de reglas y trabajar en conjunto con el resto del equipo para desarrollar una aplicación basada en reglas.

**Desarrollador:** es responsable del desarrollo, prueba, depuración y despliegue de aplicaciones de reglas de negocio. Interactúa con los modelos de objetos, APIs, y el entorno de desarrollo (Java EE servidores de aplicación o Java SE).

**Analista de Reglas ("Rule Analyst"):** es el equivalente a un "analista de negocios" o un "analista funcional" si lo comparamos con los roles de las metodologías de trabajo tradicionales. Debe poseer un fuerte conocimiento del negocio como así también

estar familiarizado con las aplicaciones basadas en reglas, conocer el funcionamiento de un motor de reglas y el proceso de descubrimiento y análisis de reglas de negocio.

**Escritor de Reglas ("Rule Writer"):** asiste a los analistas de negocio en la creación de las reglas tal como las requieren el motor de reglas.

**Administrador de Reglas ("Rule Administrator"):** gestiona la creación y despliegue de reglas, garantizando la integridad y ciclo de vida de las mismas [1][7].

#### d. Motor de Reglas de Negocio

La utilización de motores de reglas es aplicable en aquellas situaciones en que la lógica requerida (por ejemplo, reglas) es lo suficientemente dinámica para hacer que la inclusión de tal lógica en el código fuente del software resulta poco práctico, por lo tanto es conveniente llevar las reglas a un repositorio de reglas externalizado, es decir, fuera de la aplicación que las utiliza y contar con un intérprete de las mismas.

En el nivel más básico, un motor de reglas se compone de tres componentes, una base de conocimiento (se trata de normas representados en un formato computacional), y un motor de inferencia o ejecución que puede razonar acerca de los datos de entrada basados en el contenido de los conocimientos de base con el fin de generar algún tipo de salida, tal como un conjunto de instrucciones o de alerta, el tercer componente, es la facilidad de la ingeniería del conocimiento, que proporciona la capacidad de conservar el contenido de la base de conocimientos sobre una base de datos

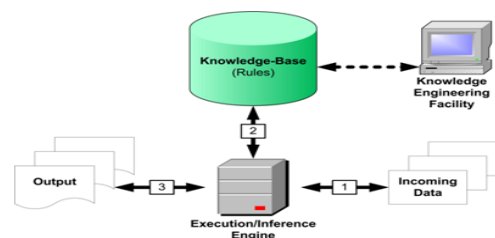


Figura 3 Arquitectura Motor de Reglas [4]

En la figura 3 se refleja la entrada de información (1) la cual es analizada por el intérprete (con los diferentes algoritmos) o motor de inferencia de reglas en conjunto con la base de conocimiento (2) y esta a su vez interactuando con la base de datos de reglas, luego de este proceso se llega a la salida de información (3).

Los algoritmos más conocidos son Rete, Secuencial y Fastpath los cuales poseen

diferentes características y es factible que los motores de reglas los combinen de acuerdo a la regla de negocio a interpretar [1][4][7][8].

#### e. Características

Un BRMS tiene las siguientes características y responsabilidades:

- El almacenamiento y el mantenimiento de un repositorio de reglas de negocio que representan las políticas y procedimientos de una empresa.

- Mantener estas reglas (la lógica de negocio) separados de los componentes de integración necesarios para los sistemas de computación distribuida. Esta característica se puede relacionar a los sistemas tradicionales con enfoque BPM (Business Process Management) en el sentido de utilizar las reglas de negocio en un ambiente distribuido y pensando los procesos desde el punto de vista de negocio

- La integración con las aplicaciones empresariales, de modo que las reglas se pueden utilizar para todos los negocios la toma de decisiones, utilizando los datos normales de negocios.

- Normas que forman en conjuntos de reglas independientes pero con conexión en cadena y las inferencias que realizan dentro de tales conjuntos de reglas

- Permitir que los analistas de negocio e incluso los usuarios para crear, entender y mantener las normas y políticas de la empresa con el mínimo de aprendizaje que requieren. Esta característica es bien diferente a los sistemas tradicionales ya que ubica al usuario en un rol preponderante en todo el ciclo de vida de la aplicación ya que el usuario de negocio participa desde al definición hasta el mantenimiento de las reglas de negocio, no siendo el caso en los sistemas tradicionales donde participa en el análisis y quizás en las pruebas de aceptación.

- La automatización y la facilitación de procesos de negocio.

- Creación de aplicaciones inteligentes que interactúan con los usuarios a través de diálogos naturales, comprensible y lógica, Esta característica quizás es el ideal de los sistemas orientados a reglas; para lograr el objetivo se requiere de madurez de las reglas como así también un fuerte conocimiento del negocio por parte de los usuarios[1][2][9].

#### LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Esta línea de investigación realiza un análisis profundo de las características de las plataformas y herramientas involucradas en el

desarrollo de proyectos basados en la administración de reglas de negocio. Para lograr dicho objetivo se seleccionaron dos plataformas:

- *WebSphere ILOG JRules*
- *Drools*

Para poder realizar la comparación de las características se elaboró una prueba de concepto que consiste básicamente en la implementación de requerimientos de negocio a través de ambas herramientas.

Es válido mencionar que para replicar o simular el proceso completo de la creación de una aplicación BRMS es necesario contar con diferentes personas, documentos, una empresa real que tenga una necesidad de negocio real y el tiempo suficiente para desarrollar el proyecto, desde el descubrimiento de las reglas de negocio hasta la implementación de las misma [2][9].

#### RESULTADOS Y OBJETIVOS

De acuerdo a la disponibilidad de recursos y objetivo de la prueba de concepto se decidió utilizar un documento o "Fuente original" elaborado de modo que sea posible investigar diferentes características de las plataformas, para que de esta forma se pueda dar inicio al proceso de desarrollo y que este sea el disparador para la obtención de las reglas de negocio, asimismo para poder desarrollar el caso de prueba fue necesaria la participación en distintos roles los cuales fueron anteriormente descriptos .

Luego de la aplicación de la prueba de concepto se realizará la comparación de la plataforma utilizando la Tabla 1.

Gran parte de las características a comparar fueron obtenidas de referencias mencionadas sobre este artículo e incluyendo algunos aspectos no mencionados.

La tabla 1 fue dividida en 4 secciones, Atributos Generales donde se incluye el desarrollo de los BRMS en Argentina midiendo su base instalada y recursos profesionales además de comparar precios y facilidades de instalación y soporte. Se incluye el IDE (*Integrated Development Environment*) el cual nos permite crear, diseñar y mantener reglas de negocio, a mi entender es muy importante ya que una interface amigable nos permite suavizar el cambio en la forma de pensar en ejecutar proyecto y alivia la "resistencia al cambio"; luego la sección *rule engine*, en donde incluimos los tres algoritmos de inferencia y también el soporte de *debugging*, muy importante también para dar herramientas que faciliten el testeado de reglas (recordemos que estamos ante un escenario diferente de pruebas del sistema tradicional), por último la sección de integración, importantísimo desde el punto de

vista de SOA, BPM y la necesidad de las grandes empresas de integrarse con diferentes sistemas [5][6].

**Tabla 1 Comparación BRMS [5]**

Atributos Generales	
Precio	
Facilidad de Instalación	
Soporte Técnico	
Servicios Profesionales disponibles (Argentina)	
Metodología de implementación	
Adherencia a la industria	
Soporte al ciclo completo de implementación	
Base instalada en Argentina	
IDE(Integrated Developed Environment)/Templates	
Disponibilidad de templates/ejemplos a reutilizar	
Soporte para formato de expresiones	
Soporte para tablas de decisión	
Referencia cruzada de reglas	
Control de versiones	
Chequeo de ambigüedad	
Rule Engine	
Soporte a RETE	
Soporte al Algoritmo Secuencial	
Soporte a Fastpath	
Soporte para debugging	
Despliegue de reglas durante la ejecución ("hot deploy")	
Capacidad para administrar un alto volumen de reglas	
Facilidad de administración y estructuración de reglas	
Capacidad para planificar deploys	
Performance / Escalabilidad	
Integración	
Facilidad de integración con herramientas de Middleware	
Integración con Webservices	
Integración con ambientes J2EE	
Integración con COBOL	
Integración con .Net	
Integración con BPM	

## FORMACIÓN DE RECURSOS HUMANOS

BRMS es una nueva forma de abordar los proyectos que requiere un ciclo de vida diferente a la implementación tradicional de las aplicaciones. Es necesario también que las implementaciones se realicen en empresas que se encuentren preparadas tecnológicamente y que dispongan de los recursos humanos adecuados para desarrollar esta forma de trabajar. Entendemos que los proyectos de BRMS no están difundidos en la Argentina, sin embargo si se puede detectar un fuerte crecimiento en la industria norteamericana de software. Conociendo esta situación real se decidió colaborar y aportar un granito de arena en la difusión de esta alternativa de implementación de aplicaciones.

## REFERENCIAS

- [1] Jerome Boyer, Hafedh Mili, "Agile Business Rule Development".
- [2] Stephen D. Hendrick, Kathleen E. Hendrick, "The Business Value of Business Rule Management Systems".
- [3] Barbara Von Halle " Business Rules Applied".
- [4] Joel Saltz, Joyce Niland, Philip Payne, Hemant Shah ,Douglas Stahl "Rules Engine Technologies Across caBIG Workspaces "
- [5] Ian Graham, " Service Oriented Business Rules Management Systems".
- [6] Daniele Di Bona, Giuseppe Lo Re, Giovanni Aiello, Adriano Tamburo, Marco Alessi, "A Methodology for Graphical Modeling of Business Rules".
- [7] Xiaoming Feng, Mani Subramanian, "Incorporating Business Rule Engine Technology in Control Center Applications"
- [8] Senlin Liang Paul Fodor Hui Wan Michael Kifer, "OpenRuleBench: An Analysis of the Performance of Rule Engines".
- [9] Saqib Ali, Ben Soh, Torab Torabi "Using Software Engineering Principles to Develop Reusable Business Rules".
- [10] Matthew L. Nelson, Robert L. Rariden, Ravi Sen, "A Lifecycle Approach towards Business Rules Management".



## Automatización de la evaluación de la Usabilidad del Software

Mascheroni, M.<sup>(1)</sup>; Greiner, C.<sup>(1)</sup>; Dapozo, G.<sup>(1)</sup>; Estayno M.<sup>(2)</sup>

(1)Departamento de Informática. Facultad de Ciencias Exactas y Naturales y Agrimensura  
Universidad Nacional del Nordeste

agustin.mascheroni@hotmail.com, {cgreiner, gndapozo}@exa.unne.edu.ar

(2)Departamento de Informática. Facultad de Ingeniería. Universidad Nacional de Lomas de Zamora  
mestayno@gmail.com

### Resumen

La usabilidad es una cualidad que se considera cada vez más importante para la calidad de los productos software. Es un atributo intangible del mismo, por lo tanto, resulta difícil de visualizar, medir y reconocer como un factor determinante de su calidad. En esta línea de trabajo se propone el desarrollo de una herramienta que facilite la evaluación de la usabilidad durante el proceso de desarrollo, basada en la aplicación de los estándares vigentes y criterios heurísticos sobre usabilidad. La herramienta formará parte de una metodología de medición cuyo objetivo es facilitar en el contexto de desarrollo de las pymes la implementación y evaluación de la usabilidad del software.

**Palabras clave:** Calidad de software, Usabilidad, Herramienta de Evaluación.

### Contexto

La línea de I/D presentada en este trabajo forma parte del proyecto F007-2009: "Modelos y métricas para la evaluación de la calidad de software", acreditado por la Secretaría General de Ciencia y Técnica de la Universidad Nacional del Nordeste (UNNE), cuya Unidad Ejecutora está integrada por docentes de la Universidad Tecnológica Nacional Regional Resistencia (UTN-FRRe), de la Facultad de Ciencias Exactas de la UNNE y de la Universidad Nacional de Lomas de Zamora (UNLZ).

El objetivo fundamental del proyecto es contribuir a la mejora en la calidad de los productos software mediante modelos y métricas aplicados al producto y al proceso de creación, diseño, desarrollo y mantenimiento de software, como medio para aumentar la competitividad de las pymes de la región NEA en el contexto de la industria del software.

### Introducción

La usabilidad es un atributo del software que indica que tan fácil es usar y comprender un determinado sistema o producto y qué tan eficiente es para realizar las funciones que debe realizar. Por lo tanto, cabe plantearse preguntas como: ¿por qué

hay que preocuparse por la usabilidad? ¿Por qué las cosas son difíciles de usar?

Un aspecto del problema, podría ser que en el desarrollo del producto se hace mayor énfasis en la tecnología, en lugar de centrarse en el usuario, la persona para la cual está hecho el sistema [1].

La usabilidad interviene también en la calidad, dado que para evaluar la calidad del software es preciso comprender el propósito para el cual el sistema se va a usar. Por tanto, la calidad no es una medida del software aislada, sino que es una medida que tiene en cuenta la relación entre el producto y su dominio de aplicación [2]. El usuario es una parte esencial de tal dominio de aplicación, por lo que la usabilidad es un componente básico de la calidad del software. Esta visión de la usabilidad como atributo de calidad no es novedosa, puesto que, a pesar de que no existe un conjunto aceptado de atributos críticos de calidad de software, la usabilidad ha estado presente como atributo en descomposiciones de calidad del software desde los años 70 [3] [4]. Por lo tanto, la usabilidad se reconoce como atributo de calidad del software.

De todas formas, que se reconozca explícitamente la usabilidad como atributo de calidad, no significa que automáticamente las prácticas de la Ingeniería de Software traten adecuadamente todo lo relacionado con la usabilidad del producto software que se está desarrollando. Por el contrario, observando las técnicas de la Ingeniería de Software que forman parte de los distintos procesos de desarrollo existentes, se pueden identificar diferencias en el modo en que son tratados distintos atributos de calidad de software [2].

Podría decirse que el cuidado por la usabilidad no está presente en las técnicas de la Ingeniería de Software más que tangencialmente, y, así, su presencia en los procesos de desarrollo es marginal. En la mayor parte de los casos no existe o está relegada a una única actividad de diseño de la Interfaz de Usuario, sin existir una preocupación por la usabilidad a lo largo de todo el proceso de desarrollo, como sí existe para otros atributos de calidad [2] [4].

## Usabilidad

Coloquialmente, se suele definir a la usabilidad como la propiedad que tiene un determinado producto para que sea “fácil de usar o de utilizar y de aprender”, tratándose de una propiedad que no es tan sólo aplicable a los sistemas de software, sino que es aplicable a otros objetos de la vida cotidiana [5]. Para el estándar ISO 9241, que trata los requerimientos ergonómicos para trabajos de oficina con visualización, la usabilidad es “el grado en el que un producto puede ser utilizado por usuarios específicos para conseguir objetivos específicos con efectividad, eficiencia y satisfacción en un determinado contexto de uso” [6].

### Atributos de Usabilidad

Para identificar y evaluar la usabilidad, esta se descompone en los siguientes atributos:

- **Facilidad de Aprendizaje:** Indica qué tan fácil es aprender la funcionalidad básica del sistema, como para ser capaz de realizar precisamente las tareas que desea realizar el usuario [7].
- **Eficiencia:** La eficiencia se determina por el número de transacciones por unidad de tiempo que el usuario puede realizar usando el sistema [8].
- **Presentación visual apropiada:** El concepto de sistema se materializa al realizar el diseño de la parte visual de la interacción, es decir, la “interfaz gráfica de usuario”. Hay una serie de normas provenientes del campo del diseño gráfico sobre cómo escoger los colores, tipo de letra, la disposición de los elementos, etc. [9].
- **Manejo de Errores:** Indica cómo el sistema previene los errores que el usuario puede cometer mientras se encuentra operándolo [10].
- **Satisfacción:** Indica la impresión subjetiva que el operador del sistema obtiene del mismo. Para ello se utilizan cuestionarios, encuestas, entrevistas [9].
- **Nivel de seguridad:** La calidad no puede existir sin seguridad [11]. La seguridad es un factor importante para determinar la usabilidad de una aplicación, puesto que los usuarios se sentirán más confiados y protegidos utilizando software seguro.

### Importancia de la usabilidad

“El gran avance en la tecnología de los ordenadores ha incrementado la potencia de éstos a la vez que ha realizado el ancho de banda de comunicación entre las personas y los ordenadores. Aun así, el impacto de la tecnología no es suficiente para realzar la usabilidad. Los principios para la interacción son independientes de la tecnología. Dependen mucho más de un

conocimiento más profundo de los elementos humanos de dicha interacción” [12].

La interfaz del sistema es la puerta del usuario a la funcionalidad del sistema y, por tanto, el hecho de que estas interfaces estén mal diseñadas es un factor que limita el uso de las funcionalidades [5]. Por esta razón, es muy importante diseñar interfaces de usuario usables.

Estudios realizados por diferentes investigadores que consideran la importancia de la usabilidad del software, constatan de una manera real los beneficios que esto supone [5] [7] [8], entre los que se destacan:

- Mejora la calidad del producto
- Reduce los costos de producción
- Reduce los costos de mantenimiento y apoyo
- Reduce los errores cometidos por los usuarios
- Aumenta la satisfacción del usuario
- Permite realizar las tareas de manera más rápida y sin pérdida de tiempo
- Desarrollar un sistema que esté diseñado como una herramienta que encaja con la forma de trabajar del usuario y que le permita realizar aquello que éste pretende hacer lo mejor posible.

### Medición de la Usabilidad

La medición de la usabilidad de una aplicación es un proceso que lleva tiempo y en muchos casos resulta muy costosa. Por ello, la mayoría de los desarrolladores de software no la abordan con la profundidad requerida.

Existen dos métodos para llevar a cabo la evaluación de la usabilidad [8] [13]:

- **Métodos heurísticos (no empíricos):** implican la participación de expertos especialistas en usabilidad.
- **Métodos empíricos:** pruebas de análisis que requieren la participación de usuarios.

### Líneas de investigación y desarrollo

La presente línea de investigación está orientada a profundizar en:

1. Desarrollo de una herramienta para la evaluación de la usabilidad durante el proceso de desarrollo de software.
2. Métodos para la evaluación de la usabilidad. Técnicas de Ingeniería de usabilidad. Criterios heurísticos para la evaluación de la usabilidad. Evaluación de la usabilidad mediante métodos de Inspección. Estándares de usabilidad vigentes. Metodologías que propician la incorporación de la Ingeniería de Usabilidad en la Ingeniería del Software.

3. Relevamiento de técnicas de evaluación de usabilidad utilizadas en empresas pymes de la región.
4. Análisis y comparación de las herramientas de usabilidad existentes en el mercado.
5. Estudio en profundidad de patrones de diseño como base para la búsqueda de soluciones a problemas comunes en el desarrollo de la herramienta. Buenas prácticas de desarrollo de software.

## Resultados y Objetivos

Para apoyar las acciones del proyecto, se realizó una investigación bibliográfica exploratoria vinculada a los conceptos de la Ingeniería de Usabilidad, y luego, se recabó información para determinar qué grado de importancia le conceden a la usabilidad las empresas que desarrollan software en la región.

Para este segundo objetivo, basado en los conceptos de la investigación bibliográfica, se diseñó un instrumento de recolección de información sobre determinados aspectos de usabilidad, a modo de guía orientativa, que se utilizó en entrevistas realizadas a los responsables del proceso de desarrollo de algunas pymes software de las ciudades de Resistencia y Corrientes.

El diseño de la guía orientativa se basó en dos aspectos principales: El perfil del usuario y las técnicas de usabilidad que se utilizan. Dentro de estos, las variables consideradas fueron: interacción con los usuarios en las distintas etapas de desarrollo de software, formas y técnicas de interacción con los usuarios, participación del usuario en el diseño de la interfaz, técnicas de prototipado utilizadas, y evaluación del producto con usuarios.

Con el instrumento de recolección de información diseñado se realizaron las entrevistas en empresas de la ciudad de Corrientes y de la ciudad de Resistencia. Los resultados permitieron concluir que las empresas no desconocen la importancia de la usabilidad, e incorporan algunas técnicas específicas [14]. Sin embargo, el grado de participación de los usuarios en las etapas de diseño y evaluación es bajo. Esta situación podría generar productos software con déficits de usabilidad.

Por esta razón, se propuso diseñar una herramienta que permita automatizar algunos aspectos de la evaluación de la usabilidad en productos software, y así mejorar la calidad de los mismos.

La herramienta se compone de dos partes: una que permite comprobar el cumplimiento de los estándares de usabilidad y la aplicación de los

criterios heurísticos, mediante formularios específicos, y otra que presenta un cuestionario de percepción del usuario, que será completado en base a las opiniones de los usuarios.

Para la primera parte se hizo un estudio en profundidad de las normas ISO, para tomar los puntos más fuertes a la hora de evaluar la usabilidad de una aplicación. Esto se complementó con los criterios heurísticos tomados de diferentes autores [15]. De esta manera, los elementos que se seleccionaron para ser evaluados por medio de la observación de un experto, el que completará el formulario, son:

- Ventanas y Botones
- Organización de la información (listas, tablas, etiquetas, etc.)
- Combinación de colores
- Fuente (tamaño y tipo)
- Elementos propios de sitios web

Cada uno de estos elementos, se evalúan individualmente siguiendo los criterios heurísticos y los estándares de usabilidad. Para hacerlo, la herramienta brinda al evaluador tres valores de conformidad en base al cumplimiento del criterio o norma:

- “Si, totalmente” = 1
- “Si, parcialmente” = 0,5
- “No” = 0

Para la elaboración del cuestionario de percepción del usuario, que forma la segunda parte de la herramienta, se hizo un estudio exploratorio de cómo un conjunto de usuarios de diferentes perfiles (novatos, intermedios y expertos) realizan una serie de tareas en una aplicación determinada y así tener en cuenta sus opiniones. Con esto se elaboró un cuestionario que permite recabar la percepción del usuario, cuyo detalle se describe en [16].

Para evaluar los puntos del cuestionario de percepción del usuario, la herramienta plantea una escala de conformidad del 1 al 5, siendo 1 el índice menos favorable. A cada valor de la escala se le asigna un peso que representa la graduación de la percepción manifestada por el usuario. El cumplimiento de cada atributo se expresa porcentualmente, es decir, el cumplimiento óptimo del atributo será 100% [16].

El proceso de evaluación de un producto software con la herramienta se basa en la intervención de un experto y un conjunto de usuarios. Para la primera parte, el evaluador irá seleccionando un valor para cada ítem del formulario y luego guardará los resultados. Para la segunda, los usuarios deberán llevar a cabo una serie de tareas establecidas y el evaluador por medio de la observación e

interrogación, irá completando un cuestionario de percepción del usuario por cada uno.

Al finalizar, la herramienta analiza los datos y genera un informe con los resultados obtenidos.

Actualmente se trabaja en la elaboración de una metodología de medición de usabilidad, que incorpora la herramienta mencionada, y se aplicará en el desarrollo de productos en ambientes reales de desarrollo como las pymes de la región.

## Formación de Recursos Humanos

Esta línea de trabajo se realiza en el marco de una beca de investigación correspondiente al programa de Becas de Estímulo a las Vocaciones Científicas, financiadas por el Consejo Interuniversitario Nacional (CIN), como parte del Plan de Fortalecimiento de la Investigación Científica, el Desarrollo Tecnológico y la Innovación en las Universidades Nacionales, bajo la supervisión de docentes investigadores del proyecto mencionado.

## Referencias

- [1] Norman, D. "The Invisible Computer: Why good Products can fail, the Personal computer is so complex, and information appliances are the solution". Editorial MIT Press Cambridge. Massachusetts, USA. 1998. ISBN: 978-0-262-14065-2.
- [2] Ferré, X. "Marco de Integración de la usabilidad en el proceso de desarrollo software". Tesis Doctoral. Departamento de Lenguajes y Sistemas Informáticos e Ingeniería del Software. Facultad de Informática. Universidad Politécnica de Madrid. Madrid, España. 2005
- [3] McCall, J. A.; Richards, P. K.; Walters, G. F. "Factors in Software Quality". 1977. Vols I, II, III. NTIS AD-A049-014, 015, 055.
- [4] Bohem, B. "Characteristics of Software Quality". North Holland Publishing Co. 1978
- [5] Granollers, T.; Lorés, V.; Cañas, J.J. "Diseño de sistemas interactivos centrados en el usuario". Editorial UOC (Universitat Oberta de Catalunya). Barcelona, España. 2011. ISBN: 9788497882675
- [6] International Organization for Standardization ISO 14598-1. Information Technology - Evaluation of Software Products - General Guide. ISO 1998.
- [7] Ferré, X. "Principios básicos de usabilidad para ingenieros software". 2000. Disponible en <http://is.ls.fi.upm.es/miembros/xavier/papers/usabilidad.pdf>
- [8] Nielsen, J. "Usability Engineering". AP Professional, 1993
- [9] Mascheroni, M.; Greiner, C.; Petris, R.; Dapozo, G.; Estayno, M. "Calidad de Software e Ingeniería de Usabilidad". XIV Workshop de Investigadores en Ciencias de la Computación. Misiones, Argentina. 2012
- [10] [Apple Computer. "Human interface guidelines: The Apple Desktop Interface". Addison-Wesley. New York, United States. 1987
- [11] International Organization for Standardization ISO 9126: Software Engineering – Product quality, Geneva, Switzerland. ISO 2001
- [12] Dix, A.; Finlay, J.; Abowd, G.; Beale, R. "Human-Computer Interaction". Editorial Prentice Hall. Londres, UK. 1993. ISBN 0-13-437211-5.
- [13] Rubin, J.; Chisnell, D. "*Handbook of usability testing: how to plan, design, and conduct effective tests*". Wiley Technical Communications, 2008. 2nd ed., chapter 1.
- [14] Mascheroni, M.; Greiner, C.; Petris, R. "Técnicas de usabilidad. Estudio exploratorio sobre su incorporación en los procesos de desarrollo de software en pymes locales". II Jornadas de Investigación en Ingeniería del NEA y países limítrofes". UTN-Facultad Regional Resistencia. Chaco, Argentina. 2012
- [15] Nielsen, J.; Molich, R. "Heuristic evaluation of user interfaces". Proc. ACM CHI'90 Conf. Seattle, WA, 1-5 April 1990, 249-256.
- [16] Mascheroni, M.; Greiner, C.; Dapozo, G.; Estayno, M. "Herramienta para automatizar la evaluación de la usabilidad en productos software". XVII Congreso Argentino de Ciencias de la Computación. Bahía Blanca, Argentina. 2012



## Procesos de Ingeniería de Requerimientos basados en modelos de madurez orientados a la certificación ISO 9001

Mariana Cecilia Arruzazabala<sup>1</sup>, Gladys Dapozo<sup>1</sup>, Pablo Thomas<sup>2</sup>

<sup>1</sup> Departamento de Informática. Facultad de Ciencias Exactas y Naturales y Agrimensura  
Universidad Nacional del Nordeste, Av.Libertad 5450, 3400, Corrientes, Argentina  
arruzazabala@gmail.com, gndapozo@exa.unne.edu.ar

<sup>2</sup> Instituto de Investigación en Informática LIDI - Facultad de Informática. Universidad Nacional de La Plata, Calles 50 y  
120, La Plata, Buenos Aires, Argentina  
pthomas@lidi.info.unlp.edu.ar

### Resumen

En el marco de la industria del software en la Argentina la certificación de la calidad de software contribuye a la competitividad del sector productivo. La Ingeniería de Requerimientos (IR) es una parte fundamental del proceso de desarrollo de software, dado que pobres prácticas en esta etapa afectan significativamente el éxito de un proyecto. En esta línea de trabajo interesa analizar en qué medida la norma ISO 9001:2008 considera las buenas prácticas de la IR, para ello se estudiaron modelos de madurez de IR y se realizó una evaluación de proceso de IR antes y después de la certificación de la mencionada norma de calidad en las empresas del Polo IT Corrientes. Se concluye que la norma mejora el proceso de IR en ciertas áreas, pero no en todas. Actualmente, en base a los resultados obtenidos, se trabaja en la definición de procesos de IR que cumplan los requisitos de la norma ISO 9001:2008 y aseguren niveles altos de madurez de IR, basado en el modelo Uni-REPM.

**Palabras clave:** Ingeniería de Requerimientos. Certificación de Calidad. Madurez en la Ingeniería de Requerimientos. ISO 9001:2008.

### Contexto

La línea de I/D presentada se desarrolla en el marco del proyecto F007-2009: "Modelos y métricas para la evaluación de la calidad de software", acreditado por la Secretaría General de Ciencia y Técnica de la Universidad Nacional del Nordeste (UNNE).

### Introducción

En el marco de la joven y dinámica industria del software en la Argentina, se promueve la certificación de calidad en el proceso de desarrollo de software como una forma de contribuir a incrementar la competitividad del sector productivo [1].

La Ley 25.922 de Promoción de la Industria del Software otorga beneficios impositivos a quienes la aplican, y tiene entre sus requisitos la certificación de calidad [2]. Entre las normas aplicables por esta

Ley se encuentran las que certifican procesos CMM, CMMI, ISO 9001, ISO/IEC 90003, IRAM 17601 (CMMI - SEI), ISO/IEC 15504, y la que certifica productos ISO/IEC 9126 [3].

Particularmente en la ciudad de Corrientes, Argentina, las empresas nucleadas en el Polo IT Corrientes, realizaron un proceso de implementación de la norma ISO 9001:2008 durante 12 meses [4].

Dentro de los procesos de desarrollo de software, la Ingeniería de Requerimientos (IR) es particularmente crítica debido a que los errores que se presentan en esta etapa originan inevitablemente problemas posteriores que afectan a todo el ciclo de vida del software [5] [6].

En los últimos años el ritmo de desarrollo de los productos se ha acelerado drásticamente, el cambio tecnológico y la rotación de personal han impactado en los niveles de experiencia de los profesionales, y el *outsourcing* y *offshoring* han cambiado el ciclo de vida del producto. Debido a esto las especificaciones de requerimientos deben ser precisas y detalladas [7]. Adicionalmente, se sabe que los requerimientos deficientes son la principal causa de fracaso de los proyectos de software [8]; y cuanto más tarde se descubre un error introducido en la etapa de requerimientos, mayor será el costo de su reparación [9].

Por todo lo expuesto, se considera relevante analizar el impacto que la certificación de la norma ISO 9001:2008 tuvo en las prácticas de IR, para lo cual se analiza el caso de las empresas del Polo IT Corrientes que certificaron dicha norma.

### Uni-REPM (Unified Requirements Engineering Process Maturity Model)

Uni-REPM es un modelo que evalúa la madurez del proceso de IR de un proyecto a través de un conjunto de buenas prácticas [11]. La evaluación consiste en mapear las actividades ideales con el trabajo real de una organización.

Las actividades del modelo se dividen en 7 áreas: Apoyo de la Organización, Proceso de Gestión de Requerimientos, Elicitación de Requerimientos, Análisis de Requerimientos, Planificación de la Entrega, Especificación de Requerimientos y

Documentación, y Validación de Requerimientos.

Cada área tiene subáreas, y cada subárea tiene acciones, que son las unidades más pequeñas de evaluación. Una acción, al ser evaluada, puede tener tres posibles estados: completa, incompleta o inaplicable. Una acción completa significa que la organización la tiene implementada, una acción incompleta significa que corresponde que la organización la implemente pero no lo hace, y una acción inaplicable es aquella que no corresponde ser evaluada en el marco de una organización específica, porque no es una actividad esencial en el negocio. En este último caso, por ejemplo, si sólo se desarrolla productos a medida, no correspondería evaluar las acciones propias de productos orientados al mercado. Se debe prestar especial atención con la asignación del valor 'inaplicable' a las acciones, ya que se podría sesgar el resultado si se la usa inapropiadamente.

El modelo propone tres niveles de madurez como resultado de la evaluación: 1-Básico, 2-Intermedio, y 3-Avanzado. El objetivo del primer nivel es asegurar que se definen procesos, y que éstos son aplicados por la organización, los *stakeholders* relevantes son involucrados en la elicitación, los requerimientos se analizan en profundidad, y se predefinen documentos estándares. El nivel Intermedio trata diferentes perspectivas y está dirigido por las estrategias de producto/objetivos, las funciones y responsabilidades están bien definidas y documentadas, las solicitudes de cambio se gestionan consistentemente a lo largo de todo el proyecto y la selección de requerimientos se realiza de manera sistemática. El nivel Avanzado representa el proceso de IR con mayor madurez, la mejora de proceso se ve en la forma avanzada de obtener los requerimientos, en asegurar su calidad, mantener las comunicaciones y la comprensión común entre los diferentes *stakeholders*, y evaluar proactivamente el proceso de toma de decisiones.

Para acceder a un nivel de madurez determinado se deben cumplir todas las acciones asociadas a dicho nivel, más las acciones de los niveles anteriores. Para un análisis más profundo se pueden analizar los niveles de madurez por área y subárea, en cuyo caso un área o subárea llegaría a un determinado nivel si se cumplen todas las acciones de ese nivel más el del nivel inmediato anterior. No todas las subáreas tienen acciones de los tres niveles, existen algunas cuyas acciones tienen como mínimo nivel dos, y otras cuyas acciones tienen como máximo nivel dos.

La creación de este modelo se basó en una revisión de literatura de los modelos REPM, CMMI e ISO 9000-TickIT y en una revisión sistemática de investigaciones de prácticas de IR orientadas al mercado. El modelo Uni-REPM fue validado en el contexto académico e industrial [12],[13].

### **Norma ISO 9001:2008**

La norma ISO 9001 especifica los requisitos para un sistema de gestión de la calidad, cuando una organización necesita demostrar su capacidad para proporcionar regularmente productos que satisfagan los requisitos del cliente, los legales y reglamentarios aplicables, y aspira a aumentar la satisfacción del cliente a través de la aplicación eficaz del sistema, incluidos los procesos para la mejora continua del sistema y el aseguramiento de la conformidad con los requisitos del cliente, los legales y reglamentarios aplicables.

Todos los requisitos de esta Norma Internacional son genéricos y se pretende que sean aplicables a todas las organizaciones sin importar su tipo, tamaño y producto suministrado.

Para verificar que se cumplen los requisitos de la norma, existen entidades de certificación que auditan la implantación y mantenimiento, emitiendo un certificado de conformidad. Estas entidades están reguladas por organismos nacionales. En Argentina, el organismo que regula la actividad es el Instituto Argentino de Normalización y Certificación (IRAM).

### **Enfoque de procesos**

Esta norma promueve la adopción de un enfoque basado en procesos cuando se desarrolla, implementa y mejora la eficacia de un sistema de gestión de calidad, para aumentar la satisfacción del cliente a través del cumplimiento de sus requisitos.

Un proceso puede ser considerado como una actividad o un conjunto de actividades que utiliza recursos, y que se gestiona con el objeto de que los elementos de entrada se transformen en resultados.

Una organización que funciona de manera eficaz tiene que determinar y gestionar numerosas actividades relacionadas entre sí.

Un *enfoque basado en procesos* es el nombre que se le puede dar a la aplicación de un sistema de procesos dentro de la organización, junto con la identificación e interacciones de tales procesos, así como su gestión para producir el resultado deseado.

Un enfoque de este tipo, cuando se utiliza dentro de un sistema de gestión de la calidad, enfatiza la importancia de: la comprensión y el cumplimiento de los requisitos, la necesidad de considerar los procesos en términos que aporten valor, la obtención de resultados del desempeño y eficacia del proceso, y la mejora continua de los procesos con base en mediciones objetivas.

Los clientes juegan un papel significativo para definir los requisitos como elementos de entrada. El seguimiento de la satisfacción del cliente requiere la evaluación de la información relativa a la percepción del cliente acerca de si la organización ha cumplido sus requisitos.

## Líneas de investigación y desarrollo

La presente línea de investigación está orientada principalmente en:

1. Estudio y evaluación de modelos de madurez aplicados a procesos de Ingeniería de Requerimientos.
2. Caracterización de las empresas de software nucleadas en el Polo IT Corrientes, en cuanto a tipo de actividad, recursos humanos, proceso definidos, entre otros.
3. Evaluación de proyectos software realizados antes y después de la certificación ISO 9001. en cuanto a las mejoras logradas luego de los procesos de certificación ISO 9001:2008.
4. Definición de procesos de desarrollo que cumplan con las recomendaciones de los modelos de madurez de la Ingeniería de Requerimientos, con vistas a procesos de certificación ISO 9001.

## Resultados y Objetivos

Para cumplir con los objetivos de la investigación, se propusieron las siguientes actividades:

- Análisis comparativo de modelos de evaluación de procesos de IR y selección de un modelo de referencia.
- Diseño e implementación de instrumentos de recolección de información para la caracterización de las empresas de software seleccionadas para la investigación.
- Evaluación de proyectos de software pre y post certificación en las empresas seleccionadas, aplicando el modelo Uni-REPM.

## Análisis comparativo de modelos de evaluación de procesos de IR

Para encontrar el modelo que mejor se adecúe a la evaluación de la madurez del proceso de IR se realizó una investigación documental exploratoria vinculada a los conceptos de IR. Los criterios de búsqueda privilegiaron los términos "Requirements Engineering Process Maturity Models", "Requirements Engineering Process Maturity", entre otros relacionados. La selección de artículos se orientó a modelos específicos de IR, y que además ofrecieran una metodología de evaluación del proceso de IR.

Los modelos estudiados fueron: REGPG, R-CMM y R-CMMi, REPM, MDREPM, Uni-REPM. El análisis comparativo resultante de dichos modelos se detalla en [10].

El resultado más importante del análisis fue la selección de Uni-REPM como modelo de referencia, por los siguientes motivos:

- *Evalúa proyectos*, esto constituye una ventaja para el estudio a realizar, dado que se tomarán proyectos de referencia para evaluar el proceso de IR en cada uno de ellos, antes y después del proceso de certificación de calidad.
- *Unifica prácticas* de IR para desarrollo a medida y desarrollo orientado al mercado, y las empresas a evaluar presentan desarrollos de ambos tipos.
- Presenta *facilidad de evaluar y calificar*, es sencillo llevar adelante un proceso de evaluación a través de una lista de verificación que otorga el modelo, y el proceso de calificación es simple.

## Caracterización de empresas

En cuanto a la caracterización de las empresas de software, se realizaron entrevistas a dichas organizaciones entre los meses de mayo y junio de 2012. En la mayoría de los casos quienes respondieron fueron los gerentes o directores de las empresas. Como resultados destacados se puede mencionar que todas las empresas son de Corrientes Capital, provincia de Corrientes, Argentina, e iniciaron sus operaciones entre el año 2000 y el 2008.

El principal servicio ofrecido por estas empresas es el desarrollo de software a medida, seguido por la implementación y puesta a punto de productos de software, y el soporte y asistencia de productos de software. También ofrecen actualización de productos de software y otros servicios relacionados, entre los que figuran capacitación, alojamiento de sitios web, y streaming de audio y video. Le sigue el desarrollo de productos de software (enlatados), servicios informáticos de valor agregado (consultorías), y en menor medida servicios brindados mediante el uso de software, y desarrollo de software embebido en equipos electrónicos.

En cuanto al rubro de productos, el 80% de las empresas se orienta al comercio, el 60% de las empresas ofrece soluciones para el sector público, un 40% lo hace para el sector primario, industria de la construcción, finanzas y seguros, transporte, comunicaciones e inmobiliarios; un 20% de las empresas cubre necesidades de software de la industria manufacturera, hotelería y restaurantes. La cantidad promedio de empleados por empresa es de diez, de los cuales el 92% se encuentra trabajando en el lugar físico de la empresa, y un 8% a distancia. De ellos un 48% trabaja a tiempo completo y un 52% a tiempo parcial, dos empresas tienen la mayor parte del personal a tiempo completo, y tres tienen la mayor parte a tiempo parcial.

En cuanto al nivel de instrucción, los recursos humanos tienen en su mayoría un título



universitario relacionado a los Sistemas de Información, los Licenciados en Sistemas representan el 30,77%, los Ingenieros en Sistemas son el 21,15%, le siguen con un 19,23% los Analistas de Sistemas o Programadores de Aplicaciones, 11,54% son Técnicos, 5,77% Diseñadores Gráficos, 5,77% estudiantes, y 5,77% corresponde a otro nivel de instrucción, relacionados a la administración, marketing, o área comercial de la empresa.

### **Evaluación de proyectos con el modelo Uni-REPM**

Para la aplicación del modelo Uni-REPM se solicitó a las empresas que seleccionen dos proyectos de software, uno realizado antes de la certificación de la norma ISO 9001:2008, y otro ejecutado después de la aplicación de la mencionada norma.

El 60% de las empresas seleccionó un proyecto de desarrollo a medida y el 40% eligió un proyecto de adaptación de un producto propio existente, todas las empresas eligieron el mismo tipo de proyecto para el caso de precertificación y poscertificación. En cuanto al punto de partida para obtener los requerimientos del proyecto, en los proyectos precertificación fueron los requerimientos del cliente en un 100% y en los proyectos poscertificación fueron en un 60% los requerimientos del cliente, y en el 40% restante el punto de partida fue una combinación de los requerimientos del cliente con las posibilidades del mercado.

En relación al modelo de ciclo de vida aplicado a los proyectos, en los de precertificación se impuso el modelo incremental con un 60%, mientras que el 40% restante siguió un modelo en cascada. Los proyectos de poscertificación se llevaron adelante con un modelo de ciclo de vida incremental (40%), iterativo (40%) y en cascada (20%).

La duración de los proyectos fue de 1 a 18 meses, pero la diferencia entre los proyectos de precertificación y de poscertificación en una misma empresa estuvo entre los 0 y 6 meses. La cantidad de personas involucradas en los proyectos varió de 2 a 6 personas por proyecto.

En referencia a los roles del equipo de desarrollo, en los proyectos de precertificación todos contaban con desarrolladores y líder de proyecto, un 60% de los proyectos incluyó el rol de analista de sistemas o requerimientos, un 40% el rol de responsable de interfaz de usuario y tester, y un 20% contó con documentador y otros roles. Para los proyectos de poscertificación todos contaron con líder de proyecto, desarrollador, analista de sistemas o requerimientos, tester, y documentador; un 80% de los proyectos cubrió el rol de responsable de interfaz de usuario, y un 20% mencionó contar con otros roles, como implementador y capacitador.

### **Análisis de los resultados**

El detalle de los resultados se describen en [15], entre estos se destaca que la certificación de la norma ISO 9001:2008 en las empresas del Polo IT mejoró las prácticas de IR, aunque tal mejora no fue suficiente para obtener un nivel de madurez mínimo del proceso de IR bajo el modelo Uni-REPM.

Sin embargo, se destaca que las empresas mejoraron significativamente en los tres niveles de madurez propuestos por el modelo Uni-REPM.

La mejora más significativa a raíz de la certificación se dio en las áreas Apoyo de la Organización y Documentación y Especificación de Requerimientos, y en las subáreas Acciones Generales y Comunicación y Negociación de Requerimientos del área Gestión de Requerimientos, en éstas, todas las empresas del Polo IT Corrientes, alcanzaron al menos el nivel 1 de madurez del proceso de IR según el modelo Uni-REPM.

Se presentaron mejoras parciales en las áreas: Elicitación de Requerimientos, Planificación del Lanzamiento y Validación de Requerimientos; y en las subáreas Gestión de la Configuración y Política de Trazabilidad de Requerimientos del área Gestión de Requerimientos, es decir, sólo algunas empresas del Polo IT Corrientes mejoraron.

La certificación no afectó el área Análisis de Requerimientos, no produjo mejora ni retroceso en las prácticas incluidas en el área.

En cuanto a las prácticas de IR, 19 de ellas fueron mejoradas luego de la certificación. Estas representan un 26% sobre el total de acciones propuestas por Uni-REPM.

La mejora en algunas prácticas se debe a que las empresas las incluyeron en la definición de sus procesos, pero otras se dieron porque la norma ISO 9001:2008 establece:

- Una serie de requisitos en torno a la comprensión y cumplimiento de los requerimientos del producto para lograr la satisfacción del cliente.
- Que la organización debe determinar los requisitos especificados por el cliente, los no especificados pero necesarios para la realización del producto, los legales y reglamentarios.
- Que la organización debe asegurarse de la disponibilidad de recursos e información necesarios para apoyar la operación y el seguimiento de los procesos.
- Que la organización debe dejar registros de los requerimientos del cliente en cuanto al producto, y de los cambios en los requerimientos del cliente.
- Que la organización debe determinar e implementar lo necesario para la comunicación



con el cliente relacionada a la información del producto.

- La revisión y validación de los requerimientos del producto.

Por todo lo expuesto, se concluye que la norma ISO 9001:2008, como norma genérica que establece ciertos requisitos, favorece el cumplimiento de las buenas prácticas más generales en cuanto a la gestión de procesos en una organización. Sin embargo, las prácticas más específicas de la Ingeniería del Software, y en particular de la IR, deben ser incluidas en la definición de los procesos de desarrollo por parte de las organizaciones, para que sean consideradas en el proceso de certificación.

Si la organización incluyera las buenas prácticas de la IR en la definición de sus procesos, entonces podría alcanzar niveles altos de madurez del proceso de IR bajo la certificación de la norma ISO 9001:2008.

En la actualidad, se trabaja en el análisis profundo de las prácticas propuestas por el modelo Uni-REPM, con el objetivo de definir un proceso de desarrollo de requerimientos de software que incorpore las buenas prácticas definidas en dicho modelo, y que además resulte aplicable a empresas de características similares a las del Polo IT Corrientes.

## Formación de Recursos Humanos

Dentro del marco de esta línea de investigación se ha desarrollado un Trabajo Final Integrador de la Especialidad en Ingeniería de Software de la UNLP, a partir del cual se plantea el plan de actividad para el desarrollo de la tesis de la Maestría en Ingeniería de Software de la misma universidad, bajo la dirección de docentes investigadores de la Facultad de Informática de la UNLP y la FaCENA-UNNE, propiciando la vinculación interinstitucional orientada a la formación de recursos humanos.

## Referencias

- [1] Cámara de Software y Servicios Informáticos - CESSI: Propuestas para el Plan de acción 2008-2011, [http://www.cessi.org.ar/documentacion/PLAN-2008-2011\\_Documento\\_Principal\\_11.pdf](http://www.cessi.org.ar/documentacion/PLAN-2008-2011_Documento_Principal_11.pdf).
- [2] Ley 25.922 - Ley de Promoción de la Industria del Software, [www.mincyt.gov.ar/multimedia/archivo/archivos/ley\\_25922\\_3.pdf](http://www.mincyt.gov.ar/multimedia/archivo/archivos/ley_25922_3.pdf).
- [3] Secretaría de Industria Comercio y de la Pequeña y Mediana Empresa: Resolución 61/2005: <http://infoleg.mecon.gov.ar/infolegInternet/anexos/105000-109999/106061/norma.htm>, (2005).
- [4] Polo IT Corrientes: Pymes correntinas de Software y Servicios Informáticos certificaran en Calidad, <http://poloitcorrientes.com/noticia.php?i=05229>, (2009).
- [5] Sommerville, I.: *Software Engineering*. Addison-Wesley (2005).
- [6] Loucopoulos, P., Karakostas, V.: *System Requirements Engineering*. McGraw-Hill (1995).
- [7] Berenbach, B., Paulish, D.J., Kazmeier, J., Rudorfer, A.: *Software & Systems Requirements Engineering: In Practice*. McGraw-Hill Osborne Media (2009).
- [8] Hofmann, H.F., Lehner, F.: Requirements engineering as a success factor in software projects. *IEEE Software*. 18, 58-66 (2001).
- [9] Mizuno, Y.: Software Quality Improvement. *Computer*. 16, 66-72 (1983).
- [10] Arruzazabala, M., Dapozo, G., Thomas, P.: Análisis comparativo de modelos de evaluación de procesos de Ingeniería de Requerimientos, *II Jornadas de Investigación en Ingeniería del NEA y Países Limítrofes*, <http://www.fre.utn.edu.ar/IJCYT/clean/files/get/item/2186>, (2012).
- [11] Nguyen, T.T.L.: The creation of Uni-REPM A universal model for assessing requirements engineering process maturity, Blekinge Institute of Technology, (2010).
- [12] Svahnberg, M., Gorschek, T., Nguyen, T.T.L., Nguyen, M.: Uni-REPM: validated and improved. *Requirements Engineering - Springer London*. 17, 19 (2012).
- [13] Nguyen, M.: Empirical Evaluation of a Universal Requirements Engineering Process Maturity Model, Blekinge Institute of Technology, (2010).
- [14] IRAM: IRAM | Instituto Argentino de Normalización y Certificación, <http://www.iram.org.ar/seccion.php?ID=3&IDS=42>. Accedido el 10 de Mayo de 2012.
- [15] Arruzazabala, M., Dapozo, G., Thomas, P.: Certificación ISO 9001:2008: Impacto en el Proceso de Ingeniería de Requerimientos. Anales de XVIII Congreso Argentino de Ciencias de la Computación. 1ª ed. - Bahía Blanca. Universidad Nacional del Sur, 2012. ISBN 978-987-1648-34-4

## Mejora de Procesos en el desarrollo de Sistemas de Software y en Procesos de Gestión. Experiencias en PyMEs.

Pesado Patricia <sup>(1,2)</sup>, Bertone Rodolfo <sup>(1)</sup>, Esponda Silvia <sup>(1)</sup>, Pasini Ariel <sup>(1)</sup>, Boracchia Marcos <sup>(1)</sup>, Martorelli Sabrina <sup>(1)</sup>, Swaels Maximiliano <sup>(1)</sup>

<sup>(1)</sup>Instituto de Investigación en Informática LIDI (III-LIDI)  
Facultad de Informática – UNLP  
50 y 120 La Plata Buenos Aires

<sup>(2)</sup> Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC)  
526 e/ 10 y 11 La Plata Buenos Aires

(ppesado, pbertone, sesponda, apasini, marcosb, smartorelli, mswaels) @lidi.info.unlp.edu.ar

### Resumen

Los conceptos de Calidad y Mejora de Procesos relacionados con el software han evolucionado fuertemente en los últimos años, debido al crecimiento constante de la Ingeniería de Software y a la necesidad de establecer modelos que sean capaces de acompañar dicho crecimiento. El avance de estos modelos está relacionado con la definición de estándares capaces de ordenar y predecir sistemáticamente la evolución de los proyectos y dar de esta forma un nivel de confianza sobre los mismos. El III-LIDI posee un grupo dedicado a la investigación y desarrollo de propuestas en temas relacionados con la mejora de los procesos de gestión y el aseguramiento de la calidad en cada una de las etapas del desarrollo de software.

### Palabras Claves

Ingeniería de Software – Calidad – Normas de Calidad – CMMI – ISO – COMPETISOFT

### Contexto

“Calidad en el desarrollo de Sistemas de Software” es una línea de investigación dentro del proyecto de “Tecnología y aplicaciones en Sistemas de Software Distribuidos. Experiencias en E-learning, E-government y Sistemas productivos”

del Instituto de Investigación en Informática LIDI acreditado por la UNLP y de proyectos específicos apoyados por la Facultad de Informática. Desde el año 2004 se trabaja en temas relacionados con el área de calidad, habiendo participado en el Proyecto Iberoamericano “COMPETISOFT - Mejora de Procesos para Fomentar la Competitividad de la Pequeña y Mediana Industria del Software de Iberoamérica” cuyos resultados han sido descritos en una publicación [12]. En el año 2012 se ha logrado la certificación IRAM-ISO 9001:2008 del Sistema de Gestión de Calidad del “Diseño y realización del curso de nivelación a distancia para el pre ingreso a la Facultad de Informática”.

Hay cooperación con varias Universidades de Argentina y del exterior, en particular con la Universidad Autónoma de México y la Universidad de Castilla-La Mancha.

Se participa en iniciativas como el Programa IberoTIC de intercambio de Profesores y Alumnos de Doctorado en el área de Informática.

Además, existen acuerdos de colaboración con empresas privadas en el tema de calidad en el desarrollo de sistemas de software.

### Introducción

La mejora de procesos en una organización se define como un conjunto de tareas llevadas adelante para obtener

productos de mejor calidad a partir de la revisión y adaptación de sus procesos. Si vemos a la organización como una empresa prestadora de servicios, el concepto de calidad está estrechamente ligado a la obtención de la satisfacción de las necesidades del cliente. En consecuencia el nivel de calidad de la organización estará ligado al nivel de calidad con que se preste el servicio. Cuanto más ordenadas y predecibles sean las organizaciones para la prestación de sus servicios, mayores serán sus niveles de calidad. Para lograr este efecto de previsibilidad, las organizaciones fueron definiendo los procesos de las actividades de cada proyecto. Cuando estos procesos fueron extensibles hacia otros proyectos, se fueron convirtiendo en estándares para la organización y la utilización de estos estándares son los que garantizan la calidad de los servicios prestados.[1]

Las organizaciones prestadoras de servicios de software, apuntan a la predictibilidad con la definición de las etapas del ciclo de vida y estableciendo los procesos necesarios para que cada etapa sea realizada según lo planificado. Estos procesos deben ser lo suficientemente ordenados para obtener la confiabilidad en las estimaciones y además ser lo suficientemente flexibles para adaptarse rápidamente a los cambios que, particularmente, la industria del software requiere. Este es el gran desafío que hoy enfrenta este tipo de organizaciones.

En los últimos 15 años y en busca de esta previsibilidad, que garantiza la calidad de los productos desarrollados, diferentes organismos relacionados con la industria del software fueron construyendo diferentes modelos y normas certificables. Los mismos se dividen en dos grandes grupos: los orientados a la calidad del producto, y los orientados a la calidad del proceso.

Dentro de los modelos y normas orientados a la calidad del producto tenemos las normas ISO/IEC 9126 [2], que apuntan a la medición de los atributos de confiabilidad, usabilidad, seguridad, disponibilidad, escalabilidad, entre otros, que aún se encuentran en vigencia en Argentina pero que serán reemplazadas en corto plazo por la familia ISO/IEC 25000 [3], también llamada SQuaRE (Requisitos y Evolución de Calidad de Producto de Software), que ya se encuentra vigente en varios países.

En el segundo grupo, de calidad del proceso, tenemos: CMMI [4] e ISO/IEC 15504 [5] en su última versión, el cual toma como base la descripción de ciclo de vida del software definido en la ISO/IEC 12207 [6]. En el marco de la gestión de proceso de TI tenemos ITIL [7] e ISO/IEC 20000 [8]. Observando al proceso de software netamente como un proceso de gestión podemos ver la opción de utilizar ISO/IEC 9001 [9] bajo las directrices que se indican en la ISO/IEC 90003 [10]. En los últimos años un número importante de empresas desarrolladoras de software en Argentina han optado por la última opción, ya que fue fuertemente fomentada por la denominada ley de promoción del software [11].

La aplicación de este tipo de herramientas requiere una importante inversión, que se verá reflejada a largo plazo, por lo que muchas veces se las considera como un gasto y no una inversión, sin percibir a la calidad como un valor en sí mismo que hará prosperar a la organización. En el caso de las PyMEs, la posibilidad de utilizar este tipo de herramientas es mucho más dificultosa [12], por el tamaño reducido de los grupos de trabajo, por los problemas de definición de roles dentro de estos grupos, por las responsabilidades poco precisas y por los recursos restringidos con que cuentan este tipo de organización. Debido a esto, aparecen modelos directamente orientados a este tipo de organizaciones: MoProSoft en

México, MPS BR en Brasil, y COMPETISOFT como un proyecto Iberoamericano [13] y la más reciente ISO/IEC 29110 [14] que se establece como estándar en las Mejoras de Procesos para PyMEs.

## Líneas de investigación y desarrollo

- Análisis y estudio de normas y modelos orientados a la Calidad del Producto, como por ejemplo ISO/IEC 9126 e ISO/IEC 25000
- Análisis y estudio de normas y modelos orientados a la Calidad del Proceso, como por ejemplo CMMI, ISO/IEC 15504, ITIL, ISO/IEC 20000, ISO/IEC 9001 e ISO/IEC 90003.
- Análisis y estudio de modelos de Mejora de Procesos para PyMEs.
- Evaluación y mantenimiento de procesos en organismos públicos y privados según los requisitos de ISO 9001.
- Investigación sobre estándares y modelos de mejoras aplicables a equipos que integren hardware y software ad-hoc para la obtención de una certificación del producto completo.
- Análisis de la problemática que presentan las PyMEs al aplicar un proceso de mejora, haciendo hincapié en las tareas o actividades con inconvenientes o faltantes.

## Resultados obtenidos / esperados

Dentro de los resultados generales esperados y obtenidos del proyecto se puede mencionar:

- Se definieron las relaciones y/o equivalencias entre los modelos de procesos para PyMEs y las directrices de la ISO 90003 para la certificación de procesos de software con ISO 9001.
- Mantenimiento de los procesos certificados ISO 9001:2004 del Pre-Ingreso a Distancia de la Facultad de Informática y aprobación, en el año 2013, de la primera auditoría de seguimiento de los procesos certificados.
- Construcción de un entorno WEB para asistir a una PyME en un proceso de mejora. Se han desarrollado las herramientas de Gestión de Cambios y Gestión de Roles, Gestión de Riesgos y Verificación y Validación (V&V).
- Se prevee la ampliación del ambiente web para PyMEs incluyendo herramientas que abarquen otros aspectos de la Gestión de la Configuración del Software. Asimismo se pretende extender el ambiente para lograr que interactúen sus propias herramientas e integrarlo con ambientes externos.
- En el período se realizó el análisis y diagnóstico, a cargo de miembros del proyecto que cuentan con la capacitación de auditores internos IRAM, de los procesos de desarrollo de Software del Departamento de Informática en Salud del Hospital Italiano (CABA). El Hospital Italiano de Buenos Aires (HIBA) es un hospital universitario de alta complejidad que posee gran parte de sus procesos informatizados por un equipo propio del hospital, dirigido por un equipo multidisciplinario de informáticos y médicos.



- Es de interés la generación de herramientas para la evaluación de PyMEs de los procesos “Project Management” y “Software Implementation” con la Norma ISO 29110 “Software engineering – Lifecycle profiles for Very Small Entities (VSEs)”
- Se avanzará en la certificación de equipos que integran hardware y software (como por ejemplo Terminales Interactivas de Voto Electrónico) a partir de la asignación de subsidios de la Facultad de Informática y del Programa de Mejora de la Gestión de la Calidad en las Unidades de Investigación, Desarrollo y Transferencia de la UNLP.

## Formación de recursos humanos

- Capacitación de los miembros del proyecto con diversos cursos del Instituto Argentino de Normalización y Certificación (IRAM).
- Uno de los miembros ha obtenido su Tesis de Maestría: “Ambiente para la ayuda a la mejora de procesos en PyMEs” en la Facultad de Informática de la UNLP
- Se desarrolla una tesis de doctorado y tesinas de grado en el área.
- Los integrantes de esta línea de Investigación participan en el dictado de asignaturas/cursos de grado/postgrado en la Facultad de Informática de la UNLP y en otras universidades del país. En

particular, en la UNLP, se dicta la asignatura “Calidad de Sistemas de Software en Pequeñas y Medianas Empresas”.

## Referencias

- [1] Pantaleo Guillermo, Calidad en el desarrollo de software, Editorial Alfaomega, ISBN 978-987-1609-23-9 Año 2011
- [2] ISO/IEC 9126, Software engineering - Product quality
- [3] ISO/IEC 25000, Software Engineering -- Software product Quality Requirements and Evaluation (SQuARE) -- Guide to SQuARE
- [4] Chrissis, Mary, CMMI 2ed Guía para la interpretación de procesos y mejora de productos, Editorial Pearson Educación, Año 2009 ISBN 978-84-7829-096-3
- [5] ISO/IEC .15504: 2004 Information Technology – Process assessment
- [6] ISO/IEC 12207, Systems and software engineering -- Software life cycle processes
- [7] ITIL “Information Technology Infrastructure Library” -- <http://www.itil-officialsite.com/> (marzo 2013)
- [8] ISO/IEC 20000 , Information technology -- Service management
- [9] IRAM-ISO. 9001:2008, “Sistema de gestión de la calidad. Requisitos” Segunda Edición 2008 IRAM
- [10] ISO/IEC 90003:2004, “Software engineering — Guidelines for the application of ISO 9001:2000 to computer” 2004, ISO
- [11] Ley 25.922 - LEY DE PROMOCION DE LA INDUSTRIA DEL SOFTWARE - [http://www.mincyt.gov.ar/multimedia/archivo/archivos/ley\\_25922\\_3.pdf](http://www.mincyt.gov.ar/multimedia/archivo/archivos/ley_25922_3.pdf)

[12] Piattini, Oktaba, Orozco, "COMPETISOFT. Mejora de procesos software para pequeñas y medianas empresas", Editorial Ra-Ma, Año 2008

[13] Esponda, Pasini, Pesado, Boracchia, "ISO 9001 en PyMEs desarrolladoras de software asistidas por el Modelo COMPETISOFT" CACIC 2012 Bahía Blanca, Año 2012

[14] ISO/IEC 29110:2011, "Software engineering -- Lifecycle profiles for Very Small Entities (VSEs)" 2011, ISO

## Ingeniería de Software en el desarrollo de Aplicaciones para Dispositivos Móviles.

Pablo Thomas<sup>(1)</sup>, Nicolás Galdamez<sup>(1)</sup>, Lisandro Delia<sup>(1)</sup>, Federico Cristina<sup>(1)</sup>, Sebastián Dapoto<sup>(1)</sup>, Fernando Tinetti<sup>(1,2)</sup>, Patricia Pesado<sup>(1,2)</sup>, Armando De Giusti<sup>(1,3)</sup>

<sup>(1)</sup>Instituto de Investigación en Informática LIDI (III-LIDI)

Facultad de Informática – UNLP

50 y 120 La Plata Buenos Aires

<sup>(2)</sup> Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC)

526 e/ 10 y 11 La Plata Buenos Aires

<sup>(3)</sup> Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)

Argentina

{ pthomas, ngaldamez, ldelia, fcristina, sdapoto, fernando, ppesado, degiusti }@lidi.info.unlp.edu.ar

### Resumen

Se presenta una línea de investigación y desarrollo que tiene por objeto estudiar temas relacionados con aplicaciones para dispositivos móviles, fundamentalmente aspectos de Ingeniería de Software orientados al desarrollo e implementación de aplicaciones móviles, sobre diversos entornos operativos y el estudio y desarrollo de aspectos de conectividad entre dispositivos móviles.

**Palabras claves:** Dispositivo Móvil – SmartPhone - Plataformas para Dispositivos Móviles – Metodologías de Desarrollo para Dispositivos Móviles – Web Services - Conectividad

### Contexto

Esta línea de Investigación forma parte del Proyecto “Tecnología y aplicaciones en Sistemas de Software Distribuidos. Experiencias en E-learning, E-government y Sistemas productivos” del Instituto de Investigación en Informática LIDI de la Facultad de Informática, acreditado por la UNLP, y de proyectos específicos apoyados por diversos organismos.

Hay cooperación con Universidades de Argentina y se participa en iniciativas como el Programa IberoTIC de intercambio de Profesores y Alumnos de Doctorado en el área de Informática.

Por otra parte, se tiene financiamiento de Telefónica de Argentina en Becas de grado y posgrado.

### Introducción

Desde hace años, con el creciente desarrollo tecnológico cada vez más personas acceden a dispositivos móviles, los cuales han evolucionado constantemente en capacidades de servicio, ofreciendo funciones de comunicación y procesamiento de datos, que van mucho más allá que las simples llamadas telefónicas o ejecución de aplicaciones básicas.

Un “dispositivo móvil” es un término que abarca una amplia gama de aparatos electrónicos surgidos en los últimos años, que se caracterizan por su tamaño reducido, su conectividad, capacidad de procesamiento y almacenamiento de datos. Ejemplo de estos dispositivos lo constituyen los teléfonos móviles más evolucionados (denominados *smartphones*), cámaras digitales, reproductores de música, consolas de videojuegos, entre otros.

Actualmente los dispositivos móviles forman parte de la vida cotidiana y son cada vez más sofisticados.

Es evidente el crecimiento de la telefonía celular y otros dispositivos móviles como los GPS's, PDA, PALM, entre otros, los cuales

requieren de software y aplicaciones que funcionen de acuerdo a las necesidades de cada usuario.

El teléfono móvil, particularmente, se ha convertido en un período de pocos años en un dispositivo masivo para la sociedad. Desde sus primeras funciones básicas de realizar llamadas y poder estar localizable, ha pasado a ser un objeto indispensable en la vida diaria de la mayor parte de la población, con funcionalidades como agenda personal, calendario, o incluso reproducción de música y cámara de fotos.

Asimismo, la tecnología ha posibilitado capacidades tiempo atrás inimaginables, como conexión a Internet de alta velocidad, GPS (Sistema de Posicionamiento Global) o grabación de video en alta definición.

Las particularidades específicas de un entorno móvil incluyen: un alto nivel de competitividad, un tiempo necesariamente corto de entrega de aplicaciones, más la dificultad adicional que implica la identificación de stakeholders y sus requerimientos. Los equipos de desarrollo deben enfrentar los desafíos de este entorno cambiante, con frecuentes modificaciones en las necesidades y expectativas de los clientes.

También existen limitaciones tecnológicas, que se reflejan en la gran variedad de dispositivos existentes, cada uno de ellos con características específicas de hardware, firmware y sistemas operativos. Algunas de esas limitaciones están vinculadas al ancho de banda, área de cobertura y seguridad, entre otras, que seguramente serán resueltas en un futuro cercano.

Existen además restricciones inherentes a la plataforma, como por ejemplo el espacio de pantalla, la capacidad de memoria y procesamiento, la reserva de energía, entre otras.

En resumen, es necesaria una metodología de desarrollo de aplicaciones, adecuada a este contexto.

Por otra parte, el número de aplicaciones móviles que requieren conectividad crece constantemente. La necesidad de compartir

información entre dispositivos móviles está presente en muchas aplicaciones. En todos los casos, el intercambio de información entre estos dispositivos involucra los mismos requerimientos: un medio para el descubrimiento de otros dispositivos móviles en una red, el establecimiento de conexiones lógicas y comunicación de datos de la aplicación, y un soporte para la determinación de la calidad de las conexiones físicas.

En este sentido es útil la definición de mecanismos que permitan resolver estos aspectos de conectividad.

### **Líneas de Investigación y Desarrollo**

- Metodologías y técnicas de la Ingeniería de Software y su aplicación en el desarrollo de software para dispositivos móviles.
- Metodologías ágiles de desarrollo para la generación de aplicaciones móviles.
- Desarrollo de web services.
- Plataformas operativas para aplicaciones móviles.
- Entornos de simulación de aplicaciones móviles.
- Lenguajes de programación para aplicaciones móviles.
- Conectividad e intercambio de información entre dispositivos móviles.

### **Resultados esperados/obtenidos**

Los resultados esperados/obtenidos se pueden resumir en:

- Avanzar en la capacitación continua de los miembros de la línea de investigación.
- Adaptar alguna metodología de desarrollo para aplicaciones móviles.
- Se ha logrado construir un prototipo que permite interactuar con la plataforma de e-learning Web-UNLP. Parte de la interface se presenta en la figura 1 y figura 2. La evolución continua de este prototipo es un objetivo específico que incluye esta línea de investigación.



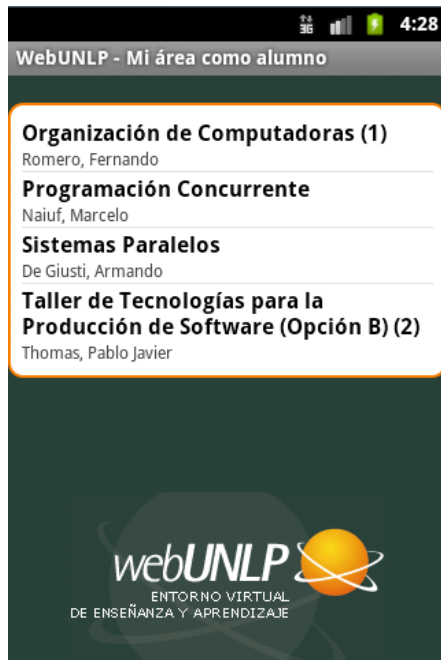


Figura 1

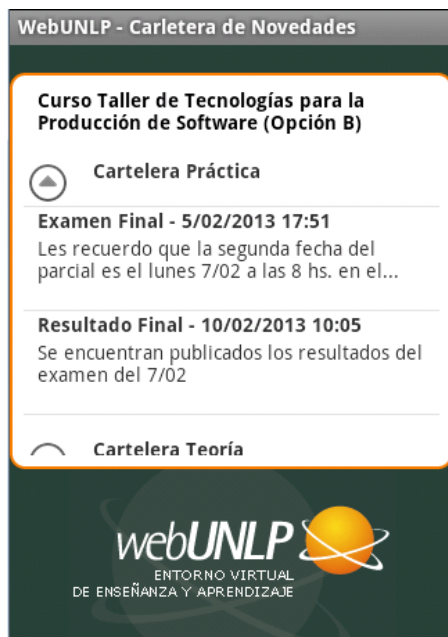


Figura 2

- Se ha desarrollado un framework *open source* – denominado *NetworkDCQ*<sup>1</sup> – para desarrolladores de aplicaciones (principalmente móviles) que actúa como una capa de soporte para el descubrimiento de hosts, la comunicación entre éstos y la caracterización de la calidad de servicio del medio (QoS). El mismo permite el desarrollo de distintos tipos de aplicaciones sobre diferentes

<sup>1</sup> <https://code.google.com/p/networkdcq/>

plataformas, tal como Android, J2ME y J2SE. La arquitectura resumida del framework se presenta en la figura 3.

Esta solución abstrae a los desarrolladores de la frecuente resolución de problemas relacionados con *networking*.

A fin de evaluar el framework, se han desarrollado dos aplicaciones que se apoyan en éste, las cuales varían en sus requerimientos de uso de red y forma de interacción. La primera es un juego en red multijugador de tipo *Asteroids*<sup>2</sup>, en la cual es necesario intercambiar información constantemente y con una alta tasa de actualización (al menos 30 veces por segundo). La segunda aplicación es la implementación multijugador del tradicional *Tic-Tac-Toe*<sup>3</sup>, en el cual se requiere una secuencia específica de interacción entre los hosts. En ambos casos el framework resultó de utilidad, simplificando la tarea de desarrollo de las aplicaciones en lo que refiere a conectividad.

- Se pretende extender el framework para otras plataformas como iOS, Windows Mobile y BlackBerry 10 OS.

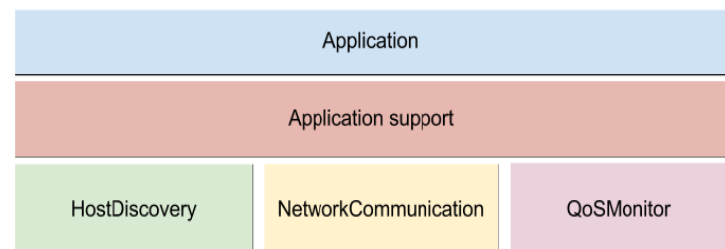


Figura 3

## Formación de Recursos Humanos

Los integrantes de esta línea de investigación dirigen Tesinas de Grado y Tesis de Postgrado en la Facultad de Informática, y Becarios III-LIDI en temas relacionados con el proyecto. Además participan en el dictado

<sup>2</sup> <https://code.google.com/p/asteroidsa/>

<sup>3</sup> <https://code.google.com/p/ticatacatoe/>

de asignaturas/cursos de grado y postgrado de la Facultad de Informática de la UNLP.

## Referencias

1. Pleeeger. Ingeniería de Software: Teoría y Práctica. Prentice-Hall. 2002
2. Anup Kumar y Bin Xie, *Handbook of Mobile Systems Applications and Services*. Editorial CRS Press, ISBN 978-1-4398-0152-9, Año 2012.
3. Sambasivan, D.; John, N.; Udayakumar, S.; Gupta, R., *Generic framework for mobile application development*, Internet (AH-ICI), 2011 Second Asian Himalayas International Conference on Computing & Processing (Hardware/Software).
4. Hammershoj, A.; Sapuppo, A.; Tadayoni, R., *Challenges for mobile application development Intelligence in Next Generation Networks*, (ICIN) 2010 14th International Conference on Computing & Processing (Hardware/Software)
5. Choi, Y.; Yang, J.-S.; Jeong, J., *Application framework for multi platform mobile application software development*, Advanced Communication Technology, 2009. ICACT 2009. 11th International Conference on Computing & Processing (Hardware/Software)
6. Grgurina, R.; Brestovac, G.; Grbac, *Development environment for Android application development: An experience report*, T.G. MIPRO, 2011 Proceedings of the 34th International Convention on Computing & Processing (Hardware/Software)
7. Anthony Wasserman, Carnegie Mellon Silicon Valley, *Software Engineering Issues for Mobile Application Development*, 2º Annual Workshop on Software Engineering for Mobile Application Development, MobiCASE '11, Santa Monica, California, USA, October 2011.
8. Ray Bareiss, Todd Sedano, Carnegie Mellon Silicon Valley, *Improving Mobile Application Development*, 2º Annual Workshop on Software Engineering for Mobile Application Development, MobiCASE '11, Santa Monica, California, USA, October 2011.
9. Luis Corral, Alberto Sillitti, Giancarlo Succi, Free University of Bozen-Bolzano, *Preparing Mobile Software Development Processes to Meet Mission-Critical Requirements*, 2º Annual Workshop on Software Engineering for Mobile Application Development, MobiCASE '11, Santa Monica, California, USA, October 2011.
10. Josh Dehlinger, Jeremy Dixon, Towson University, *Mobile Application Software Engineering: Challenges and Research Directions*, 2º Annual Workshop on Software Engineering for Mobile Application Development, MobiCASE '11, Santa Monica, California, USA, October 2011.
11. Ivo Salmre, *Writing Mobile Code Essential Software Engineering for Building Mobile Applications*, Addison Wesley Professional, ISBN: 0-321-26931-4.
12. Abrahamsson, P. (2007). Agile Software Development of Mobile Information Systems. In *Advanced Information Systems* (pp. 1-4). Berlin: Springer.
13. Asymco (2012). *The Rise and Fall of Personal Computing*. <http://www.asymco.com/2012/01/17/the-rise-and-fall-of-personal-computing/>.
14. China Internet Network (2012). China Internet Development Statistics Report. China Internet Network Information Center, <http://www.cnnic.cn/research/bgxz/tjbg/201207/P020120719489935146937.pdf>.
15. Meeker, M. (2012). D10 Conference. Internet Trends. Kleiner Perkins Caufield Byers <http://www.kpcb.com/insights/2012-internet-trends>.

## Extracción de Información Estática de Programas Escritos Usando el Paradigma Orientado a Objetos

Arnaldo Ceballos, Hernán Bernardis, Enrique Miranda, Mario Berón, Daniel Riesco

Departamento de Informática/ Facultad de Ciencias Físico Matemáticas y Naturales/  
Universidad Nacional de San Luis

Ejercito de Los Andes 950 - San Luis - Argentina

arnaldoceballos@gmail.com, {hbernardis, eamiranda, mberon, driesco}@unsl.edu.ar

### Resumen

La Comprensión de Programas (CP) es una disciplina de la Ingeniería del Software cuyo objetivo es facilitar el entendimiento de los sistemas. Esta disciplina está influenciada en gran medida por el tamaño de los mismos. Es decir, mientras más grande es el código del sistema, más complejo se hace su entendimiento.

Comprender un sistema de manera correcta y rápida disminuye claramente el costo de actividades tales como: mantenimiento, reingeniería, evolución; lo cual representa una de las características más importantes de esta disciplina.

Entre los principales desafíos en la CP se destaca lo siguiente: lograr reconstruir la relación entre el Dominio del Problema y el Dominio del Programa. El primero hace referencia a la salida del sistema y el segundo a las componentes utilizadas para producir dicha salida.

Para el caso particular del Dominio del Programa, uno de los lugares más significativos desde donde se puede extraer información es el código fuente. La información que se obtiene de esta extracción se clasifica en estática, si es obtenida sin

ejecutar el programa y dinámica, si es obtenida en tiempo de ejecución.

En este artículo se describe una línea de investigación centrada en la extracción de información estática de los sistemas para facilitar el proceso de comprensión de los mismos.

**Palabras claves:** *Comprensión de Programas, Dominio del Programa, Extracción de la Información Estática.*

### Contexto

La línea de investigación descrita en este artículo se encuentra enmarcada dentro del proyecto: Ingeniería de Software: Aspectos de alta sensibilidad en el ejercicio de la profesión de Ingeniero de Software de la Universidad Nacional de San Luis. Dicho proyecto, es reconocido por el programa de incentivos y es la continuación de diferentes proyectos de investigación de gran éxito a nivel nacional e internacional.

Este proyecto se desarrolla en el marco del Área de Programación y Metodologías de Desarrollo de Software y del Laboratorio de Calidad e Ingeniería de Software, de la Facultad de Ciencias Físico Matemáticas y Naturales de la Universidad Nacional de San Luis (Argentina) y del Grupo de

Procesamiento de Lenguajes del departamento de informática de la Universidade do Minho (Portugal).

## Introducción

La Comprensión de Programas [1,2,3] es un área de la Ingeniería del Software destinada a elaborar métodos, técnicas y herramientas, basadas en un proceso cognitivo y un proceso de ingeniería, para lograr un conocimiento profundo de un sistema de software.

El proceso cognitivo [4,5] implica el estudio y análisis de las fases y pasos seguidos por los programadores para comprender un sistema. El proceso de ingeniería [1] tiene la finalidad de representar la información del sistema de manera que tal que enfatice sus principales aspectos. Este proceso implica el estudio de áreas tales como: Visualización de Software, Extracción de la Información, Administración de la Información.

El proceso de comprensión de programas se traduce en la habilidad de entender una pieza de código escrito en un lenguaje de alto nivel. El lector de un programa consigue extraer el significado del mismo cuando comprende de que forma el código cumple con la tarea para la cual fue creado [6].

Por lo expuesto en el párrafo anterior, se puede afirmar que el principal desafío en la CP consiste en lograr relacionar correctamente el Dominio del Problema y el Dominio del Programa [1]. Una forma de alcanzar este objetivo consiste en:

1. Extraer información del Dominio del Problema y del Dominio del Programa.
2. Proveer representaciones para ambos dominios a partir de esta información.
3. Definir una estrategia de vinculación que permita unir ambas representaciones.

Para poder construir la representación de cada dominio, previamente es necesario extraer información (EI) de cada uno de ellos. La Figura 1 visualiza los pasos previamente mencionados.

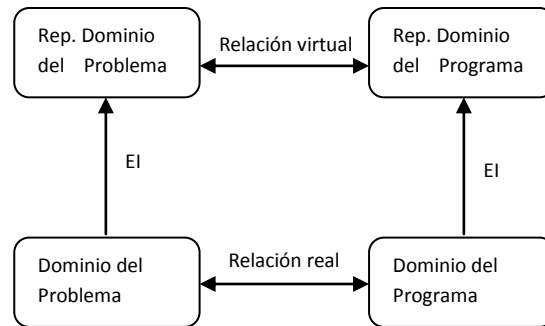


Figura 1. Modelo de Comprensión de Programas.

El proceso de extracción de información desde el Dominio del Problema es un gran desafío en la Ingeniería del Software, debido a su elevada complejidad. Dicha complejidad generalmente se debe a:

- La carencia de conocimiento del programador del sistema provocada, entre otras cosas, por la falta de documentación o cuando aquellos que intentan comprenderlo no son los creadores del mismo.
- El olvido de los aspectos relacionados al mismo por parte de los creadores luego de un tiempo.

Por otro lado, el Dominio del Programa es más palpable y concreto que el Dominio del Problema, porque está relacionado a las componentes que generan el comportamiento del sistema, presentes en el código fuente. Por lo tanto, el análisis sobre el código fuente permite obtener una gran cantidad de información del sistema y, con ella, crear representaciones del Dominio del Programa. La información que puede extraerse desde éste, se puede clasificar en dos tipos:



- Dinámica: es la que se obtiene a través de la ejecución del programa. Un ejemplo de estrategias de Extracción de la Información Dinámica es la Instrumentación de Código [7,8]. Esta estrategia consiste en insertar instrucciones dentro del código fuente del programa de modo que cuando éste se ejecute, proporcione información relacionada al comportamiento del mismo durante su ejecución.
- Estática: es aquella que se adquiere a través del uso de técnicas de compilación tradicionales, como por ejemplo: análisis lexicográfico, análisis sintáctico y análisis semántico del código fuente. Este tipo de información es la que comúnmente manejan los compiladores para: determinar la correctitud de los programas, generar código, optimizar código, entre otras tantas.

En este artículo se describe una línea de investigación que se centra en el estudio y creación de Técnicas de Extracción de la Información Estática desde el Dominio del Programa.

La organización de este artículo se expone a continuación. La sección 3 describe la línea de investigación. La sección 4 presenta los resultados obtenidos y esperados. Finalmente, la sección 5 describe brevemente las tareas realizadas en el contexto de los recursos humanos.

### 3. Línea de Investigación y Desarrollo

La información estática es obtenida sin ejecutar el programa. Una forma posible de hacerlo es usando técnicas de compilación. Dichas técnicas permiten capturar nombres de variables, tipos de variables, nombres de

procedimientos/métodos, variables locales a un procedimiento/método, etc.

Los procesos de extracción de información estática son fundamentales para la Representación del Dominio del Programa. Éstos permiten detectar errores lexicográficos, sintácticos y semánticos, relaciones entre las componentes del sistema, entre otras.

Básicamente esta línea de investigación se centra: i) En el estudio de diferentes herramientas de extracción de la información, tales como: Analizadores Lexicográficos, Analizadores Sintácticos, Depuradores, Profilers, entre otros; ii) En la elaboración e implementación de representaciones del Dominio del Programa basadas en la información estática. A modo de ejemplo se pueden mencionar: El grafo de funciones, el grafo de dependencias de módulos, el grafo de dependencias del sistema, etc. iii) En la definición de métodos para seleccionar la herramienta de extracción más apropiada para un Dominio del Programa específico, iv) En el análisis y creación de técnicas de compilación.

Además de las temáticas mencionadas en el párrafo precedente, el grupo de investigación también estudia diferentes formas de visualizar la información extraída. Esta tarea tiene como finalidad proporcionar un conjunto de técnicas y herramientas que disminuyan sustancialmente la brecha existente entre el conocimiento del ingeniero y el subyacente al sistema que se desea entender.

### 4. Resultados y Objetivos

Los resultados obtenidos hasta el momento se vinculan estrechamente al lenguaje de programación Java y a un generador de analizadores sintácticos que ha demostrado

ser robusto. Dichos resultados se mencionan a continuación:

- Se analizó una especificación de la gramática del lenguaje Java y se pudo concluir que es posible insertar acciones semánticas, para obtener mediante atributos sintetizados y heredados la información requerida.
- Se estudiaron herramientas de creación automática de Analizadores Lexicográficos y Sintácticos que permiten utilizar los conceptos anteriormente mencionados (atributos sintetizados, heredados y acciones semánticas) en la gramática. Del estudio previamente mencionado se pudo concluir que AnTLR [9] es la herramienta más apropiada para llevar a cabo la inspección estática del código.
- Se usó AnTLR con gramáticas de ejemplo para realizar inspecciones de prueba.

Dentro de los objetivos planteados a corto plazo se encuentran:

- La construcción de una herramienta que genere distintas vistas a partir de la información estática obtenida de programas escritos en Java.
- Ampliar la cantidad de información extraída, y de esta manera poder crear representaciones aún más completas.
- Crear ontologías sobre la información extraída de manera de facilitar la reconstrucción virtual de la relación real Dominio del Problema - Dominio del Programa. Se pretende crear ontologías para describir los conceptos y las relaciones pertenecientes al Dominio del Problema, con el objetivo de sistematizar una representación de este dominio, y facilitar la

identificación del mapeo de estos conceptos con secuencias de ejecución.

Todos los trabajos futuros mencionados previamente permiten percibir la importancia de la línea de investigación presentada en este trabajo para la Comprensión de Programas y la Ingeniería Inversa.

## 5. Formación de Recursos Humanos

Actualmente, los temas abordados por esta línea de investigación están siendo desarrollados como parte de trabajos de Licenciatura en Ciencias de la Computación.

Además se está trabajando en conjunto con integrantes que abordan temas fuertemente relacionados con el de la temática descrita en este artículo. Este grupo de investigadores se encargan de:

1. Extraer información para la construcción de ontologías que permitan representar el Dominio del Programa.
2. Extraer información dinámica para representar el Dominio del Programa.
3. Elaborar estrategias de visualización de software.
4. Elaborar métodos de evaluación para facilitar la toma de decisiones en un amplio rango de herramientas y métodos usados en Comprensión de Programas.
5. Estudiar y definir lenguajes específicos del dominio.
6. Estudiar lenguajes orientados a la web.

El objetivo principal de este trabajo colaborativo es: Proporcionar resultados relevantes para el área de la Comprensión de Programas, y a partir de estos resultados definir tesis de maestría y doctorado. Se pretende que las tesis, antes mencionadas, se puedan llevar a cabo en el ámbito nacional

como así también en el internacional. Fomentando, de esta manera, la cooperación inter-universitaria.

### Referencias

[1] Mario Berón; Roberto Uzal; Pedro Rangel Henriques; Maria João Varanda Pereira. Inspección de Código para relacionar los Dominios del Problema y Programa para la Comprensión de Programas. X Workshop de Investigadores en Ciencias de la Computación. 2008.

[2] Mario M. Berón, Daniel Riesco, Germán Montejano, Pedro R. Henriques, Maria J. Pereira. Estrategias para Facilitar la Comprensión de Programas. XII Workshop de Investigadores en Ciencias de la Computación. 2010.

[3] Liberman H.; Fry C. Bridging the Gulf Between Code and Behavior in Programming. ACM Conference on Computers and Human Interface. Denver, Colorado. 1994.

[4] M.-A.D. Storey, F. D. Fracchia, and H. A. Mueller. 1997. Cognitive Design Elements to Support the Construction of a Mental Model during Software Visualization. In Proceedings of the 5th International Workshop on Program Comprehension (WPC '97).IEEE Computer Society, Washington, DC, USA, 17-.

[5] Wang Kechao; Wang Tiantian; Su Xiaohong; Ma Peijun. "Overview of Program Comprehension", Computer Science and Electronics Engineering (ICCSEE), 2012 International Conference on, vol.1, no., pp.601,605,23-25 March 2012. doi: 10.1109/ICCSEE.2012.285.

[6] Walenstein, A., "Theory-based analysis of cognitive support in software comprehension tools", Program Comprehension, 2002.

Proceedings. 10th International Workshop on, vol., no., pp.75,84,2002.

[7] Hernán Bernardis. Instrumentación de Programas Escritos en Java para Interconectar los Dominios del Problema y del Programa. 40° Jornadas Argentinas de Informática e Investigación Operativa (JAIIO). 2011.

[8] Berón Mario; Henriques, Pedro; Pereira, Maria João; Uzal, Roberto. Instrumentación de programas escritos en C para interrelacionar las vistas comportamental y operacional de los sistemas de software. In XV CACIC'09-Argentine Congress on Computer Science. S. S. de Jujuy. 2009.

[9] ANTLR, ANOther Tool for Language Recognition, <http://www.antlr.org/>.

## Métodos y Procesos para Sistemas Distribuidos. Experiencias

Patricia Pesado<sup>(1,2)</sup>, Pablo Thomas<sup>(1)</sup>, Rodolfo Bertone<sup>(1)</sup>, Hugo Ramón<sup>(1)</sup>, Ariel Pasini<sup>(1)</sup>,  
Eduardo Ibañez<sup>(1)</sup>, Cesar Estrebow<sup>(1)</sup>, Luciano Marrero<sup>(1)</sup>, Lisandro Delia<sup>(1)</sup>, Nicolás  
Galdamez<sup>(1)</sup>, Alejandra Ripodas<sup>(1)</sup>, Verónica Aguirre<sup>(1)</sup>

<sup>(1)</sup>Instituto de Investigación en Informática LIDI (III-LIDI)  
Facultad de Informática – UNLP  
50 y 120 La Plata Buenos Aires

<sup>(2)</sup>Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC)  
526 e/ 10 y 11 La Plata Buenos Aires

{ppesado, pthomas, pbertone, hramon, apasini, eibanez, cesarest, lmarrero, ldelia, ngaldamez, aripodas, vaguirre, gcaseres}@lidi.info.unlp.edu.ar

### Resumen

Se presenta la continuidad de la línea de investigación y desarrollo, que tiene por objeto estudiar temas relacionados con aspectos de Ingeniería de Software orientados al desarrollo e implementación de proyectos concretos de Sistemas Distribuidos, en particular Sistemas Web, Sistemas Sensibles al Contexto y aplicaciones de E-Government, considerando las extensiones necesarias en las metodologías y herramientas clásicas.

**Palabras claves:** Sistemas Distribuidos – Ingeniería de Requerimientos – Metodologías de Desarrollo – Sistemas Web – Bases de Datos Distribuidas – E-Government

### Contexto

Esta línea de Investigación forma parte del Proyecto “Tecnología y aplicaciones en Sistemas de Software Distribuidos. Experiencias en E-learning, E-government y Sistemas productivos” del Instituto de Investigación en Informática LIDI de la Facultad de Informática, acreditado por la UNLP, y de proyectos específicos apoyados por diversos organismos.

### Introducción

Un sistema distribuido consiste en un conjunto de computadoras autónomas conectadas por una red y con soporte de software distribuido. Permite que las computadoras coordinen sus actividades y compartan recursos de hardware, software y

datos, de manera tal que el usuario percibe una única facilidad de cómputo integrada, aunque ésta pueda estar implementada por varias máquinas en distintas ubicaciones [1].

Los entornos distribuidos persiguen los siguientes objetivos:

- 1) Portabilidad de aplicaciones y de servicios (ejecución en diferentes computadoras, con hardware y software de base diferentes)
- 2) Interoperabilidad, posibilidad de comunicar diferentes equipos y aplicaciones
- 3) Integración, intercambio de información sin necesidad de intervención de agentes externos.
- 4) Transparencia, la utilización de la información por parte de los usuarios no requiere conocer su ubicación dentro de la red.
- 5) Facilidad de crecimiento del sistema, agregando o cambiando el hardware de base
- 6) Seguridad de los datos.

Los sistemas abiertos son sistemas distribuidos donde además de las características de estos últimos se pretende generar interfaces entre sus componentes que respeten un conjunto de normas de programación, comunicación, e interfaces aceptadas por múltiples plataformas.



La Ingeniería de Software comprende la aplicación de principios científicos para generar una solución elaborada de software que resuelva un problema determinado, y el mantenimiento subsecuente de ese sistema de software hasta el final de su vida útil [2]. La utilización de estas prácticas para resolver Sistemas Distribuidos, de Tiempo Real, Sensibles al Contexto o Aplicaciones para Dispositivos Móviles, hace necesaria su adaptación en función de las características de dichos sistemas.

La adopción de un enfoque de ingeniería para el desarrollo de software, genera una serie de fases o estados conformando un ciclo de vida. Este ciclo de vida está guiado por una planificación que incluye el conjunto de acciones a realizar, y los productos generados por la aplicación del plan (inclusive el mismo plan) están administrados por diferentes Metodologías de Gestión y Desarrollo [3].

En el recorrido del ciclo de vida del desarrollo del software, la fase inicial comprende a la Ingeniería de Requerimientos que permite comprender, documentar y acordar sobre el alcance del problema, teniendo esto impacto directo sobre la Planificación y la Gestión del Proyecto de acuerdo a la Metodología de desarrollo seleccionada para el mismo [4]. Este no es el único impacto que justifica el énfasis en la Ingeniería de Requerimientos, ya que esta disciplina permite establecer claramente *que, porque, en que contexto*, a quien está dirigido y como va a ser utilizado, posponiendo el *cómo* a etapas posteriores en el ciclo de vida, constituyendo el fundamento de la construcción de un Sistema de Software [5].

El modelo de procesos elegido para el desarrollo de software define las actividades a realizar para la generación de productos de acuerdo a los objetivos planteados [6].

Otro modelo de desarrollo lo constituyen los denominados métodos ágiles. El desarrollo ágil de software es un marco de trabajo conceptual de la ingeniería de software que promueve iteraciones en el desarrollo a lo largo de todo el ciclo de vida del proyecto. Existen diversos métodos de desarrollo ágil;

la mayoría minimiza riesgos desarrollando software en cortos lapsos de tiempo.

El software desarrollado en una unidad de tiempo es llamado una iteración, la cual debe durar un tiempo acotado. El propósito de cada iteración es generar una versión que responda al reducido conjunto de requerimientos definidos para ella. Al final de cada iteración el equipo vuelve a evaluar las prioridades del proyecto.

Los métodos ágiles enfatizan las comunicaciones cara a cara en vez de la documentación. También enfatizan que el software funcional es la primera medida del progreso.

Los principios esenciales del desarrollo ágil son: participación del cliente, entrega incremental, focalización en personas más que en procesos, aceptar el cambio y mantener la simplicidad como meta para cada iteración.

Asociado con la Ingeniería de Software de Sistemas Distribuidos, está el problema de utilizar un entorno WEB para los servicios que ofrece el Sistema. El desarrollo de arquitecturas centradas en un servidor (o un conjunto de servidores distribuidos) que ofrecen una interfaz WEB a los usuarios, ha generado un importante desarrollo en la Investigación de metodologías y herramientas orientadas a Sistemas WEB, así como ha generado la necesidad de establecer nuevas métricas y parámetros de aseguramiento de la Calidad para tales Sistemas [7] [8] [9].

Por otra parte, los Sistemas Sensibles al Contexto representan una rama emergente de los Sistemas de Software, que si bien es incipiente, tiene una amplia aceptación y varias áreas, tales como Inmótica, Domótica, E-commerce, entre otras.

La característica fundamental de los Sistemas Sensibles al Contexto es su adaptación constante a lo que sucede en el entorno de la aplicación; es decir, la respuesta del sistema se ajusta a los cambios que suceden en su contexto. La necesidad de aplicar un proceso de Ingeniería de Requerimientos es razón de análisis e investigación de la comunidad de

Ingeniería de Software para este tipo de sistemas.

Otro tema de estudio es la Estimación de esfuerzo en el desarrollo de sistemas de software a partir de Especificación de Requerimientos, particularmente modificando el método que utiliza Puntos de Casos de Uso [28] [29].

Finalmente, es de sumo interés el estudio de procesos de gobierno electrónico. El gobierno electrónico consiste en el uso de las tecnologías de la información y el conocimiento en los procesos internos de gobierno en la entrega de los productos y servicios del Estado tanto a los ciudadanos como a la industria.

Muchas de las tecnologías involucradas y sus implementaciones son las mismas o similares a aquellas correspondientes al sector privado del comercio electrónico (e-business), mientras que otras son específicas o únicas en relación a las necesidades del gobierno.

Se basa principalmente en la implantación de herramientas como portales, redes sociales o comunidades virtuales y otras, buscando una mejora en la eficiencia y eficacia de los procesos internos y de vinculación con la sociedad.

El gobierno electrónico describe el uso de tecnologías para facilitar la operación de gobierno y la distribución de la información y los servicios del mismo. Se debe proporcionar al exterior información acerca de sí mismo y los procesos que lleva a cabo. El nivel de transparencia mide el esfuerzo para hacer disponible la información a través de su sitio web. La transparencia no se refleja sólo en la cantidad de información; también en la calidad de la misma. Por otro lado, la interactividad mide el grado de facilidad con la que los usuarios pueden acceder a la información proporcionada por el gobierno y utilizarla. Esto no sólo implica el uso de la información digital, sino también la transacción de servicios en un sentido eminentemente práctico.

El gobierno electrónico debe centrarse en la inclusión de los ciudadanos de una manera participativa a través de las tecnologías de la información y la comunicación.

### **Líneas de Investigación y Desarrollo**

- Conceptos de procesamiento distribuido. Arquitectura, comunicaciones y software. Middleware.
- Metodologías ágiles de desarrollo utilizando frameworks propios y disponibles de uso libre con diferentes tecnologías.
- Ingeniería de Requerimientos para Sistemas Sensibles al Contexto.
- Estimación de esfuerzo en el desarrollo de software.
- Bases de Datos Distribuidas.
- Herramientas de modelización de Bases de Datos.
- Herramienta para Gestión de Riesgo en Proyectos de Software.
- Herramientas de integración y mantenimiento de proyectos distribuidos.
- Reingeniería de sistemas complejos que migran por downsizing a esquemas cliente-servidor distribuidos.
- E-Government

### **Resultados esperados/obtenidos**

Los resultados esperados/obtenidos se pueden resumir en:

- Avanzar en la capacitación continua de los miembros de la línea de investigación.
- Desarrollar soluciones a problemas concretos de software de sistemas distribuidos, poniendo énfasis en el desarrollo de metodologías y herramientas específicas para clases de aplicaciones.

Se mencionan algunas transferencias realizadas en el año:

#### **3.1. Gobierno electrónico**

Desde el año 2003 se trabaja en aplicaciones en esta área, entre los cuales se pueden mencionar:

- Desarrollo de terminales interactivas de consultas, utilizadas en las elecciones de la Facultad de Informática desde el año 2007, para los claustros de alumnos, graduados y profesores [21] [22].
- Utilización permanente de las terminales como medio de consultas de actividades de la Facultad de Informática, por parte de los alumnos.
- Desarrollo de soluciones informáticas con identificación biométrica de personas (huellas dactilares, iris).
- Desarrollo de urnas electrónicas utilizadas para las elecciones de alumnos, docentes y no docentes de la UNNOBA.
- Análisis de sistema de votación distribuida y transmisión vía internet, aplicado a las elecciones del CONICET.
- Desarrollo de un sistema de voto electrónico utilizado en la elección de Cadenas Productivas prioritarias, realizada en el marco del Plan Estratégico Productivo de la provincia de Buenos Aires.

### **3.2. Sistema para gestión académica de carreras de PostGrado. Facultad de Ciencias Económicas. UNLP**

Se ha desarrollado un Sistema que permite la gestión académica integral de carreras de postgrado.

El sistema permite administrar todas las carreras, los diferentes planes de estudio de cada una de ellas, los cursos, docentes, alumnos y todos los requerimientos para lograr un control automático absoluto del estado académico de cada alumno.

La aplicación fue desarrollada sobre una plataforma LAMP (Linux, Apache, MySql y Php) utilizando el framework de desarrollo Symfony.

Además está integrado con un sistema de control de asistencia de alumnos a cursos, utilizando las huellas digitales de los alumnos.

### **3.3 Sistemas para relevamiento de ofertas tecnológicas**

Se han desarrollado sistemas de software relacionados con la gestión de información de centros, laboratorios e institutos de investigación. En ese sentido se puede mencionar:

- Sistema para administrar información de los Grupos de Investigación de la Universidad Nacional de La Plata.
- Sistema para la gestión de la red de vinculación tecnológica (Red ViTec) de las universidades nacionales argentinas.
- Sistema de gestión de grupos de investigación y/o transferencia tecnológica en la provincia de Buenos Aires.

### **3.4. Sistema para la gestión de Unidades de Información del Tribunal de Cuentas de la Pcia. de Buenos Aires.**

- Las teorías actuales marcan una evolución en la concepción de la Administración Pública, enfocándose en la necesidad de Estados inteligentes y eficientes.
- Por otra parte la Gestión del Conocimiento es el conjunto de actividades realizadas con el fin de utilizar, compartir y desarrollar los conocimientos de una organización y de los individuos que en ella trabajan, encaminándolos a la mejor consecución de sus objetivos.
- En este sentido, cobra valor la aplicación de las teorías de Gestión del Conocimiento que permitirán visualizar, compartir y utilizar los recursos intangibles existentes, por parte de los empleados y funcionarios públicos, ciudadanos, empresas y todos los actores sociales en la medida de sus necesidades de conocimiento y comprensión de la realidad, en pos del progreso y

modernización de las Organizaciones Públicas.

- En este contexto se dirigió una Tesina de Grado cuyo objetivo fue la construcción de una plataforma informática colaborativa que permite la gestión sistematizada de Unidades de Información (tales como, jurisprudencia, fallos, normativas, doctrinas) de uso en el Tribunal de Cuentas de la Pcia. de Bs. As.

### 3.5. Consultoría para la Secretaría de Comunicaciones de la Nación.

Se evalúan técnica y económicamente proyectos presentados al Plan Nacional de Servicio Universal, en conjunto con las Fac. de Ingeniería y Ciencias Económicas de la UNLP.

### Formación de Recursos Humanos

Los integrantes de esta línea de investigación dirigen Tesinas de Grado y Tesis de Postgrado en la Facultad de Informática, y Becarios III-LIDI en temas relacionados con el proyecto. Además participan en el dictado de asignaturas/cursos de grado y postgrado de la Facultad de Informática de la UNLP.

### Referencias

[1] G. Coulouris. Distributed Systems – Concepts and Design. Addison-Wesley. 1994.

[2] R. Pressman. Ingeniería de Software: Un Enfoque Práctico. McGraw-Hill. 2002

[3] R. Wysocki. Effective Project Management: Traditional, Adaptive, Extreme, .Wiley .2003

[4] Loucopoulos, P; Karakosas, V.. Systems Requirements Engineering. .McGraw Hill. Book Company. 1995

[5] G. Kotonya and I. Sommerville, Requirements Engineering: Processes and Techniques, Wiley. 1998

[6] Pleegeer. Ingeniería de Software: Teoría y Práctica. Prentice-Hall. 2002

[7] Stephen Kan. Metrics and Models in Software Quality Engineering (2nd Edition). Addison Wesley. 2003

[8] Offutt J., “Quality Attributes of Web Software Applications”. IEEE Software: Special, Issue on Software Engineering of Internet Software 19 (2):25-32, Marzo / Abril 2002.

[9] Wu, Y. y Offutt, J. “Modeling and testing web-based Applications”. <https://citeseer.ist.psu.edu/551504.html>: 1-12, Julio 2004

[10] Silberschatz A et al: “Fundamentos de Bases de Datos”, Tercera Edición Mc Graw Hill 1998

[11] Ozsu M. Valduriez, P.: “Principles of Distributed Database Systems”, Segunda Edición. Prentice Hall 1999

[12] Piattini, M; Oktaba, H; Pino, F; Orozco, M; Alquicira, C. COMPETISOFT. Mejora de Procesos Software para Pequeñas y Medianas Empresas y Proyectos. Editorial RaMa. ISBN; 978-84-7897-901-1. 2008

[13] Elmasri, Navathe, “Fundamentos de Sistemas de Bases de Datos”, Tercera Edición, Pearson, 2006

[15] Everdingen, Y., Hillegersberg, J., Waarts, E. “ERP Adoption by European Midsize Companies”. Communications of the ACM, vol. 43, n°. 4, pp. 27-31. 2000.

[16] Sprott D. “Enterprise Resource Planning: Componentizing the Enterprise Application Packages”. Communications of the ACM, vol. 43, n°. 4, pp. 63-69. 2000.

[18] Estrebou C., Romero A., Galdamez N., Moralejo L. “Sistema Web para Planeamiento de la Producción de una empresa con Plantas distribuidas”. Jornadas AUGM. Campinas (Brasil). 2006.

[19] Feierherd G., De Giusti A., Pesado P., Depetris B. “Una aproximación a los requerimientos del software de voto electrónico de Argentina”. CACIC 2004.

[20] Pesado P., Feierherd G., Pasini A. “Especificación de Requerimientos para Sistemas de Voto Electrónico”. CACIC 2005.

[21] Feierherd G. “Voto por Internet”. Reporte Técnico UNPSJB.

[22] Pesado P., Pasini A., Ibáñez E., Galdámez N., Chichizola F., Rodríguez I., Estrebou C., De Giusti A. “E-Government- El voto electrónico sobre Internet”. CACIC 2008.

[23] Carri J., Pasini A., Pesado P., De Giusti A. “Reconocimiento biométrico en aplicaciones de E-Government. Análisis de confiabilidad / tiempo de respuesta.” CACIC 2007.

[24] Pasini A., Ibáñez E., Galdamez N., Estrebou C., Rodríguez I., Pousa A., De Giusti A. “Análisis Urna Electrónica ALTEC S.E.”. Informe Técnico III-LIDI. 2007.

[25] Estrebou C., Galdamez N., Pasini A., Pousa A., De Giusti A. “Análisis Urna Electrónica ALTEC S.E. Actualización 2008”. Informe Técnico. III-LIDI. 2008.

[26] “Acuerdo de creación de un ESICENTER en Argentina, entre European Software Institute, Grupo Tekne, Universidad Nacional de La Plata, Universidad Nacional de San Martín”. 2007.

[27] Castelli V., Thomas P., Bertone R., Oliveros A., “A Requirements Engineering Process extended to Context Information Management”, RCIS 2011 (Fifth IEEE International Conference on Research Challenges in Information Science). Mayo 2011. IEEE (Print Version ISBN #978-1-4244-8671-7).



- [28] Anda B., Benestad HC, "A Multiple-Case Study of Effort Estimation based on Use Case Points", Empirical Software Engineering 2005, International Symposium.
- [29] Kusumoto, S., Matsukawa, F., Inoue, K., Hanabusa, S., Maegawa, Y.: "Estimating Effort by Use Case Points: Method, Tool and Case Study". In: Proceedings of the 10<sup>th</sup> International Symposium on Software Metrics, pp. 292--299 (2004).

## Análisis comparativo de modelos de calidad orientado al desarrollo de software en pymes

Marianela Llaneza<sup>1</sup>, Gladys Dapozo<sup>1</sup>, Cristina Greiner<sup>1</sup>; Marcelo Estayno<sup>2</sup>

(1)Departamento de Informática. Facultad de Ciencias Exactas y Naturales y Agrimensura  
Universidad Nacional del Nordeste

mary\_llaneza@hotmail.com, {gndapozo,cgreiner}@exa.unne.edu.ar;

(2)Departamento de Informática. Facultad de Ingeniería. Universidad Nacional de Lomas de Zamora  
mestayno@gmail.com

### Resumen

Las pymes de software actualmente constituyen un sector muy importante en el crecimiento informático del país, pero por su organización y tamaño tienen dificultades para acceder a certificaciones de calidad. Para contribuir a mejorar esta problemática, se propone estudiar los modelos de calidad CMMI y COMPETISOFT para proponer una guía de buenas prácticas que orienten el proceso de desarrollo en pymes locales con el objetivo de consolidar criterios de calidad con vistas a futuros mecanismos de certificación.

**Palabras clave:** Calidad de software, Buenas Prácticas, Pymes de software.

### Contexto

La línea de I/D presentada en este trabajo forma parte del proyecto F007-2009: "Modelos y métricas para la evaluación de la calidad de software", acreditado por la Secretaría General de Ciencia y Técnica de la Universidad Nacional del Nordeste (UNNE). El objetivo fundamental del proyecto es contribuir a la mejora en la calidad de los productos software mediante modelos y métricas aplicados al producto y al proceso de creación, diseño, desarrollo y mantenimiento de software, como medio para aumentar la competitividad de las pymes de la región NEA en el contexto de la industria del software.

### Introducción

A lo largo del tiempo han surgido diferentes modelos para evaluar la calidad del software, que intentan descomponer la calidad en una categoría de características más sencillas. El objetivo de estos modelos consiste en mejorar los procesos de desarrollo de modo que los proyectos sean más predecibles en tiempo y costo. Se busca además la reducción de los riesgos en el proceso de desarrollo dado que este afecta directamente al ahorro del costo. En la actualidad, existe una diversidad de modelos, enfocados en diferentes realidades y contextos. Las principales organizaciones que lideran los estándares más difundidos son: la Organización de Normalización

Internacional (ISO) y el Instituto de Ingeniería de Software (SEI), definiendo normas asociadas a los modelos desarrollados.

Por otra parte, las pequeñas y medianas empresas de software, cumplen un papel fundamental en el sector informático del país. Las mismas buscan progresar, y para eso, ven como opción transitar el camino hacia una certificación con la convicción de que la mejora de sus procesos y metodologías de trabajo son un camino más directo y seguro hacia el éxito comercial futuro.

### Calidad de Software

Pressman define la calidad del software como "Concordancia con los requisitos funcionales y de rendimiento explícitamente establecidos, con los estándares de desarrollo explícitamente documentados, y con las características implícitas que se espera de todo software desarrollado profesionalmente" [1].

Al definir el concepto de calidad del software se debe diferenciar entre la calidad del Producto de software y la calidad del Proceso de desarrollo del mismo.

En este trabajo, se hará hincapié en la calidad del proceso, en particular, las cuestiones referida a la mejora de procesos.

### Modelos de Calidad

Un modelo de calidad es un conjunto de buenas prácticas vinculadas a los procesos de gestión y desarrollo de proyectos. Este modelo supone una planificación para alcanzar un impacto estratégico, cumpliendo con los objetivos fijados en lo referente a la calidad del producto o servicio.

La aplicación de modelos de calidad favorece a la mejora continua, establece procesos estándares con insumos y resultados medibles, reduce costos y promueve la eficiencia. Las empresas se ven beneficiadas al poder ofrecer a sus clientes productos de mayor calidad y seguridad cumplidos en los tiempos previstos [2].

Es importante resaltar que en un programa de mejora se involucran diferentes tipos de modelos entre los que se encuentran:

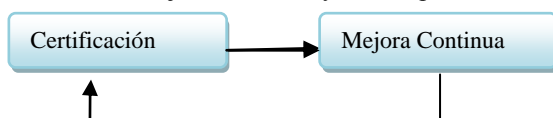
- El modelo que gestiona y conduce la mejora (Modelo de procesos): Describe la

infraestructura, actividades, ciclo de vida y consideraciones prácticas para guiar la mejora de procesos software en una organización.

- El método de evaluación de procesos (Modelo de evaluación): Especifica la ejecución de la evaluación formal para producir un resultado cuantitativo que caracterice el estado de la capacidad del proceso o la madurez de la organización.
- El modelo de referencia de procesos a seguir (Modelo de mejora): Describe qué actividades son reconocidas como las mejores prácticas que una organización debe implementar para la producción de software.

### Importancia de la certificación

Según el IRAM, la certificación es la demostración objetiva de conformidad con normas de calidad, seguridad, eficiencia, desempeño, gestión de las organizaciones y buenas prácticas de manufactura y comerciales. La certificación contribuye al desarrollo tecnológico de las organizaciones, genera un mejor posicionamiento, facilita la apertura de nuevos mercados [3]. Es la puerta de entrada a la Mejora Continua y la Competitividad.



Las certificaciones en los últimos años se han centrado en la aplicación de dos normas internacionales: el Modelo CMMI (*Capability Maturity Model Integration for Development-CMMI-DEV*), y la Norma Internacional ISO 9001. Si bien ha aumentado notablemente la cantidad de empresas certificadas con la Norma ISO 9001 y el Modelo CMMI, las mismas presentan diversas dificultades en su implementación. En cuanto a las Norma ISO 9001, certifica el Sistema de Gestión de la Calidad en una organización y puede ser aplicada en cualquier empresa tanto manufacturera como de servicios, en tanto que la Norma ISO 90003 de Gestión de la Calidad y aseguramiento de la Calidad, define las directrices para la aplicación de la norma ISO/IEC 9001 para el desarrollo, suministro, instalación y mantenimiento del software [4], [5].

En relación al Modelo CMMI su adopción conlleva costos muy elevados y resulta difícil de implementar por parte de las pymes del sector, debido a su organización y tamaño. Como consecuencia, las empresas tienden a elegir el uso de otro modelo de mejora, o en el peor caso, ninguno [6].

Existen distintas iniciativas que promueven modelos de calidad orientados a las pymes que pretenden abordar la mejora de procesos.

Además del modelo CMMI y la norma ISO 9001, podemos mencionar MOPROSOFT basado en ISO

12207, y el método de evaluación de procesos para la industria de software (EvalProSoft), basado en ISO 15504.

Para Iberoamérica, se ha desarrollado el proyecto COMPETISOFT, basado en ISO 12207, ISO 15504, CMMI, MANTEMA, Métrica V3, Agile SPI, y principalmente en MoProsoft y EvalProsoft.

### Organizaciones que avalan las certificaciones

El Modelo CMMI está avalado por el SEI (*Software Engineering Institute*) y la Carnegie Mellon University.

El SEI es quien habilita oficialmente a una persona para funcionar como “*Lead Appraiser*” o evaluador. El *Lead Appraiser* junto con un Team entrenado especialmente (que generalmente incluye a recursos de la propia compañía) es quien hace la evaluación, y para la misma, se utiliza el método SCAMPI.

En el caso de las Normas ISO, conforma una red de organizaciones nacionales de normalización de 156 países (siendo miembro uno por cada país); y desde su sede central se coordina todo el sistema.

En Argentina se cuenta con el instituto IRAM (Instituto Argentino de Normalización y Certificación), donde se participa activamente en 75 comités [4].

A continuación, se realizará una descripción de los estándares a estudiar en el presente trabajo:

### CMMI

CMMI (*Capability Maturity Model Integration*) es un modelo de madurez de mejora de los procesos para el desarrollo de productos y de servicios. Abarca un conjunto de las mejores prácticas que tratan las actividades de desarrollo y de mantenimiento que cubren el ciclo de vida del producto, desde la concepción a la entrega y el mantenimiento.

El propósito de CMMI para desarrollo es ayudar a las organizaciones a mejorar sus procesos de desarrollo y de mantenimiento, tanto para los productos como para los servicios.

Para dar soporte a aquellos que utilizan la representación por etapas, todos los modelos CMMI reflejan niveles de madurez en su diseño y contexto. Un nivel de madurez consta de prácticas relacionadas específicas y genéricas para un conjunto predefinido de áreas de proceso que mejoran el rendimiento global de la organización. El nivel de madurez de una organización proporciona un camino para predecir el rendimiento en una disciplina dada o en un conjunto de disciplinas.

La experiencia ha mostrado que las organizaciones toman la mejor decisión cuando centran sus esfuerzos de mejora de procesos en un número controlable de áreas de proceso a la vez y que

dichas áreas requieren aumentar su complejidad cuando la organización mejora.

Un nivel de madurez es una meseta evolutiva definida para la mejora de procesos de la organización. Cada nivel de madurez madura un subconjunto importante de procesos de la organización, preparándola para pasar al siguiente nivel de madurez. Los niveles de madurez se miden mediante el logro de metas específicas y genéricas asociadas a cada conjunto predefinido de áreas de proceso.

Un “área de proceso” es un conjunto de mejores prácticas relacionadas a un área, que cuando se implementan colectivamente satisfacen un conjunto de objetivos considerados importantes para lograr una mejora significativa en esa área.

Existen cinco niveles de madurez, siendo cada uno de ellos una capa en la cimentación de la mejora de procesos en curso, denominados por los números 1 a 5.

- Inicial.
- Gestionado.
- Definido.
- Gestionado cuantitativamente.
- En optimización.

Los niveles de madurez 2 a 5 utilizan los mismos términos que los niveles de capacidad 2 a 5. Esto se ha hecho de forma intencionada porque los conceptos de los niveles de madurez y niveles de capacidad son complementarios. Los niveles de madurez se utilizan para caracterizar la mejora de la organización relativa a un conjunto de áreas de proceso, y los niveles de capacidad caracterizan la mejora de la organización relativa a un área de proceso individual.

#### *Nivel de madurez 1: Inicial*

La organización generalmente no proporciona un entorno estable para dar soporte a los procesos. El éxito en estas organizaciones depende de la competencia del personal de la organización y no del uso de procesos probados. A pesar de este caos, las organizaciones de nivel de madurez 1 a menudo producen productos y servicios que funcionan; sin embargo, frecuentemente exceden sus presupuestos y no cumplen sus calendarios.

#### *Nivel de madurez 2: Gestionado*

En el nivel de madurez 2, los proyectos de la organización han asegurado que los procesos se planifican y realizan de acuerdo a políticas; los proyectos emplean personal con habilidad que dispone de recursos adecuados para producir resultados controlados; involucran a las partes interesadas relevantes; se monitorizan, controlan y revisan; y se evalúan en cuanto a su adherencia a sus descripciones de proceso. Énfasis en el proceso.

#### *Nivel de madurez 3: Definido*

En el nivel de madurez 3, los procesos son bien caracterizados y comprendidos, y se describen en estándares, procedimientos, herramientas y métodos. El conjunto de procesos estándar de la organización, que es la base del nivel de madurez 3, se establece y mejora a lo largo del tiempo. Énfasis en la estandarización de los procesos.

#### *Nivel de madurez 4: Gestionado cuantitativamente*

En el nivel de madurez 4, la organización y los proyectos establecen objetivos cuantitativos en cuanto al rendimiento de calidad y del proceso, y los utilizan como criterios en la gestión de los procesos. Los objetivos cuantitativos se basan en las necesidades del cliente, usuarios finales, organización e implementadores del proceso. El rendimiento de calidad y del proceso se comprende en términos estadísticos y se gestiona durante la vida de los procesos. Énfasis en la predictibilidad del rendimiento del proceso, utilizando técnicas cuantitativas.

#### *Nivel de madurez 5: En optimización*

En el nivel de madurez 5, una organización mejora continuamente sus procesos basándose en una comprensión cuantitativa de las causas comunes de variación inherentes a los procesos.

El nivel de madurez 5 se centra en mejorar continuamente el rendimiento de procesos mediante mejoras incrementales e innovadoras de proceso y tecnológicas [7].

### **COMPETISOFT**

El Modelo de Procesos de COMPETISOFT está dirigido a las empresas o áreas internas dedicadas al desarrollo y/o mantenimiento de software. Las organizaciones que no cuenten con procesos establecidos pueden usar el modelo como la primera versión de sus procesos e ir ajustándolo de acuerdo a sus necesidades y la experiencia adquirida. Mientras que las organizaciones, que ya tienen procesos establecidos, pueden usarlo como punto de referencia para identificar los elementos que les hace falta cubrir.

El mismo, fue desarrollado y mejorado por personas que tienen amplio conocimiento en el contenido de los modelos internacionales y cuentan con la experiencia en la implantación de estos modelos en PyMEs. Los requisitos para su elaboración se definieron a partir de las entrevistas con personas que trabajan en la industria de desarrollo de software.

El modelo está conformado por tres categorías que agrupan procesos de acuerdo a la estructura típica de una organización:

Categoría de Alta Dirección: establece la razón de ser de la organización, lo que se desea alcanzar y las estrategias para lograrlo.



Categoría de Gerencia: establece planes de acción para instrumentar las estrategias en cuanto a proyectos, procesos y recursos necesarios realiza monitoreo de la categoría de Operación y retroalimenta a la categoría de Alta Dirección.

Categoría de Operación: realiza proyectos de desarrollo o mantenimiento de software que cubran necesidades del cliente en el tiempo y costo esperados y reporta los resultados a la categoría de Gerencia [8].

### **Comparación y Análisis entre Modelos de Calidad**

Todos los estándares comparten casi de manera equivalente principios y criterios, estando en permanente revisión y mejora. Difieren en algunos contenidos dentro de cada criterio, y algunos se encuentran más actualizados que otros, pero comparten objetivos comunes: Aprendizaje, innovación y mejora continua.

Se han encontrado muchas comparaciones entre los diferentes modelos y normas de calidad, entre otros, podemos mencionar:

En [9], se realiza un modelo unificado entre ISO 9001 y CMMI, el cual sirve como ayuda a empresas certificadas con normas ISO, que en un futuro quieran implementar CMMI.

En [10], se elabora un análisis comparativo entre COMPETISOFT e ISO 9001:2000. Su propósito es examinar las posibilidades de que pequeñas empresas guiadas por la mejora de procesos de COMPETISOFT puedan obtener una certificación ISO 9001.

Siguiendo la misma línea de comparaciones, se analiza las características coincidentes entre el modelo COMPETISOFT y CMMI:

Ambos modelos caen en la categoría de modelos descriptivos, puesto que no prescriben métodos o prácticas específicas para realizar las diversas actividades. Sin embargo se percibe que al definir una estructura de procesos, con actividades, roles y caracterización de productos, el modelo COMPETISOFT avanza un poco en prescribir la manera de diseñar los procesos y por lo tanto representa una guía útil para una organización.

Para justificar que el modelo COMPETISOFT es un apoyo a la implementación de CMMI, se hará unas consideraciones:

CMMI define que los procesos deben estar alineados para cumplir los objetivos estratégicos pero no establece un proceso para definirlos. COMPETISOFT considera el proceso de Gestión de Negocio para definir la visión estratégica de la organización.

La descripción de los roles con sus competencias y los diversos artefactos con sus contenidos, así como el patrón de procesos que propone el modelo de COMPETISOFT, es una referencia útil para la revisión de la manera como la organización define sus procesos.

El tratamiento que da el modelo COMPETISOFT a las actividades que son transversales a diferentes procesos (Verificación, Validación, Mediciones, Gestión de Configuración) de articularlas a los diferentes procesos puede verse como una ventaja del modelo y a la vez como una debilidad. En ventaja en el sentido en que, al encontrarse ya articuladas en los respectivos procesos, las organizaciones encuentran más natural la aplicación de estas prácticas. Sin embargo, si se apunta a niveles de madurez CMMI iguales o superiores a 3, se ve muy conveniente que estas actividades puedan ser formalizadas y definidas como procesos de apoyo que establecen unas políticas y lineamientos para toda la organización [8].

La idea esencial del presente trabajo es realizar un análisis comparativo entre CMMI y COMPETISOFT, y poder extraer las buenas prácticas preponderantes entre ambos modelos.

### **Líneas de investigación y desarrollo**

En el marco de la línea de investigación se plantean las siguientes actividades:

- Estudio del modelo Competisof y elaboración de una herramienta de autoevaluación para pymes de software basada en el modelo.
- Análisis comparativo de los principales modelos de calidad (CMMI y COMPETISOFT), a fin de extraer las buenas prácticas para la elaboración de una guía para el desarrollo de software.
- Aplicación de la guía de buenas prácticas en alguna de las pymes desarrolladoras de software a fin de retroalimentar la propuesta con vista a consolidar una guía automatizada para ser transferidas a las empresas.

### **Resultados y Objetivos**

Se elaboró una aplicación web con el objetivo de ofrecer a las empresas de software un mecanismo de autoevaluación basado en el modelo COMPETISOFT. La aplicación se concentra en la parte operacional del modelo Competisof, presentando el cuestionario de auto-evaluación de los procesos Administración de un Proyecto Específico y Desarrollo de Software y Mantenimiento de Software.

Para evaluar el comportamiento de la aplicación, se solicitó la colaboración de empresas desarrolladoras de software de la ciudad de Corrientes.. Previo a la evaluación, se realizó una entrevista con los responsables para comentarles acerca del objetivo del trabajo y del modelo de calidad Competisof. Los resultados obtenidos se detallan en [11].

De la evaluación realizada, se pudo observar que en los procesos de Desarrollo y Mantenimiento de Software se obtuvieron niveles de capacidad que dan cuenta de la incorporación de buenas prácticas, mientras que en el proceso de Administración de Proyectos Específicos obtuvieron el nivel Realizado, que denota una incipiente incorporación de buenas prácticas. Estos resultados son compatibles con otros estudios que indican que las empresas ponen mayor énfasis en los procesos propios de la ingeniería de software y no tanto en los procesos relacionados con la administración de los mismos [12].

En la actualidad se trabaja en un análisis comparativo de las prácticas propuestas por los modelos COMPETISOFT y CMMI, para conformar una guía de buenas prácticas para aplicar al desarrollo de software en pymes locales.

### Formación de Recursos Humanos

En esta línea de trabajo se desarrolla el plan de actividades de una becaria de investigación de pregrado de la Secretaria de Ciencia y Técnica de la UNNE (SECYT-UNNE), bajo la supervisión de docentes integrantes del proyecto.

### Referencias

1. Pressman, R.S. (1998) Ingeniería del Software. Un enfoque práctico. Cuarta edición. McGraw-Hill. Madrid.
2. Estayno, M.; Dapozo, G.; Greiner, C.; Cuenca Pletch, L.; Pelozo, S. (2009) CACIC. Congreso Argentino de Ciencias de la Computación. Caracterización de las Pymes de software de la región NEA orientada hacia un marco de la mejora de calidad. Universidad Nacional del Nordeste.
3. IRAM. Instituto Argentino de Normalización y Certificación. Consulta: 5 de marzo del 2013. Disponible en: <http://www.iram.org.ar/seccion.php?ID=3>
4. Calidad del Software. Preguntas Frecuentes. Cámara de Empresas de Software & Servicios Informáticos de la República Argentina. Consulta: 7 de marzo de 2013. Disponible en: [www.cessi.org.ar/documentacion/Calidad-Software-faq.doc](http://www.cessi.org.ar/documentacion/Calidad-Software-faq.doc)
5. ISO/IEC 90003:2004. Quality management and quality assurance standards. Part 3: Guidelines for the application of ISO 9001 to the development, supply, installation and maintenance of computer software. International Organisation for Standardization, 2004.
6. Staples, M.; Niazi, M.; Jeffery, R.; Abrahams, A.; Byatt, P.; Murphy, R. An exploratory study of why organizations do not adopt CMMI. *Journal of Systems and Software*, Volume 80, Issue 6, June 2007, Pages 883-895, ISSN 0164-1212, 10.1016/j.jss.2006.09.008. (<http://www.sciencedirect.com/science/article/pii/S0164121206002573>)
7. Chrissis, M. B.; Konrad, M.; Shrum, S. (2009). CMMI, Guía para la integración de procesos y la mejora de productos. Pearson Educación. ISBN: 9788478290963
8. Oktaba, H.; Piattin, M.; Pino, F. J.; Orozco, M. J.; Alquicira, C. (2009).COMPETISOFT. Mejora de procesos para pequeñas y medianas empresas y proyectos. Alfaomega. México. ISBN: 9786077686170
9. Chanwoo, Y.; Junho, Y.; Byungjeong, L.; Chongwon, L.; Jinyoung, L.; Seunghun, H.; Chisu, W. (2006). A unified model for the implementation of both ISO 9001:2000 and CMMI by ISO-certified organizations. School of Computer Science. University of Seoul. South Korea.
10. Esponda, S.; Pasini, A.; Pesado, P.; Borachia, M. (2012) CACIC. Congreso Argentino de Ciencias de la Computación. ISO 9001 in software-developing VSEs assisted by the COMPETISO FT model. Universidad Nacional de La Plata.
11. Leonardo Fernández, Gladys Dapozo, Cristina Greiner. "Aplicación para la autoevaluación de Capacidad de Proceso orientada a pymes software del NEA". Anales del 1º Seminario Argentina Brasil de Tecnologías de la Información y la Computación (SABTIC2011). ISSN: 2237-2970. Pontificia Universidad Católica. Rosario. Santa Fe. 3 y 4 de Noviembre 2011. <http://www.uca.edu.ar/index.php/site/index/es/uca/facultades/rosario/quimica-ingenieria/congreso/>
12. Pino, F., García, F., Piattini (2009). M. Key processes to start software process improvement in small companies. SAC'09, March 8-12, 2009, Honolulu, Hawaii, U.S.A. Copyright 2009 ACM 978-1-60558-166-8/09/03.

# Plantilla CUPIDO: Automatización y avances

Laura Aballay<sup>1</sup>, María Inés Lund<sup>1</sup>, Emilio Ormeño<sup>1</sup>, Sabrina Cruz Introini<sup>1</sup>, Cecilia Marcuzzi, Gerardo Jofré

Instituto de Informática – Fac. Cs. Exactas F. y Naturales –  
Universidad Nacional de San Juan

Ignacio de la Roza y Meglioli. Complejo CUIM. UNSJ  
Tel 0264-4265101

<sup>1</sup> laballay, mlund, eormeno, sci{ @iinfo.unsj.edu.ar}

## Resumen

Las plantillas de casos de uso CUPIDO se están convirtiendo, dentro de las carreras de Licenciatura en Ciencias de la Computación y Licenciatura en Sistemas de Información de la FCEFN de la UNSJ, en un artefacto de uso cotidiano para facilitar la documentación de un proyecto de software, tanto en las cátedras como en sus proyectos de software para sus tesis de licenciatura. Por lo tanto, automatizar la documentación de las plantillas de casos de uso incorporando en el flujo de eventos esta sintaxis, permitirá no solo documentar sino generar en forma semiautomática los diagramas de secuencia, de comunicación, hasta un modelo de clases inicial del sistema.

**Palabras clave:** Plantillas, Documentación de Casos de Uso, Flujo de Eventos, Escenario, Patrones de Diseño.

## Contexto

Se presenta el estado de avance del proyecto de investigación ForCUPIDO-“Formalización de Descripciones de Casos de Uso a Través de Metamodelos” (Proy N° E-883), de la FCEFN de la UNSJ (Argentina). El proyecto, aprobado por evaluación externa y financiado por la UNSJ, se encuentra inserto en el marco de las líneas de investigación del Gabinete de Ingeniería de Software del Instituto de Informática. Además se encuentra vinculado a cátedras de las carreras de Licenciatura en Ciencias de la Computación y Licenciatura en Sistemas de Información, de la FCEFN de la Universidad Nacional de San Juan.

## Introducción

Actualmente, si se desea construir una aplicación de tamaño considerable, un componente fundamental en el proceso de desarrollo de software es un diseño consistente y bien documentado. Es necesario contar con modelos del sistema que acompañen y guíen el proceso de desarrollo. Dentro de las herramientas de modelado de un sistema encontramos los diagramas de casos de uso. Ivar Jacobson, creador de los casos de uso, indica que además de ser un recurso para capturar requisitos, ayudan en la delimitación del sistema y son un medio de comunicación entre los usuarios finales y clientes, ya que permiten describir de forma clara para ambas partes (cliente-desarrollador) las funcionalidades de un sistema [1]. Ésta es una de sus principales ventajas ya que resultan muy fáciles de entender por el cliente y posibilita llegar a un acuerdo con el desarrollador sobre los requisitos del sistema; sin embargo no siempre brindan la precisión necesaria si no se acompañan con una descripción textual u otro diagrama que lo detalle. Generalmente esta descripción se expresa en plantillas que permiten guiar, portar o construir un esquema predefinido.

Las plantillas de casos de uso son documentos estructurados que facilitan el proceso de documentación de un proyecto de software así como la descripción de los requisitos funcionales. Es necesario que un caso de uso sea bien descrito para que sea fácil de leer y de interpretar. En base a esta necesidad y al contexto de desarrollo de software local y regional, este grupo de investigadores trabajó en proyectos anteriores donde se analizó y elaboró una nueva plantilla que integra diferentes

aspectos considerados en distintas plantillas examinadas y se la denominó CUPIDO (Casos de Uso, Plantilla Integradora para Documentarlos) [2].

El Flujo de Eventos (FE), también llamado escenario, es la parte principal de la plantilla de casos de uso, por lo que merece especial cuidado y atención. Está compuesto por una serie de sentencias escritas en una sola forma gramatical y donde cada oración es un paso, a través del cual un actor llega al resultado esperado.

Motiva esta investigación la dificultad en la construcción del FE evidenciada en alumnos de 4º año de la materia “Diseño de Software” de las carreras de Informática de la Facultad de Ciencias Exactas Físicas y Naturales de la Universidad Nacional de San Juan (Argentina). En esta materia los alumnos utilizan plantillas para describir sus casos de uso en las clases prácticas, desde el año 2005, y desde el año 2010 usan la plantilla CUPIDO, generada en el 2009 en el proyecto CUProSoft (Proy N° E-822) por este mismo grupo de investigación.

La experiencia de todos estos años demostró que, al momento de documentar en las plantillas, los estudiantes expresan tener confusión y dificultades al completar el campo referido al FE, dato que fue corroborado a través de encuestas de usabilidad de las plantillas. Investigando este conflicto, se plantea la hipótesis de que puede ser debido a la falta de estructuración y guía para completar el FE. Por este motivo se propone definir una sintaxis que favorezca a este fin.

Se plantea estructurar el FE, a través de reglas gramaticales que guíen y formalicen su escritura, por medio de una sintaxis que se denominará UCEFlow (Estructuración del Flujo de Eventos de Casos de Uso, se lee “iu si i flou”, que sería como decir “you see the e-flow” o tú ves el flujo de eventos). Expresar el FE con esta sintaxis posibilitará la generación semiautomática de los diagramas de secuencias y, en un futuro, por medio de transformaciones automáticas obtener la prototipación del sistema, fundamentándose en la teoría de Desarrollo de Software Dirigido por Modelos (MDD)[3].

Las plantillas de casos de uso CUPIDO se están convirtiendo, dentro de las carreras de Licenciatura en Ciencias de la Computación y Licenciatura en Sistemas de Información de la FCEFN de la UNSJ, en un artefacto de uso cotidiano para facilitar la documentación de un proyecto de software, tanto en las cátedras como en sus proyectos de software para sus tesis de licenciatura. Por lo tanto, automatizar la documentación de las plantillas de casos de uso incorporando en el flujo de eventos esta sintaxis, permitirá no solo documentar sino generar en forma semiautomática los diagramas de secuencia, de comunicación, hasta un modelo de clases inicial del sistema.

En esta línea se está trabajando, y actualmente en proceso de implementación de un software que permita documentar un proyecto de software a través de las plantillas CUPIDO. Se pretende utilizar técnicas que permitan optimizar el desarrollo de software. Partiendo de la definición de nombres de cada uno de los elementos del sistema, se utilizará un framework para estandarizar la forma de trabajo, el uso de mapeos para los objetos de negocio, lo cual permitirá lograr una mayor independencia de la capa de negocios respecto de la capa de acceso a datos. El framework trabajará mediante el esquema MVC (Model View Controller) logrando que la aplicación pueda ser desarrollada en paralelo, es decir que una vez definidos la nomenclatura de los elementos del sistema y el esquema de trabajo, fácilmente puedan repartirse en diferentes partes del sistema a desarrollar; en este caso, tres componentes fundamentales, el modelo, la vista y el controlador. Buenas prácticas sugieren controladores delgados y modelos anchos, esto quiere decir que toda la lógica de negocio debe ser manejada por el modelo, para así permitir al controlador manejar peticiones fácilmente, esto ha sido tenido en cuenta en el desarrollo de este proyecto. El sistema se está desarrollando en PHP, como motor de base de datos se está utilizando MySQL, y el framework de base es Zend. También se incorporan algunos lenguajes de soporte como, JS, CSS etc.

En forma paralela, la estructuración del flujo de eventos a través de la sintaxis generada para tal fin, se encuentra en una etapa avanzada de definición. La sintaxis se planteó en

## Líneas de investigación y desarrollo



(Extended Backus-Naur Form) que comprende todos aquellos conceptos relevantes a un flujo de eventos, y se realizó un modelo conceptual en UML de la estructuración del FE. En un principio en el desarrollo de la investigación, el lenguaje generado se comenzó probando con un graficador léxico denominado ANTLR el cual generaba un árbol de derivación de un FE ingresado, a partir de la sintaxis propuesta. Luego de varias pruebas y debido a la poca flexibilidad del graficador para aceptar los conceptos definidos por el grupo de investigación se comenzó a trabajar con un parser desarrollado para tal fin en lenguaje JAVA en el IDE NetBeans, el cual brindó mayor flexibilidad, ya que se pudieron introducir cambios en el parser para que acepte la estructuración que se formuló. Con esto se consiguió analizar el FE y obtener un árbol de derivación que, si bien no es gráfico, se puede observar claramente la identificación de elementos claves dentro de un FE como por ejemplo actores, sistema, clases, acciones. A partir de esta identificación se pretende desarrollar una aplicación en la cual se ingrese el FE de la plantilla de casos de uso CUPIDO y se obtenga la clasificación de elementos, que sirva a los alumnos para tener desde la redacción del FE una somera idea de las posibles clases que corresponden al sistema que se desea desarrollar. Por supuesto que esta clasificación que realizará la herramienta no es definitiva, ya que el trabajo de modelado y diseño de un sistema es algo subjetivo, que depende mucho de la persona que lo realiza y por lo tanto requerirá que se realice un posterior análisis y depuración de las clases identificadas por parte del analista o el diseñador.

Por otro lado, se incorporaron patrones de usabilidad y de interacción, a la última versión de la plantilla CUPIDO, para considerar los aspectos de la interfaz de usuario, en etapas tempranas del desarrollo de software. Se incorporó un nuevo campo en la plantilla, para permitir identificar los patrones a aplicar en el diseño de la interfaz de ese caso de uso. Fue necesario indicar cómo esos patrones identificados se incorporan en el flujo de eventos de cada una de las plantillas de los casos de uso. Para tal fin se generó una nueva versión de la plantilla CUPIDO, obteniendo

como resultado la versión 1.5 de la plantilla, denominada “CUPIDO+Pi”. Actualmente, se está realizando la búsqueda y recopilación de documentación de nuevos patrones de diseño de interfaz existentes, para extender la antología ya generada. Se espera en 2013 poder difundir los patrones de interfaz a través de un taller a alumnos de 4° y 5° año de las carreras de LSI y LCC, en el mes de mayo. Como también generar una implementación de referencia de patrones de interfaz para ejemplos de aplicación.

## Resultados y Objetivos

Los miembros del proyecto asistieron y aprobaron los siguientes cursos y congresos afines a la temática del proyecto:

- Curso de Posgrado – “Metodología de la Investigación. Redacción Científica”. Dictado por el Dr. Gerardo Matturro, Universidad ORT de Uruguay, en la FCEFNU-UNSJ. Aprobado 45hs. L. Aballay, E. Ormeño, M.I. Lund.
- Curso de Posgrado – “Modelado de Procesos de Desarrollo de Software”. Dictado por el Dr. Julio Ariel Hurtado Alegría, Universidad del Cauca, Colombia, en la FCEFNU-UNSJ. Aprobado 45hs. L. Aballay, E. Ormeño, M.I. Lund.
- Curso de posgrado del Doctorado de la Universidad de La Plata “Interfaces Avanzadas” modalidad On-Line. L. Aballay
- Realización de curso “Cima Enterprise Fundamentals” – (Agosto-Octubre) – Dictado por la empresa CIMA Tech. Sabrina Cruz Introini y Cecilia Marcuzzi
- Realización curso de “Desarrollo de sitios web dinámicos con PHP y MYSQL” - (Agosto- Noviembre) – Dictado por la UNSJ – Facultad de Ingeniería. S. Cruz Introini y C. Marcuzzi
- Aprobado Examen Final del último nivel. Intensivo de Adultos – año 2012. Título: Capacitación en Idioma Inglés. Inst. Saint Thomas – Resolución N° 3.310. C. Marcuzzi.
- XVIII Congreso Argentino de Ciencias de la Computación (CACIC) – Octubre 2012 Universidad Nacional del Sur – Bahía Blanca. S. Cruz Introini y M.I. Lund.

Se realizaron las siguientes publicaciones en congresos y revistas del área:

- "Formalización de Descripciones de Casos de Uso a través de Metamodelos (ForCUPIDo)". Poster en WICC 2012 (Argentina). Autores: L.Aballay, MI.Lund, E.Ormeño, S.Oviedo, C.Marcuzzi, S.Cruz Introini, V.Alferillo. ISSN 978-950-766-082-5
- "An UML profile for modeling RESTful services". ASSE 2012 - 13th Argentine Symposium on Software Engineering – 41 JAIIO. Autores: E.Ormeño, MI.Lund, L.Aballay, S.Aciar. La Plata, Agosto. ISSN: 1850-2792
- "CODILA+A. Modelo de apoyo para la preparación de actividades experimentales destinadas a la enseñanza de Ingeniería de Software en ambientes colaborativos y distribuidos geográficamente". I Workshop de Innovación en Educación en Informática", dentro de CACIC 2012. Autores: MI.Lund, C.Collazos, E.Ormeño. Bahía Blanca, 9 al 11 de octubre. E-book: ISBN 978-987-1648-34-4
- Ponencia: "Patrones de Interacción aplicados a Plantillas de casos de uso y Diseño de interfaces" en el 1º Encuentro de Investigadores Jóvenes de San Juan y 2º Jornada de Becarios de la UNSJ. Res N° 242. 13 y 14 de Diciembre 2012. Cecilia Marcuzzi
- Ponencia: "Formalización del flujo de eventos de casos de uso. Identificación de conceptos del dominio de descripción de casos de uso" en el 1º Encuentro de Investigadores Jóvenes de San Juan y 2º Jornada de Becarios de la UNSJ. Res N° 242. 13 y 14 de Diciembre 2012. Sabrina Cruz Introini

Se dictaron los siguientes cursos:

- Curso Capacitación de Posgrado "Desarrollo Dirigido por Modelos con Spring Roo". 45 hs. P.Resp: Emilio G.Ormeño, M.I.Lund. Colaboradores: Laura Aballay y Sabrina Cruz Introini. Res N° 456/2012-CEF. N° 44/2012-CD-FCEF. N.

Objetivos a Futuro:

- Realizar pruebas del Software para

automatizar la plantilla CUPIDo con las restricciones de formalización impuestas por el lenguaje.

- Someter la primera versión del lenguaje de FE a referato nacional/internacional.
- Adaptación de la primera versión del lenguaje a la plantilla CUPIDo.
- Implementación, en lenguajes estándares, de transformadores para la generación de modelos de menor nivel de abstracción tales como diagramas de secuencias, a partir del flujo de eventos de los casos de uso.
- Someter la implementación de los métodos y estrategias de transformación a Referato nacional/internacional.
- Implementación de nuevos transformadores específicos a plataformas tales como J2EE o .NET.
- Difusión de los resultados, experiencias y recomendaciones en publicaciones con referato.

Los resultados parciales y finales del proyecto tienen transferencia directa a los alumnos de grado, especialmente los alumnos cursantes de la cátedra Diseño de Software, de las carreras Licenciatura en Ciencias de la Computación (LCC) y Licenciatura en Sistemas de Información (LSI), en donde utilizan como artefacto de especificación de requisitos, las plantillas de Casos de Uso. Se planea dictar cursos o talleres para mostrar nuevas técnicas para el diseño, como las brindadas por MDD, y los resultados obtenidos en el proyecto.

## Formación de Recursos Humanos

- Obtención del Título Magister en Informática, Universidad Nacional de La Matanza. Tesis: "Modelo de apoyo para la preparación de actividades experimentales destinadas a la enseñanza de Ingeniería de Software en ambientes colaborativos y distribuidos geográficamente". Autor: María Inés Lund. Director: Dr. César Collazos. Co-Director:

Mg. Emilio Ormeño.

Tres alumnos de la carrera Licenciatura en Ciencias de la Información, adscriptos al proyecto, poseen becas y trabajan para realización de su tesis de grado:

- Beca de Investigación y Creación categoría Estudiantes Avanzados. Tema: “Caracterización del flujo de eventos de casos de uso. Identificación y generación de una gramática formal”, alumna Sabrina Cruz Introini, para ser desarrollada en el Instituto de Informática de la Universidad Nacional de San Juan. Directora: M.I. Lund.

- Beca TIC – UNSJ, para finalización de estudios de la alumna Sabrina Cruz Introini, para ser desarrollada en el Instituto de Informática de la UNSJ. Directora: M.I. Lund

- Trabajo Final de Tesis: “Caracterización del flujo de eventos de casos de uso. Identificación y generación de una gramática formal”. S. Cruz Introini.

- Beca TIC – UNSJ, Tema "Antología de patrones de usabilidad para el diseño de interfaces. Casos de Estudio”, para finalización de estudios de la alumna Verónica Cecilia Marcuzzi, para ser desarrollada en el Instituto de Informática de la UNSJ. Directora: M.I. Lund.

- Trabajo Final de Tesis: "Antología de patrones de diseño de interfaces. Implementación de referencia de patrones para ejemplos de aplicación.". C.Marcuzzi.

- Beca TIC – UNSJ, para finalización de estudios del alumno Gerardo Jofré, para ser desarrollada en el Instituto de Informática de la UNSJ. Director: E.Ormeño.

## Referencias

- [1] I. Jacobson, G. Booch, y J. Rumbaugh, «Chapter 1: The Unified Process: Use-Case Driven, Architecture-Centric, Iterative, and Incremental», in *The Unified Software Development Process*, Addison-Wesley Professional, 1999, p. 5.
- [2] M. I. Lund, C. Ferrarini, L. Aballay, y E. Meni, «CUPIDO - Plantilla para Documentar Casos de Uso», presented at the V Congreso de Tecnología en Educación y Educación en Tecnología, El Calafate, Santa Cruz, Argentina, 2010.
- [3] C. Pons, R. Giandini, y G. Pérez, *Desarrollo de Software Dirigido por Modelos. Conceptos teóricos y su aplicación práctica*. EDUNLP and McGraw-Hill Education, 2010.

## REFACTORING: SECUENCIA DE APLICACIÓN

(<sup>1</sup>) Ing. Pablo Vilaboa; (<sup>2</sup>) Dra. Alejandra Garrido;

(<sup>1</sup>) CAETI, Facultad de Tecnología Informática, UAI; (<sup>2</sup>) LIFIA, Fac. Cs. Exactas, UNLP;

(<sup>1</sup>) Pablo.vilaboa@uai.edu.ar; (<sup>2</sup>) garrido@sol.info.unlp.edu.ar;

### RESUMEN

Este trabajo intenta demostrar la posibilidad de diseñar un orden lógico o una metodología coherente que consienta la teoría para formalizar algún criterio práctico o racional que determine un orden en la ejecución de refactoring definido en un catálogo. Dicha secuencia debe asegurar transformaciones sucesivas respetando cambios legibles y mantenibles que aseguren los niveles de comprensión de modo tal que el responsable de transformar el código legacy pueda lograr un resultado óptimo al culminar el proceso.

**Palabras Claves:** Refactoring, Fortran, Secuencialidad, legibilidad, Mantenimiento

### CONTEXTO

El trabajo descrito se encuentra enmarcado en la aplicación de un enfoque sistémico, disciplinado y cuantificable del mantenimiento del software subscripta a la línea de investigación de Algoritmos y software del centro de altos estudios en tecnología informática.

### INTRODUCCIÓN

La claridad del código fuente está íntimamente vinculado con la comprensión, el grado de claridad es directamente proporcional a la distancia que existe entre el desarrollador del código y quien lo interpreta. Por lo tanto la evaluación que pudiese hacerse sobre la claridad del código tendría un alto grado de susceptibilidad, y es con el uso de refactoring que se puede disminuir la brecha de susceptibilidad. Todos los profesionales que se desempeñan en el área de sistemas han experimentado horas de frustración al heredar soluciones

informáticas. El mantenimiento y la escalabilidad del producto logran involucrar miles de horas hombre para alcanzar muchas veces soluciones inútiles. En la actualidad los sistemas por diferentes motivos se vuelven ilegibles y difíciles de mantener, sobre todo cuando los desarrolladores son algo indiferentes a las arquitecturas de diseño, buenas prácticas y criterios adecuados de programación. El código producido en un ambiente poco estandarizado se convierte en código inmanejable (Foote & Yoder, 2000)

Foote en su trabajo titulado "*Big ball of Mud*" expone que las empresas deben conocer las causas que comprometen la ocurrencia de sistemas enmarañados y buscar los caminos alternativos para solucionarlos. El autor trata de encontrar una respuesta a la realidad del por qué buenos programadores terminan creando programas ilegibles, y es allí donde recae la necesidad de encontrar esquemas que permitan recuperar el código fuente de los sistemas de información de la empresa.

El código heredado, es algo que posiblemente haya pasado de moda, pero funciona y la organización no desea desprenderse de él. El software legado no necesariamente está íntimamente vinculado a un lenguaje obsoleto, seguramente un sistema de algunos pocos años haya pasado por tanto programadores que puede convertirse en un legacy mucho más pronto de lo que uno puede imaginarse (Demeyer, Ducasse, & Nierstrasz, 2009) Según lo planteado por Foote en su trabajo *Big ball of mud*, se pueden clasificar algunas razones por las cuales los sistemas Legacy se convierten en una enorme bola de barro. Dichas razones no son triviales y enumeran



los pecados capitales en los cuales incurren generalmente las empresas desarrolladoras de software.

Históricamente los desarrolladores han madurado junto a cientos de metodologías en el desarrollo de software, las modas fueron cambiando y los protagonistas que forman parte del grupo de trabajo también. Esto lleva a largas discusiones sobre el cumplimiento real de los requerimientos, para asegurar que el producto final es exactamente lo que el cliente desea.

La generación de productos bajo una demanda exigida fomenta la creación de código rápido que luego debería ser desechado, en general si funciona no se toca y rara vez es documentado. El código rápido convive con el resto de la aplicación arrastrándose a lo largo del tiempo y sufriendo adaptaciones de todo tipo. La acumulación de código rápido perjudica la claridad y disminuye la posibilidad de escalar la aplicación y su mantenimiento.

Los sistemas miden su complejidad en función de diversos factores: como la cantidad de áreas que abarcan, su tamaño y la cantidad de personas que participan. Los sistemas de información complejos promueven un alto impacto por la cantidad de recursos necesarios que son utilizados, exigiendo una coordinación adecuada entre todas las tareas. (Laudon & Laudon, 2000). Las empresas para mejorar su nivel competitivo utilizan los recursos en otros países con mejores ofertas económicas (Porter, 1991). Estos países tienen un alto grado de rotación en el personal, y sumado a la falta de documentación se pierde un contexto de interrelaciones que constituyen la base de conocimiento de la solución.

Comúnmente el uso de la programación orientada a objetos permite el aporte de uno de sus pilares fundamentales, la encapsulación. El uso de encapsulación permite ocultar la implantación desarrollada en el interior a los ojos observadores del exterior. El uso de

fachadas colabora en el ordenamiento de una solución basada en capas eliminando las dependencias complejas (Gamma, Helm, & Johnson, c2003). A pesar de las ventajas que promueve el uso de patrones, es utilizado también para ocultar el código desprolijo.

El ciclo de vida de los sistemas de información impone cambios naturales en la evolución del software. (Lehman, 2000) Propone un conjunto de leyes que fueron identificadas en el estudio de sistemas entre 1968 y 1985. Los mecanismos que aseguran el crecimiento y madures del software son con el tiempo justificantes en el aumento de complejidad de los legacy.

En base a que los sistemas tienden a evolucionar según las leyes de Lehman, el software, en su mayoría, no respetan las buenas prácticas aplicando técnicas que favorecen los riesgos de aumentar el mal olor en el código fuente (Mendez, Overbey, & Garrido, 2010). Lo anteriormente mencionado identifica varias de las prácticas que, a lo largo del tiempo, promueven las razones por las cuales el código heredado puede ser difícil de mantener. La falta de buenas prácticas en el ciclo de vida de un sistema de información convierte el código fuente en una malla compleja de interrelaciones que lo vuelven difícil de mantener y que junto a la falta de testeos de unidad catapultan los modelos de refactoring como soluciones indispensables.

Con el afán de asegurar aplicaciones legibles, varios autores (Larman, c2003) (Gamma, Helm, & Johnson, c2003) (Kerievsky, 2009) (Fowler, Rice, & Foemmel, 2010) han trabajado sobre la confección de un catalogo de patrones que predominan en la comunidad con el fin de mejorar el código de los sistemas. Sin embargo, los criterios que mueven el uso de patrones no siempre son adecuados, el uso indiscriminado de patrones no siempre es considerado como buenas prácticas (Poveda Villalon, Suárez-Figueroa, & Suárez-Figueroa, 2009) y sumado al hecho que las

aplicaciones con cierto grado de madurez no lo aplican desde sus orígenes transforma las aplicaciones Fortran en un verdadero problema de mantenibilidad. Los cambios cognitivos y las diferentes habilidades que pueden tener los programadores se vuelve un multiplicador del problema. A pesar que estos programadores basen su andamiaje de conocimiento en los tres conceptos básicos finito, preciso y definido (Joyanes Aguilar, 1998) no asegura un código legible sin la auditoría del trabajo que realizan. Es aquí donde aparecen los refactoring, los mismos permiten asegurar transformaciones en el código sin modificar los resultados obtenidos en dicho código (Overbey, Negara, & Johnson, 2009).

Luego de varios años de evolución, Fortran se ha convertido en un lenguaje de programación aplicable en disciplinas como meteorología, física y matemáticas. Las aplicaciones desarrolladas en este lenguaje han evolucionado variando sus requerimientos y ajustándolos a las necesidades actuales. Fortran a evolucionado en los últimos años asegurando la convivencia de todas las versiones del lenguaje, cada fragmento de código es un módulo complejo que depende del momento que fue programado y la versión del lenguaje que se utilizó. La magnitud de las tareas de mantenimiento se incrementan por la evolución de las versiones del lenguaje y por los ajustes de los requerimientos funcionales convirtiendo las actividades de mantenimiento en un verdadero reto (Mendez, Overbey, & Garrido, 2010). Por esto último, La legibilidad pasó a ser considerada un atributo importante de las piezas de programación.

La preocupación por transformar la mantenibilidad, legibilidad o el riesgo de un sistema de información en valores cuantificables no es una tarea sencilla debido a la compleja interoperabilidad entre diferentes elementos poco similares que conviven con el código fuente como

estilos de programación, cambios en la conceptualización de las buenas prácticas de moda, evolución en las versiones de la semántica del lenguaje o la discontinuidad u obsolescencia del conjunto de instrucciones. Estas relaciones complejas se reflejan en los estilos de programación creando zonas de conflicto al momento de escalar o mantener los servicios de la aplicación. El uso de métricas es un camino factible para alcanzar el objetivo de cuantificar la secuencia de aplicación de Refactoring.

La secuencialidad puede expresarse con la combinación de dos factores, la legibilidad y la mantenibilidad. Raymond expone que la legibilidad está sujeta a diversas características descriptivas, tan relativas como el propósito del software, la lógica de programación y la documentación que acompaña al código. Uno de las consideraciones del autor es analizar la legibilidad en pequeños fragmentos de código, algo que en nuestro modo de entender se aplicaría perfectamente a los ejemplos de Refactorización. (Raymond & Westley, 2008). El estudio de campo de estudiantes de ingeniería centrado en el comportamiento compulsivo al programar y las interpretaciones intelectuales de la semántica facilitan el hallazgo de estadísticas de legibilidad.

La IEEE define mantenibilidad como La facilidad con la que un sistema o componente de software puede ser modificado para corregir fallos, mejorar su funcionamiento u otros atributos o adaptarse a cambios en el entorno (Institute of electrical and Electronics Engineers, 1990). Considerar mantenible un fragmento del código fuente asegura la factibilidad de cambiar el código después de concluir su implementación. El mantenimiento del código esta sujeto según Sommerville en 3 tipos claramente diferenciados 1) por fallas, 2) Por Adaptaciones, 3) Por nuevas funcionalidades (Sommerville, 2002).

A diferencia de la subceptibilidad relacionada a la legibilidad, la mantenibilidad puede calcularse en función del índice de mantenibilidad expresado como una conjunción de varias métricas (Oman & Hagemester, 1992) (Kaur & Singh, 2011) (Aldekoa, Trujillo, Sagardui, & Diaz, 2006) (Samoladas, Stamelos, Lefteris, & Apostolos, 2004).

$$MI = 171 - 5,2 \ln(\text{avgV}) - 0,23 \text{ avgV}(g) - 16,2$$

En donde:

**avgV:** es el Volumen promedio por modulo de Halstead.

**avgV(g):** es el promedio extendido por modulo del cálculo de complejidad ciclomática

**avgLOC:** es el promedio de la cantidad de líneas de código por modulo

**PerCM:** es el promedio del porcentaje por modulo de líneas para comentarios.

Originalmente el cálculo de MI fue definido sin considerar las líneas de comentarios, pues en función de su participación en la ecuación el valor es insignificante. Con lo cual el índice de mantenibilidad podría calcularse como:

$$MI = 171 - 5,2 \ln(\text{avgV}) - 0,23 \text{ avgV}(g) - 16,2 \ln(\text{avgLOC})$$

Esta métrica analizada por los ingenieros de HP, pretende ser un índice porcentual que indica la complejidad de mantener el código de un modo general, con el fin de poder localizar de manera factible los posibles problemas. El valor obtenido por las formular de esfuerzo de Halstead brinda los valores posibles para determinar el índice de mantenibilidad, sin embargo el Cálculo del índice se complementa con los resultados de una combinación de métricas. (Coleman, Ash, Lowther, & Oman, 1994).

## LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Este proyecto tiene como objetivo principal encontrar una vinculación entre la

legibilidad y la mantenibilidad para determinar secuencialidad en la aplicación de Refactoring.

Nuestro estudio elabora un sentido para la construcción de un algoritmo que permita esta estimación basado en la generación de criterios que asocien la legibilidad y la mantenibilidad de una pieza de código apoyado en estadísticas de legibilidad y métricas precisas que expresen numéricamente el índice de mantenibilidad.

Buscamos focalizar los requerimientos para reducir la incertidumbre que promueve la necesidad de participación humana en la toma de decisiones de modo tal que la secuencia de aplicación de Refactoring en ciclos iterativos asegure una continuidad en los valores de legibilidad y mantenibilidad desde el inicio del proceso hasta su culminación.

## RESULTADOS Y OBJETIVOS

Durante este proyecto se trabajó en 1) selección de un catálogo de refactoring para Fortran, 2) Formalizar un estudio comparativo de las estadísticas de Legibilidad y las métricas de mantenibilidad para establecer una secuencia de aplicación.

Este análisis dio lugar a la confección de una lista de refactoring ordenada en función de los objetivos planteados. El estudio comparativo permite disminuir la brecha de incertidumbre en la implementación de un criterio para aplicar los Refactoring en una solución legacy.

Actualmente, sólo con el índice de mantenibilidad es posible ordenar los refactoring, la experiencia de evaluar conceptos y compartimientos referidos a la legibilidad puede mejorar el orden de aplicación de los Refactoring.

## FORMACIÓN DE RECURSOS HUMANOS

Los trabajos elaborados en la presente

investigación forman parte del desarrollo de tesis para optar al grado de Magister en ingeniería de sistemas de software (UAI).

## REFERENCIAS

- Aldekoa, G., Trujillo, S., Sagardui, G., & Diaz, O. (2006). Experience Measuring Maintainability in Software Product lines. Brcelona: XV Jornadas de Ingeniería del software y Base de Datos.
- Coleman, D., Ash, D., Lowther, B., & Oman, P. (1994). Using Metrics to Evaluate Software System Maintainability. 27 (8).
- Demeyer, S., Ducasse, S., & Nierstrasz, O. (2009). *Object-Oriented Reengineering Patterns*.
- Fowler, M., Rice, D., & Foemmel, M. (2010). *Patterns of enterprise application architecture* (Vol. 1). Boston, MA: Addison Wesley.
- Foote, B., & Yoder, J. W. (2000). Big ball of mud. In *Pattern Languages of Program Design* (Vols. volume 4, pages 654–692.). Addison Wesley.
- Gamma, E., Helm, R., & Johnson, R. (c2003). *Patrones de diseño : elementos de software orientado a objetos reutilizable* (Vol. 1). Madrid: Pearson Educación.
- Institute of electrical and Electronics Engineers. (1990). *IEEE Standar Computer dictionary: A compilation of IEEE Standard Computer Glossaries*. New York, USA: IEEE.
- Joyanes Aguilar, L. (1998). *Programación en turbo /borland : Pascal 7* (Vol. 3a). Madrid: Osborne/McGraw-Hill.
- Kaur, K., & Singh, H. (2011). Determination of Maintainability Index for Object Oriented systems. 36 (2).
- Kerievsky, J. (2009). *Refactoring to patterns* (Vol. 1). Boston, MA : Addison Wesley.
- Laudon, K. C., & Laudon, J. P. (2000). *Sistemas de información gerencial: administración de la empresa digital* (Vol. 8). Mexico, DF: Pearson Educación.
- Larman, C. (c2003). *UML y patrones : una introducción al análisis y diseño orientado a objetos y al proceso unificado* (Vol. 2). Madrid: Pearson Educación.
- Lehman, M. (2000). *Rules and tools for software Evolution Planning and Management FEAST 2000 International Workshop: Feedback in software and Business Processes*. London: Imperial College.
- Mendez, M., Overbey, J., & Garrido, A. (2010). *A Catalog and Classification of fortran Refactoring*. La Plata: Universidad de la Pplata.
- Overbey, J., Negara, S., & Johnson, E. (2009). Refactoring and the evolution of Fortran. (IEEE, Ed.) *roceedings of the 2009 ICSE Workshop on Software Engineering for Computational Science and Engineering, ser. SECSE '09* , 28-34.
- Oman, P., & Hagemeister, J. (1992). *Metrics for Assesing a Software system's Maintainability*. Moscow: University of Odaho.
- Poveda Villalon, M., Suárez-Figueroa, M., & Suárez-Figueroa, G. (2009). *Malas prácticas en ontologías*. Sevilla, España.: CAEPIA-TTIA.
- Porter, J. (1991). *La ventaja competitiva de las naciones*. Buenos aires: Javier Vergara.
- Samoladas, I., Stamelos, I., Lefteris, A., & Apostolos, O. (2004). Open Source software Development Should Strive for Even Greater Code Maintainability. 47 (10).
- Sommerville, I. (2002). *Ingeniería de Softaware*. México: Pearson.
- Raymond, P. B., & Westley, R. W. (2008). *A metric for software readability* (Vol. 10.1145/1390630.1390647). Seattle: International symposium on software testing and analysis.



## Análisis de disciplinas de las Ciencias de la Conducta como aporte a la Ingeniería de Requerimientos

Silvia M. Aranguren<sup>1</sup>, Marcelo G. Estayno<sup>2</sup>, Gladys N. Dapozo<sup>3</sup>

1 Departamento de Sistemas. Facultad de Ciencia y Tecnología. Universidad Autónoma de Entre Ríos, Ruta 11 Km. 10,5.  
3100 Paraná, Entre Ríos, Argentina  
sarangur@gmail.com

2 Departamento de Informática. Facultad de Ingeniería. Universidad Nacional de Lomas de Zamora, Ruta 4 Km 2, 1832  
Lomas de Zamora, Buenos Aires, Argentina  
mestayno@fibertel.com.ar

3 Departamento de Informática. Facultad de Ciencias Exactas y Naturales y Agrimensura  
Universidad Nacional del Nordeste, Av. Libertad 5450, 3400, Corrientes, Argentina  
gndapozo@exa.unne.edu.ar

### RESUMEN

La Ingeniería de Requerimientos (IR) juega un papel relevante en el desarrollo de software dado que contribuye a entender mejor el problema en cuestión. En este marco, un punto crítico de la IR es la comunicación entre las personas, por tanto, es importante incorporar mecanismos que posibiliten una comunicación eficaz entre los clientes y los profesionales de Sistemas. La IR constituye una compleja mezcla de aspectos tangibles e intangibles, dado que aborda cuestiones técnicas, como la trazabilidad o la medición, así como cuestiones referidas al comportamiento individual o grupal. En esta línea de trabajo se propone considerar a la IR como una disciplina sociotécnica, e incorporar nuevos recursos provenientes de otros campos, tales como la Programación Neurolingüística, el Coaching Ontológico, la Inteligencia Emocional y la Mediación, de manera de fortalecer el proceso de comunicación y contribuir a la obtención de requerimientos más precisos y efectivos..

### Palabras clave

Ingeniería de Requerimientos (IR), Programación Neurolingüística (PNL), Coaching Ontológico (CO), Inteligencia Emocional (IE), Mediación

### Contexto

El presente proyecto se desarrolla en el marco de una propuesta de tesis para la Maestría en Ingeniería de Sistemas de Información de UTN – Santa Fe.

### INTRODUCCIÓN

En una época de cambios cada vez más acelerados, los aportes de diferentes ciencias y teorías han colocado a la Ingeniería de Requerimientos (IR) en una posición inmejorable para aplicar estos nuevos conocimientos.

La IR juega un papel relevante pues ayuda, a los responsables del proyecto a entender mejor el problema en cuya solución trabajarán y a utilizar herramientas que les permitan disminuir los riesgos y los sobrecostos en el proceso de desarrollo.

En este contexto, la comunicación entre las personas es compleja, cuando se trata de requerimientos para sistemas o productos software. Para paliar estos inconvenientes hay que hallar mecanismos o herramientas que permitan que la comunicación establecida entre el cliente y el profesional de sistemas sea efectiva y convierta lo emitido por el cliente y/o usuario en información fiable.

### Conceptos sobre IR

La IR es un conjunto de actividades en las cuales, utilizando técnicas y herramientas, se analiza un problema y se concluye con la especificación de una solución.

Algunas de las definiciones más generales de la misma son:

"Ingeniería de Requerimientos es la disciplina para desarrollar una especificación completa, consistente y no ambigua, la cual servirá como base para acuerdos comunes entre todas las partes involucradas y en dónde se describen las funciones que realizará el sistema" [1].

"Ingeniería de requerimientos es un enfoque sistémico para recolectar, organizar y documentar los requerimientos del sistema; es también el proceso que establece y mantiene acuerdos sobre los cambios de requerimientos, entre los clientes y el equipo del proyecto" [2].

"La Ingeniería de Requerimientos es la ciencia y disciplina a la cual le concierne el establecer y documentar los requerimientos" [3].

Como se puede apreciar en cada una de estas definiciones, todos los procesos involucrados con la Ingeniería de Requerimientos están relacionados con identificar, modelar, comunicar y documentar los requerimientos de un sistema o producto de software y los contextos en los cuales este sistema o producto está envuelto.

Los requerimientos deben describir lo que se debe hacer y cómo se debe llevar a cabo. Esto en la vida real es algo muy difícil de realizar. Por esto existen muchas técnicas disponibles para la aplicación de la Ingeniería de Requerimientos, con el fin de asegurar que los requerimientos obtenidos cuenten, al final del proceso de Ingeniería de

Requerimientos, con las características necesarias para ser implementados.

Por tanto, lo que se busca al aplicar un proceso de Ingeniería de Requerimientos es ayudar a la totalidad de los participantes del proyecto a conocer que desean construir antes de empezar la construcción del mismo.

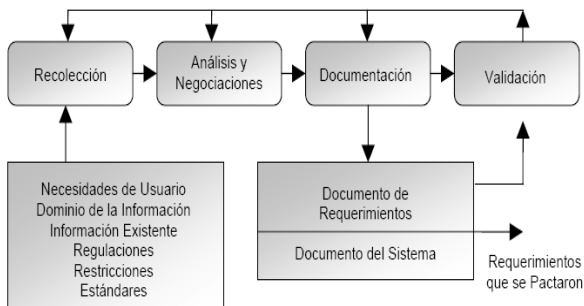
En este sentido, el proceso de Ingeniería de Requerimientos [4] describe de manera detallada y precisa cada uno de los aspectos del ciclo de vida de un conjunto de requerimientos.

La operatoria de los usuarios, los procesos de negocio y dispositivos son elementos típicamente complejos. Por extensión, los requerimientos de software son una combinación compleja de requerimientos de diferentes personas, con diferentes niveles dentro de una organización y restricciones del entorno en que el software funcionará.

La Ingeniería de Requerimientos [5], [6], [7], [8] facilita el mecanismo apropiado para comprender lo que quiere el cliente, analizando necesidades, confirmando su viabilidad, negociando una solicitud razonable, especificando la solución sin ambigüedad, validando la especificación y gestionando los requerimientos para que se transformen en un sistema operacional.

### Nuevos recursos que aportan a la IR

Se debe considerar a la IR como una disciplina sociotécnica<sup>1</sup> que requiere hoy de diversas habilidades y conocimientos [9], [10].



**Fig.1.** El proceso iterativo de los requerimientos

En la figura 1 se esquematiza el proceso iterativo de los requerimientos, es un proceso que involucra actividades complejas, totalmente acopladas (secuenciales y concurrentes), no se trata de una actividad discreta, y no sólo se desarrolla al comienzo de un proyecto de desarrollo de software sino a través de todo su desarrollo.

Y esto se da porque los requerimientos cambian durante todo el proceso; los participantes adquieren mejor comprensión del sistema, los requerimientos originales estaban incompletos, hay cambios en el

ámbito y la tecnología, hay cambios en el entorno de negocios y en entorno técnico del sistema, entre otros.

### LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

El objetivo principal de esta línea de investigación es fortalecer la Ingeniería de Requerimientos utilizando técnicas asociadas a la PNL y el CO, con dinámica de grupos y el manejo de conflictos, para obtener mejores especificaciones de requerimientos reduciendo en gran parte la problemática que surge al definir los requisitos del sistema y logrando un buen clima de trabajo y de comunicación entre los involucrados en esta etapa del proceso de desarrollo de software.

Para ello, los ejes temáticos a abordar son:

- Análisis de las técnicas comunicacionales que forman parte de las Ciencias de la Conducta (PNL, CO, Mediación, Inteligencia Emocional, etc.).
- Análisis de la situación actual de la Ingeniería de Requerimientos, sus métodos y técnicas utilizados, su evolución y efectividad desde su incorporación en el proceso software.
- Estudio de la factibilidad de la incorporación de la PNL, el CO y las técnicas de Mediación, con el objetivo de disminuir la distancia intelectual entre lo que se desea o está en los pensamientos de los usuarios y clientes – a nivel consciente, inconsciente e incluso subconsciente, para establecer mecanismos de comunicación efectivos.

### RESULTADOS OBTENIDOS/ ESPERADOS

En esta primera etapa del proyecto, se analizaron las distintas herramientas que promueven y mejoran la comunicación entre las personas. A continuación se describen las características de estas:

#### Coaching Ontológico:

El modelo del Coaching (CO) ofrece un buen marco de referencia que puede usarse para descubrir requerimientos. Los orígenes del CO se dan en los últimos años con el aporte de nuevos enfoques científicos como la Física Cuántica (David Bohm y Fritjof Capra), la Biología del Conocimiento (Humberto Maturana y Francisco Varela), el Pensamiento Sistémico (Heinz von Foerster) y ciertas corrientes filosóficas (a partir de Heidegger), la Lingüística (a partir de Austin), ciertas corrientes psicológicas (Constructivismo, Gestalt), el Management Moderno (Peter Senge, Stephen Covey) que han contribuido al surgimiento de una nueva interpretación del Ser Humano.

Podemos decir que el CO es fundamentalmente cambio, creatividad y aprendizaje, lo que provoca una gran cantidad de definiciones válidas sobre el mismo. Se puede definir al CO como una disciplina que aporta una manera diferente de interpretar a los

<sup>1</sup> Enfoque social, fundamentado en técnicas y herramientas de las Ciencias de la Conducta (Mediación, CO, PNL, IE) y un enfoque técnico, soportado en la Ingeniería de Requerimientos.

seres humanos, su modo de relacionarse, de actuar y de alcanzar los objetivos que se proponen para sí mismos, para sus empresas y para la sociedad. Uno de los postulados que lo caracterizan es que el lenguaje no sólo describe la realidad, sino que por medio de él se genera la realidad. Un proceso orientado a la transformación personal y profesional, cuyo objetivo es aumentar el poder de acción de las personas, a través del desarrollo de la capacidad de aprendizaje, habilidades de inteligencia emocional y de destrezas de creatividad, liderazgo y comunicación efectiva [11].

Pedro Flores [12] explica la definición básica de la coherencia del ser, “se basa en la coherencia específica y vinculante de las palabras, las emociones y el cuerpo, que influyen recíprocamente y están interrelacionadas, por lo que, si una cambia ésta afectará a que cambien las demás”. En la figura 2 se muestra la interacción de los dominios primarios del observador: su lenguaje, la emocionalidad y el cuerpo.



Fig. 2. Coherencia del ser

### La Mediación

La Mediación, provee un gran potencial en la resolución de conflictos o búsqueda de consensos. En todo conflicto entre dos o más partes, básicamente se entremezclan un sinnúmero de factores que no sólo son los que aparecen a simple vista, están los hechos (como cada parte los interpreta), pero además están presentes emociones, prejuicios, percepciones, suposiciones, opiniones, frustraciones, creencias y hasta complejos y fallas de la personalidad. La mediación es el proceso voluntario por el cual un tercero neutral, MEDIADOR, facilita la comunicación entre dos o más personas para que logren llevar adelante una negociación colaborativa con el objeto de zanjar el conflicto que los enfrenta (con el menor costo de tiempo y desgaste emocional) y llegar a un acuerdo satisfactorio para todos los *stakeholders*.

### La Programación Neurolingüística

Así como el Coaching y la Mediación aportan habilidades para la ingeniería de requerimientos, la PNL ofrece herramientas que ayudan a los ingenieros de requerimientos a mejorar la comunicación con los clientes o usuarios. Según autores de la PNL

(O'Connor y Seymour) “Si continuamos haciendo lo que siempre hemos hecho, continuaremos obteniendo lo que siempre hemos obtenido. Para conseguir algo diferente hagamos algo diferente” [13]. Lo que significa que es importante utilizar nuevas estrategias en la Ingeniería de Requerimientos.

La PNL [14], [15] provee herramientas y habilidades para el desarrollo de estados de excelencia en comunicación y cambio, la cual promueve flexibilidad del comportamiento, pensamiento estratégico y la comprensión de los procesos mentales. PNL significa:

-Programación término tomado de la Informática para sugerir que los pensamientos, emociones y acciones son programas, y que, como tales, pueden ser cambiados. El modo de codificar lenguaje y conducta para crear un “programa” personal.

-Neuro se refiere al sistema nervioso que regula el funcionamiento del cuerpo humano, los circuitos mentales de los cinco sentidos, a través de los cuales las personas pueden ver, oír y sentir. El modo de filtrar la información y procesar lo experimentado a través de los sentidos.

-Lingüística hace referencia a la habilidad para utilizar el lenguaje verbal y no verbal, así como el modo en que las palabras y frases específicas reflejan el mapa mental de las personas. El modo de interpretar las experiencias a través del lenguaje.

La aplicación de la técnica del Metamodelo del Lenguaje [16], basado en los trabajos de Korzybsky y de la lingüística, ayuda al comunicador a reconocer y explicitar la manera en que las distintas personas utilizan su lenguaje verbal.

### Escucha activa

La “escucha activa” es la plena disposición del receptor (responsable del proyecto), a estar presente en tiempo y espacio (aquí y ahora), para recibir, decodificar y procesar los mensajes expresados verbal, postural y gestualmente por el cliente; para luego codificar en palabras lo procesado y darlo como respuesta o *feedback* al emisor.

“No se nos ha enseñado la ventaja que tiene escuchar” [17].

“Todo lo dicho es dicho por un observador” (Heinz Von Foerster) Maturana también dirá “Todo lo dicho es dicho por un observador a otro observador, o a sí mismo” [18].

En toda comunicación existe información verbal y no verbal. La información se transmite no sólo con las palabras, sino a través de los movimientos corporales, las posturas, la mirada, la tensión del cuerpo, las posiciones, las distancias, la forma de sentarse, de andar. Se cree que en toda comunicación, se da un 10 o 15% de información VERBAL (palabras escritas y orales) y un 85 a 90% de información NO VERBAL (compuesta por lo visual, lo gestual, postura y tono de voz) [19].



## La Inteligencia Emocional

La investigación realizada a nivel mundial por *The Consortium for Research on Emotional Intelligence in Organizations*, arrojó un resultado sorprendente y vinculado a nuestro Cociente de Éxito (ver Fig. 3), el mismo se debe un 23% a nuestras capacidades intelectuales, y un 77% a nuestras aptitudes emocionales.

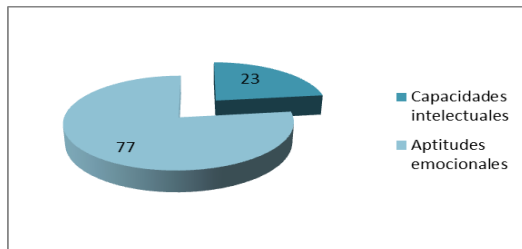


Fig. 3. Cociente de Éxito

Respecto a la emocionalidad y corporalidad, Susana Bloch [20] menciona seis emociones básicas, universales y aculturales: alegría, pena, rabia, miedo, erotismo y ternura y a reconocerlas genuinamente a través de la tríada “respiración + postura + gesto facial”.

El arte de las relaciones se basa, en buena medida, en la habilidad para relacionarnos adecuadamente con las emociones ajenas [21].

Respecto al *rapport*, entendido como la armonía o el entendimiento entre personas que se comunican, tan importante para una comunicación eficaz, Goleman dice “El *rapport* sólo existe entre los seres humanos y se halla presente en cualquier relación afectuosa, comprometida y amable”. La sensación que acompaña al *rapport* es muy positiva y genera la armonía que jalona la simpatía, en donde los distintos implicados experimentan la cordialidad, la comprensión y la autenticidad del otro. Aunque sólo sea de un modo provisional, se trata de una sensación que fortalece los vínculos interpersonales.

“El *rapport* es tan completo que sus posturas y movimientos encajan como si estuvieran ejecutando deliberadamente una danza en la que, cuando uno avanza, el otro retrocede” [22].

## Conclusiones

Para conseguir el éxito en cualquier desarrollo de software es esencial la comprensión total de los requerimientos.

El análisis y la especificación de los requerimientos puede parecer una tarea relativamente sencilla, pero, en realidad, el contenido del análisis es muy denso y abundan las malas interpretaciones o la falta de información. Es muy difícil evitar la ambigüedad.

La elicitación de requerimientos es la fase más importante en el desarrollo de un proyecto software, ya que es en esta fase en la que el usuario indica las especificaciones del futuro sistema, porque de una

correcta elicitación y análisis dependerá la correcta implementación de la aplicación.

La tarea del análisis de requerimientos es un proceso de descubrimiento, refinamiento, modelado y especificación y, por tanto, el desarrollador y el cliente tienen un papel activo en la obtención de estas necesidades.

El responsable de proyecto – mediador necesita recurrir a estas modernas disciplinas del campo de las Ciencias de la Conducta como: la Programación Neurolingüística (PNL), las Ciencias de la Comunicación, el desarrollo de la Inteligencia Emocional y Alfabetización Emocional, el Coaching Ontológico, el desarrollo de la creatividad y la heurística, con dinámica de grupos y el manejo de conflictos.

El aporte de todas estas disciplinas, en síntesis integradora ayudarán a fortalecer la Ingeniería de Requerimientos, reduciendo en gran parte la problemática que surge al definir los requisitos del sistema y logrando un buen *rapport* entre los involucrados en esta etapa del proceso de software permitiendo la coherencia en la especificación de requerimientos y a no dejar cabos sueltos.

Actualmente se estudia la incorporación de estas técnicas en el proceso de IR para fortalecer la definición de los requerimientos de software.

## FORMACIÓN DE RECURSOS HUMANOS

En esta línea de investigación, se encuentran en curso el desarrollo de un plan de trabajo de la Tesina de Grado para obtener el título de Licenciado en Sistemas de la UADER FCyT, y el plan de tesis de la Maestría en Ingeniería en Sistemas de Información de UTN Regional Santa Fe. Asimismo, se espera fortalecer la formación profesional de los matriculados del COPROCIER (Colegio de Profesionales de Ciencias Informáticas de Entre Ríos), mediante talleres de capacitación en temas de Ingeniería de Requerimientos.

## REFERENCIAS

- [1] Boehm B. *Software Engineering Economics*. New Jersey: Prentice Hall, 1981.
- [2] Oberg R., Probasco L., Ericsson M. *Rational Software. Applying Requirements Management with Use Cases* [online]. Rational Software Corporation, 2003, Disponible en Internet: <<http://www.uml.org.cn/RequirementProject/pdf/apprmuc.pdf>>
- [3] Thayer R., Dorfam M. *Software Requirements Engineering*. 2 ed. Los Alamitos, California: IEEE Computer Science Press, 2000.
- [4] Pressman R. S. *Ingeniería del software: un enfoque práctico* MacGraw-Hill, 7ª ed., 2010. ISBN: 978-607-15-0314-5.



- [5] Durán Toro A., Bernárdez Jiménez B. *Metodología para la elicitación de Requerimientos de Sistemas Software*, Versión 2.3, Informe Técnico LSI-2000-10 (revisado), Universidad de Sevilla.
- [6] Robertson S., Robertson J. *Mastering the requirements process*. Londres: Addison – Wesley, 1999.
- [7] Sommerville I. *Ingeniería del software*. Pearson. 7ª ed., 2008. ISBN: 978-84-7829-074-1.
- [8] Sommerville I., Sawyer P. *Requirements engineering: A good practice guide*. 3 ed. Chinchener, Inglaterra: John Wiley & Sons Ltd., 2000.
- [9] Rupp C. *Requirements and Psychology* IEEE Software, 2002.
- [10] Robertson S. *Learning from Other Disciplines*. IEEE Software, 2005, vol. 22.
- [11] Echeverría R. *Ontología del Lenguaje*. Editorial Granica, 2011.
- [12] Flores P. *El Líder Coach: el cambio en 540°* Editorial Forja, 2009.
- [13] Andreas S. y Faulkner C. *PNL: La Nueva Tecnología del Éxito*, Urano, 1998. ISBN: 84-7953-221-1.
- [14] Bandler R. *Use su cabeza para variar – submodalidades en Programación Neurolingüística*. Cuatro vientos editorial. 1º Edición, 1997.
- [15] Cudicio C. *PNL y Comunicación*, Granica 2ª.ed., 2010. ISBN: 978-950-641-569-3.
- [16] Bandler R., Grinder J. *La Estructura de la Magia I*. Cuatro vientos editorial. 4º edición, 1989.
- [17] Echeverría R. *Actos del Lenguaje*. Volumen I: La escucha. Editorial Granica, 2008.
- [18] Echeverría R. *El Observador y su mundo*. Volumen I. Editorial Granica, 2009.
- [19] Mehrabian A. *Comunicación no verbal*. Edit. Transaction Pub, 2009.
- [20] Bloch S. *Al alba de las emociones* Editorial Grijalbo S.A., 2002, reeditado en 2007 por Uqbar Ediciones.
- [21] Goleman D. *Inteligencia Emocional* Editorial Kairós. 82ª edición, 2010.
- [22] Goleman D. *Inteligencia Social La nueva Ciencia de las Relaciones Humanas*, Editorial Cairo, 2006.

## Persistencia de mediciones como apoyo a la gestión de proyectos de software

Greiner, C. (1); Dapozo, G.(1); Acosta, J. (1); Domínguez, M. (1) Chiapello, J. (1); Estayno, M.(2);

(1) Departamento de Informática. Facultad de Ciencias Exactas y Naturales y Agrimensura  
Universidad Nacional del Nordeste

{gndapozo, cgreiner}@exa.unne.edu.ar

(2)Departamento de Informática. Facultad de Ingeniería. Universidad Nacional de Lomas de Zamora  
mestayno@gmail.com

### RESUMEN

La medida de la calidad del software es una necesidad para las empresas de Software y Servicios Informáticos (SSI), porque representa una ventaja estratégica al proporcionar el conocimiento de los procesos productivos y permitir mejorar las tareas menos eficientes. La calidad del software está estrechamente vinculada con la medición del mismo. La implementación exitosa en una organización de una nueva práctica o el uso de una herramienta depende en gran medida de la automatización de la misma, a fin de no significar una pérdida de tiempo o uso de recursos adicionales. En este trabajo se presenta una línea de investigación que apunta a la elaboración de metodologías de medición de atributos, recolección de datos, y diseño e implementación de un repositorio de mediciones, que contribuya a una más eficiente gestión de proyectos de software y a la toma de decisiones en aspectos referidos a la calidad de software.

**Palabras clave:** Calidad de software. Herramientas de medición de software. Sistematización de la medición. Gestión de proyectos de software.

### Contexto

La línea de I/D presentada forma parte de las actividades definidas en el marco del proyecto F007-2009: "Modelos y métricas para la evaluación de la calidad de software", acreditado por la Secretaría General de Ciencia y Técnica de la Universidad Nacional del Nordeste (UNNE).

En el proyecto se trabajan diversas líneas, entre ellas una orientada a la calidad del producto software. El objetivo fundamental es brindar soluciones a las empresas de desarrollo de software de la región NEA en términos de modelos y herramientas que colaboren en la gestión de proyectos de software, particularmente en el proceso de mejora de atributos de calidad del producto final, mediante modelos y herramientas de automatización de medición de software.

### Introducción

El software se ha convertido en parte de la vida cotidiana. Hay una constante interacción con él, y los usuarios necesitan garantías de que el software utilizado cumple con determinados estándares de calidad, los cuales le garantizan un correcto

desempeño y el cumplimiento de sus requerimientos. A lo largo de los años se ha buscado permanentemente la forma de garantizar la corrección del software que se construye. Esto es tanto más importante cuanto más crítica sea la aplicación de que se trate, es decir, cuyas fallas pueden ocasionar daños de importancia (pérdida de vidas humanas, grandes pérdidas financieras, catástrofes nucleares, etc.). Por ello la implementación de procesos de mejora, gestión de atributos de calidad, mediciones de software y certificaciones de distintos estándares juegan cada vez un papel más importante en el ámbito del desarrollo de software.

La medida de la calidad del software es una necesidad para las empresas de Software y Servicios Informáticos (SSI), porque representa una ventaja estratégica al proporcionar el conocimiento de los procesos productivos y permitir mejorar las tareas menos eficientes.

La calidad del software está estrechamente vinculada con la medición del mismo. Autores como Piattini et.al [1] aseguran que la medición de atributos internos del software es el primer indicador de cumplimiento de atributos externos, como la mantenibilidad, funcionalidad etc.

En [2] se señala que atributos como la mantenibilidad y comprensibilidad son evaluados utilizando métricas de software, las cuales proveen un modo de representar en números atributos abstractos como la complejidad y el tamaño. Los mismos autores mencionan que la utilización de una sola métrica es insuficiente para analizar efectivamente atributos de calidad, por lo que sugieren utilizar un conjunto de métricas para evaluar cada atributo externo de calidad.

En [3] se destaca que el esfuerzo del testing debe estar enfocado en las clases cuyas métricas están por fuera de los valores recomendados. Identificar estas clases constituye un desafío y una ardua tarea, para lo cual se han probado muchas métricas, técnicas y modelos.

Por lo tanto en esta línea de investigación se han enfocado los esfuerzos al estudio de modelos y métricas de atributos de calidad del producto, y a la automatización de esta tarea, con el propósito de apoyar la gestión de proyectos de software para el desarrollo de productos de calidad.

### **Métricas de software orientado a objetos**

En el proceso de desarrollo de software, la medición de los atributos que definen la calidad del mismo permite a los desarrolladores analizar la eficacia de los modelos de análisis, diseño y código, así también como de la efectividad de los casos de prueba y la calidad general del software mientras se diseña o construye [4].

Las métricas para aplicaciones Orientadas a Objetos (OO) deben ajustarse a las características que distinguen al software de este paradigma del software convencional, por lo que hacen hincapié en conceptos tales como encapsulamiento, herencia y polimorfismo. Sin embargo, comparten los objetivos de las métricas convencionales: comprender mejor la calidad del producto, estimar la efectividad del proceso y mejorar la calidad del trabajo realizado a nivel del proyecto [5].

Dependiendo de la tecnología utilizada para el desarrollo del software, se encuentran disponibles diversas herramientas para la medición de atributos en aplicaciones OO, con diferentes características y nivel de desarrollo. Las mismas ayudan a comprender mejor el código fuente de los proyectos, entender su arquitectura, auditar la calidad del código, detectar las posibilidades de refactorización y reducción de la complejidad

### **Métricas de Software OO en JAVA**

Existe una gran variedad de herramientas para la medición de atributos en programas escritos en Java, que trabajan de forma independiente, o como plugin de otros entornos de desarrollo, tales como Eclipse metrics plugin, Chidamber and Kemerer Java Metrics, Coqua, JRefactory, EclipsePlug-in to calculate CKmetrics, JDepend, etc.

Entre ellas, destaca RefactorIT, que está disponible como un plugin de Eclipse, NetBeans, JDeveloper, o JBuilder. Su valor deriva de la baja curva de aprendizaje y no tiene costo adicional. Presenta en forma gráfica las dependencias entre paquetes, y permite exportar las mediciones en archivos de diferentes formatos, como CSV, HTML y XML.

### **Métricas de Software OO en .NET**

Del relevamiento y análisis de herramientas que permiten la medición de aspectos inherentes al diseño e implementación de software del entorno .NET, se seleccionaron SourceMonitor y NDepend. Estas implementan un gran número de métricas, incluidas las de Chidamber y Kemerer, uno de los conjuntos de métricas de software OO a los que se hace más ampliamente referencia [6]. Ambas permiten comparar las mediciones de diferentes versiones, proporcionando información sobre la evolución entre etapas de refactorización del código. También permiten exportar el resultado de las mediciones a archivos de distintos formatos, como HTML y XML.

### **Métricas de Software OO en PHP**

PHP es uno de los lenguajes más utilizados en el desarrollo de aplicaciones web. Forma parte de las tecnologías LAMP (Linux Apache, MySQL, PHP/Python/Perl), siendo esta plataforma una de las más extendidas en cuanto software open source para desarrollo web se refiere [7], y en sus últimas actualizaciones se introdujo un modelo de objetos, lo cual dio lugar a la necesidad de mediciones de atributos OO en aplicaciones PHP. Un relevamiento de las herramientas de medición para aplicaciones de estas características permitió establecer que PHPDepend es la que mayor cantidad de información brinda, sin embargo, no aporta una cantidad significativa de información en comparación a herramientas para otros lenguajes y entornos de desarrollo.

### **Métricas de software como apoyo a la gestión de proyectos**

Los ambientes cambiantes en el área de desarrollo de software y la competencia globalizada han cambiado el modo en que se lleva a cabo actualmente la administración de proyectos de software. Las nuevas tecnologías, la mejora de la calidad, generación y distribución de conocimiento, coordinación de varios proyectos, etc. evidencian la necesidad de nuevas herramientas e iniciativas de apoyo a la administración de proyectos, las cuales deben adoptarse considerando las características y objetivos propios de la organización. Las tecnologías de información juegan un papel clave en esta evolución.

La mejora en las prácticas de la administración de proyectos de desarrollo de software, dadas las características de la competitividad actual, solo puede lograrse por medio de la integración de múltiples herramientas, el uso de nuevas técnicas, etc, sin olvidar el fomento al aprendizaje organizacional [8].

Para la implementación exitosa en una organización de una nueva práctica o el uso de una herramienta se debe tener en cuenta la obligatoriedad del carácter sistemático y automatizado que esta debe presentar. Incorporar la medición en un proyecto de software, no debe significar una pérdida de tiempo o uso de recursos adicionales, sino que debe ser vista como una actividad que aporte información automática, rápida, sencilla y estandarizada de las distintas etapas por las que pasa un proyecto software para contribuir a la toma de decisiones.

Esta línea de investigación se enfoca en metodologías de medición de atributos, recolección de datos, estado de proyectos etc., y al diseño e implementación de un repositorio de mediciones, que contribuya la toma de decisiones en la gestión de proyectos de software.

## Líneas de investigación y desarrollo

Las principales líneas del proyecto están orientadas hacia el análisis, estudio y discusión de modelos de evaluación de calidad, estándares, metodologías y herramientas, especialmente enfocadas a la aplicabilidad en las pymes de software.

En particular, se trabaja actualmente en las siguientes líneas:

- Medición de atributos de calidad OO en aplicaciones desarrolladas en los lenguajes de programación Java, PHP y .NET
- Herramientas software que automatizan el cálculo de las métricas OO y generan información para su análisis.
- Persistencia de los valores de las distintas mediciones de los atributos de calidad definidos, integrándolos en una base de datos.
- Conformación de una base de conocimiento para contribuir a la toma de decisiones en la gestión de proyectos de software.

## Resultados y Objetivos

En esta etapa del desarrollo del proyecto algunos resultados obtenidos fueron:

- Se elaboró una metodología para la medición de atributos de calidad y determinación de nivel de complejidad en aplicaciones orientadas a objeto desarrolladas en Java, utilizando la herramienta RefactorIT [9].
- Se realizó un relevamiento y análisis de herramientas de medición de atributos de calidad de aplicaciones OO desarrolladas en .NET [10].
- Se elaboró una metodología de medición de atributos característicos de la programación orientada a objetos sobre aplicaciones escritas en PHP 5, con el objetivo de conocer qué tan bien se implementan los conceptos de la POO en aplicaciones PHP. La misma fue aplicada para evaluar cuatro frameworks de desarrollo PHP, que soportan POO, utilizando la herramienta PHPDepend cuyos resultados se describen en [11].

Los trabajos que se encuentran en curso son:

- Diseño y desarrollo de una herramienta de medición de atributos de calidad orientados a objeto para evaluación de código PHP ya que, como se mencionó en [11], las herramientas relevadas proveen información generalizada sobre un conjunto de atributos, que no permite identificar fehacientemente el elemento que presenta alguna falencia. No especifican la clase o método exacto que contiene el código que mayormente influye en el resultado negativo de la medición, representado con un

valor de medición fuera del rango de aceptación. Por tal motivo se propone el desarrollo de una herramienta de medición con el objetivo de proporcionar a los desarrolladores y directores de proyectos, información sobre el estado de calidad del software de sus proyectos y un informe detallado señalando las porciones de código donde se producen resultados de mediciones fuera de rango.

- Se encuentra en etapa avanzada de desarrollo una aplicación orientada a la persistencia de las mediciones de complejidad de software orientadas a la generación de un repositorio de mediciones provenientes de diversos proyectos, cuyas características se describen más adelante.
- Finalizada la herramienta, se espera contar con la colaboración de empresas de desarrollo del medio, para la recolección de datos reales, y su posterior análisis y emisión de informes, que contribuyan a la administración de proyectos de software.

## Persistencia de las mediciones

Con este objetivo se encuentra en desarrollo una herramienta que permite registrar mediciones de atributos de calidad de software en un repositorio de datos cuantitativos, de modo tal de generar un registro histórico como apoyo a la gestión de proyectos de desarrollo de software, que posee las siguientes funcionalidades:

- **Importar medición:** Permite la selección del archivo que contiene la medición que provee la herramienta de cálculo de las métricas. Analiza la sintaxis de la información adaptando al modelo subyacente, y registra en la base de datos.
- **Informes:** Permite la generación de reportes basados en las mediciones registradas.
- **Gestión:** Permite la administración de las métricas disponibles, como ser su edición o consulta.

Las mediciones de los atributos de calidad son generadas por herramientas existentes en el mercado para el cálculo automatizado. Actualmente la aplicación permite la manipulación de información generada por *RefactorIT*, para software desarrollado en java. Merced a la arquitectura e implementación de patrones detallados más adelante es posible extender la aplicación permitiendo la importación de mediciones generadas por otras herramientas de medición como por ejemplo las que ofrece PHP Depend, entre otras.

Los informes proveen información detallada, señalando clases con posibles dificultades y sugerencias para su corrección, e información



resumida, indicando en qué porcentaje las clases se encuentran dentro de los valores recomendados.

Esta información permite un análisis puntual del código, permitiendo detectar situaciones que pueden afectar la calidad del software. De este modo, se podrá tomar mejores decisiones respecto a la implementación de mejoras correctivas.

Dada la necesidad de hacer sistemas más modulares, que permitan la reutilización de componentes, reduciendo de ese modo el costo de desarrollo, en el diseño de la presente aplicación se prestó especial cuidado de la arquitectura de la misma. Una arquitectura bien implementada permite mantener un bajo acoplamiento entre los elementos del sistema, con muy alta cohesión, dado que se define una clara estructura de los componentes, sus formas de comunicarse y las relaciones que existen entre ellos.

Para el desarrollo de esta aplicación se optó por el Modelo Vista Controlador (MVC), patrón de arquitectura de software que separa los datos, la interfaz de usuario y la lógica de control de una aplicación en tres componentes distintos. Esta arquitectura permite desarrollar software cuyo mantenimiento y evolución son fáciles de realizar, contribuyendo con estas características a la calidad del software.

También se tuvo en cuenta el uso de patrones, utilizándose los siguientes:

- **DAO:** (Objeto de Acceso a Datos) permite abstraer a la aplicación de la tecnología de persistencia subyacente a través de una interfaz común, logrando un bajo acoplamiento e independencia con respecto al modo y fuente de acceso a datos (base de datos, archivos, servicios externos, etc.).
- **Strategy** (patrón de diseño): Permite definir un grupo de clases que representan un conjunto de posibles comportamientos, los cuales pueden ser fácilmente intercambiables en una aplicación, modificando la funcionalidad en cualquier instante. Este patrón se utiliza en la funcionalidad de importación de las mediciones, brindando flexible respecto a la incorporación de mediciones provenientes de otras herramientas, no previstas inicialmente.

Para la implementación de la aplicación se eligió el lenguaje de programación orientado a objetos Java debido que ofrece portabilidad y robustez, y la persistencia de los datos se realiza en una base de datos relacional, con el sistema de gestión de bases de datos MySQL, efectuando el mapeo de los datos con el ORM Hibernate.

Se utilizan un conjunto de librerías para optimizar el desarrollo, entre ellas se cuentan: Hibernate, Spring, Jasper Reports, IReport, Jdom y Swing.

## Formación de Recursos Humanos

En esta línea desarrollan su plan de actividad tres becarios de pregrado de la SECYT-UNNE, una estudiante de la Licenciatura en Sistemas de Información de la UNNE quien desarrolla su Trabajo Final de Aplicación y una docente de la misma carrera avanza con su tesis de Maestría Ingeniería de Software.

## Referencias

- [1] Piattini, M.; Garzas, J.; García, F.; Genero, M. "Medición y estimación del software". ISBN: 9789701514139 - Editorial ALFA OMEGA. 2008.
- [2] Herbold, S.; Grabowski, J.; Waack, W. "Calculation and optimization of thresholds for sets of software metrics". Empir Software Eng. 2011.
- [3] Kpodjedo, S.; Ricca, F.; Galinier, P.; Guéhéneuc, Y.; Antoniol, G. "Design evolution metrics for defect prediction in object oriented systems". Empir Software Eng. 2011.
- [4] Pressman, R. "Ingeniería de Software. Un enfoque práctico". McGraw-Hill – 2005.
- [5] González, D. "Las Métricas de Software y su uso en la Región". Tesis Licenciatura. Ingeniería en Sistemas Computacionales. Universidad de las Américas Puebla. 2001.
- [6] Chidamber, S., Kemerer, C., "A Metrics Suite for Object-Oriented Design." M.I.T. Sloan School of Management E53-315, 1993.
- [7] Fernández, A. Comparativa proyectos PHP para e-commerce. Revista Todo Linux. N°. 112, 2010, págs. 26-30.
- [8] Caballero Cervantes, O.H. "Tecnologías de información y herramientas para la administración de proyectos de software". Revista Digital Universitaria. 2006 • Volumen 7 Número 6 • ISSN: 1067-6079.
- [9] Demchum, D., Greiner, C., Dapozo G.; Cuenca Pletsch L., Estayno M. "Medición de Atributos de Calidad en Aplicaciones Orientadas a Objeto". Publicado en la revista "La UTN en el NEA: Investigación y Desarrollo en la Regional Resistencia". 2011.
- [10] Chichi López, Elías D.; Greiner, Cristina L.; Dapozo, Gladys N. Métricas Técnicas de Software. "Análisis comparativo de herramientas de medición". Reunión de Comunicaciones Científicas y Tecnológicas 2012. Universidad Nacional Del Nordeste (UNNE). Modalidad: Póster y Exposición oral. Con referato. 27, 28 y 29 de Junio de 2012. Campus Universitario. Corrientes. Argentina.
- [11] Acosta, J.; Greiner, C.; Dapozo, G.; Estayno, M. "Medición de atributos POO en frameworks de desarrollo PHP". CACIC-2012. Publicado en anales. ISBN 978-987-1648-34-4.
- [12] Stelting, S.; Maassen, O. "Patrones de diseño aplicados a java". ISBN: 9788420538396. Prentice Hall. 2003.
- [13] Spring Source Feature Tour – Sitio oficial del proyecto. <http://www.springsource.org/features>

# Herramientas Software Para Una Optima Gestión del Proceso de desarrollo de software, a utilizarse en las Empresas de Software de la Provincia de Catamarca

María Carolina Haustein, Diana Palliotto, María Isabel Korzenieswki

Departamento Informática/Facultad de Tecnología/Universidad Nacional de Catamarca

Maximio Victoria N° 55- San Fernando del Valle de Catamarca

0383 4435112,int 165

[Carolina.haustein@gmail.com](mailto:Carolina.haustein@gmail.com); [diana.palliotto@gmail.com](mailto:diana.palliotto@gmail.com); [marisa\\_kb2004@yahoo.com.ar](mailto:marisa_kb2004@yahoo.com.ar)

## Resumen

Una de las metas siempre vigente en las empresas desarrolladoras de software, es mejorar el proceso de desarrollo, para lo cual, se fueron fijando y ejecutando distintas estrategias, la más recurrente fue la creación de nuevas metodologías. Hasta que a principios del siglo XXI, se advirtió que lo que fallaba en la gran parte de las empresas, era la gestión en el desarrollo de software, tarea esta, sumamente importante, ya que se encarga de la planificación, asignación de recursos, la documentación y fundamentalmente el control en el proceso desarrollo. Existen varias herramientas que ayudan a un jefe de equipo de proyecto de desarrollo de software, a realizar una la gestión del mismo. El presente trabajo de investigación busca llegar a las empresas de desarrollo de software de la provincia de Catamarca, para transmitir los resultados del estudio de dichas herramientas, indicándoles cuáles son las óptimas de acuerdo al software a desarrollar; con el objetivo de brindarles los conocimientos necesarios, al momento de realizar una gestión estratégica que, las ayude a poder seguir vigentes en el mercado, y hacer frente a un contexto globalizado en el que se encuentran insertas.

**Palabras claves:** Gestión de desarrollo, estrategias, gestión de control, herramientas de gestión

## Contexto

El presente trabajo se desprende del Proyecto “Estudio y comparación de las herramientas para una gestión estratégica del software.” En el área de Ingeniería de Software. Este proyecto se desarrolla dentro de la Facultad de Tecnología y Ciencias Aplicadas de la Universidad Nacional de Catamarca, y está acreditado por la Secretaría de Ciencia y Tecnología, de la Universidad Nacional de Catamarca.

## Introducción

El avance de las TIC (Tecnología de la Información y la Comunicación) es muy vertiginoso, dado que nos encontramos un contexto de trabajo donde lo único constante es el cambio, y obviamente la gestión de desarrollo de software no exenta de dicha situación. Sumado a esto, la situación de la provincia de Catamarca, en la que ubicación geográfica hace que el acceso a determinadas tecnologías sea más lento. Es que se busca estudiar y poder transmitir a las empresas del medio, las herramientas que le sean útiles para llevar adelante la gestión de desarrollo de software.

En un comienzo las herramientas aportadas por las TIC para el desarrollo de software, fueron las técnicas tradicionales, donde se permitía organizar el proyecto en función de las

tareas y subtareas, la asignación de recursos financieros, humanos y temporales.

Toda empresa desarrolladora de software debe definir sus objetivos y las estrategias para alcanzar los mismos. Al momento de definir sus estrategias, es sumamente importante, que las autoridades de las empresas transmitan las mismas hacia todos los integrantes del equipo, o de la empresa. El uso de herramientas para ayudar a la gestión de un proceso de desarrollo, son muy importante, ya que, ayudan a las autoridades a llevar un control de los procesos. Se define a una aplicación de administración de proyecto, aquella que ayuda a crear y realizar el seguimiento de los planes de proyectos, a la asignación de recursos, a realizar la planificación temporal, de manera que se pueda cumplir con el calendario y el presupuesto.

Se busca cada vez optimizar las herramientas de administración de proyectos de desarrollo de software, del avance de este estudio y con la llegada a las empresas desarrolladoras de software de la capital de la Provincia de Catamarca y se detectó que el principal problema es la falta de gestión de los mismos, situación está que no encuentra su solución sólo con el desarrollo de herramientas de ayuda, sino con la gestión del personal que tiene a su cargo la función de controlar que el software sea entregado al cliente en el tiempo establecido, y con el presupuesto asignado.

Se entiende por gestión de proyectos de desarrollo software al conjunto de actividades: planificación, la implantación y el control todo el proceso de desarrollo del producto. Realiza el seguimiento de la forma de trabajo de manera que se cumplan los estándares establecidos, se siga la agenda prevista y no se sobrepase el presupuesto ni el calendario. Por esta razón, la gestión de proyectos de software es una rama de la ingeniería de software que emplea metodologías bien definidas, realiza medidas repetibles y confiables, estima costos y tiempos, da elementos para la gestión de los proyectos, replantea resultados para ajustar la información disponible.

Según Pressman [1] la gestión de Proyectos de Software consta de principios básicos tales como una eficaz gestión de la cuatro P's: Personal, Producto, Proceso y Proyecto.

**Personal:** es contar con personal capacitado, preparado y motivado, que posea una estructura organizativa y un buen sistema de comunicación que permita mantener informado al equipo de los compromisos adquiridos y resultados obtenidos, a lo largo del proyecto.

**Producto:** Son el conjunto de artefactos y resultados que se crean durante la vida del proyecto, como los modelos, el código, los ejecutables, la documentación, versiones de productos, entre otros. Antes de poder planificar un proyecto se deben establecer los objetivos y el ámbito del producto, se debe considerar soluciones alternativas e identificar las dificultades técnicas y de gestión.

Proceso: Está constituido por las actividades necesarias para lograr el objetivo, se pueden clasificar en actividades de gestión y actividades de ingeniería. Las actividades de gestión son la planificación, seguimiento y control del proyecto. Las actividades de ingeniería son obtención de requerimientos, análisis, diseño, implementación y pruebas. Proyecto: Es el elemento organizativo a través del que se gestiona el desarrollo del software. Las actividades más relevantes en la Gestión de proyectos de software son [2]:

Planificación: Permite organizar el proyecto en función de tareas y subtareas, con asignación y control de tiempos y recursos materiales y humanos. Idealmente el sistema de planificación debe permitirnos también hacer el seguimiento y reajustar la planificación en función de la evolución del proyecto.

Organización: Referida a las relaciones entre las unidades de trabajo para el cumplimiento de objetivos y el otorgamiento de responsabilidad y autoridad para obtener esos objetivos.

Staffing: Selección y entrenamiento de personas para el proyecto de software.

Dirección: Creación de una atmósfera que apoye y motive a la gente para alcanzar los resultados finales deseados.

Control: Establecimiento, medición y evaluación del desempeño de las actividades a través de los objetivos planeados. Es una actividad continua. Esta actividad le permite al gestor conocer el progreso del proyecto con los costos actuales y los planificados. También, es normal tener varias revisiones formales de su gestión. Se hace una revisión completa del progreso y de los desarrollos técnicos del proyecto, teniendo en cuenta el estado del proyecto. El resultado puede dar lugar a una cancelación.

Las TIC son el conjunto de herramientas que han sido desarrolladas para gestionar información y, que la misma pueda ser utilizada en distintos lugares, abarcan un gran número y muy variados de soluciones, incluyendo las tecnologías para almacenar información y recuperarla después, enviar y recibir información de un sitio a otro, o procesar información para poder calcular resultados y elaborar reportes. De esta manera podemos afirmar que los usos y aplicaciones de las TIC en la gestión de proyecto de software impactan de manera positiva permitiendo responder a un mercado altamente competitivo y cambiante [3].

Actualmente en el mercado existen diversas soluciones que pueden ser utilizadas para la gestión de proyectos. Las primeras herramientas que aportaron las TIC en la gestión de desarrollo de software, emplean métodos y técnicas tradicionales, la nueva generación de herramientas combinan las tres S's: scope, scheduling y status. Es decir, herramientas para administrar el alcance, la programación de tareas y el estado en que se encuentran [2]. El contexto en el se ha desarrollado los proyectos de software está fuertemente ligado a cinco décadas, a continuación se muestra su evolución

Décadas	Gestión de proyectos de Desarrollo de Software	Tecnologías de la Información y las Comunicaciones
1950-1960	El software se desarrollaba de manera artesanal. Se utilizaban lenguajes de bajo nivel. El procesamiento se	Se caracteriza por una mecanización de transacciones; surge una relación indisoluble entre el progreso tecnológico y económico.

	realizaba por lotes. La mayoría del software desarrollado era utilizado por la misma persona u organización. El desarrollo de software carecía de metodologías y se realizaba sin ninguna planificación. La documentación no existía y era muy dependiente del hardware.	
1960-1970	Aparecen los sistemas multiusuario, el procesamiento en tiempo real, la primera generación de sistemas de base de datos. Se acuñó el término "crisis del software" en la primera conferencia de la Organización del Tratado del Atlántico Norte (OTAN) sobre desarrollo de software y con él se etiquetaron los problemas que surgían en el desarrollo de sistemas de software. También se utilizó por primera vez el término "ingeniería del software" para describir el conjunto de conocimientos que existían en aquel estado inicial.	En 1969 aparece un invento que cambiaría muchos aspectos de la vida: Internet. El desarrollo de Internet ha significado que la información está ahora en muchos sitios.
1970-1980	Se caracteriza por la difusión de computadoras lo que lleva a que se pueda trabajar con proyectos medianos y grandes que permiten hacer gran cantidad de estudios que derivan en modelos y metodologías con el objetivo de mejorar el desarrollo de software y la administración del	La revolución electrónica iniciada en la década de los 70 constituye el punto de partida para el desarrollo creciente de la Era Digital. Los avances científicos en el campo de la electrónica tuvieron dos consecuencias inmediatas: la vertiginosa suba de los precios de las materias primas y la preponderancia de las Tecnologías de la



	mismo.	Información (Information Technologies) que combinaban esencialmente la electrónica y el software.
1980-1990	Se confrontan los modelos y las metodologías con grandes proyectos de software, de esta confrontación sobreviven la técnica de puntos de función de Albrecht, el modelo COCOMO de Bohem, entre otros. Aparecen nuevos paradigmas de programación y producción de programas como la orientación a objetos, sistemas gestores de bases de datos y sistemas operativos.	Las investigaciones desarrolladas a principios de los años 80 han permitido la convergencia de la electrónica, la informática y las telecomunicaciones posibilitando la interconexión entre redes. De esta forma, las TIC se han convertido en un sector estratégico para la "Nueva Economía".
1990-Actualidad	Es la época de la normalización de las metodologías de modo que se pueden obtener resultados consistentes y comparables. El software libre se está convirtiendo en una tendencia importante.	Las TIC conforman el conjunto de recursos necesarios para manipular la información: los ordenadores, los programas informáticos y las redes necesarias para convertirla, almacenarla, administrarla, transmitirla y encontrarla.

Tabla 1. Evolución del desarrollo de software y las TIC.  
Fuente: Elaboración Equipo de Proyecto

Las herramientas de software para la gestión de proyectos pueden permitir [4]:

- Trabajar en colaboración en los proyectos (Collaborative software).
- Realizar un seguimiento de las tareas (Issue tracking system).
- Realizar la programación del proyecto (Scheduling).
- Gestionar la cartera de proyectos (Project portfolio management).
- Gestionar los recursos asignados a cada proyecto (Resourcemanagement).
- Gestionar los documentos de cada proyecto (Documentmanagement).


El software de gestión de proyectos puede ser:

- Software Libre o
- Software Propietario.

## Líneas de Investigación


En este trabajo nos abocamos al estudio de herramientas de software libre. Este ha cobrado muchos adeptos a nivel mundial, por las ventajas, tanto filosóficas como prácticas, que ofrece a sus usuarios y desarrolladores. Las ventajas de este movimiento se derivan de las cuatro libertades (uso, estudio, modificación y distribución) que promueven en sus sistemas, donde se encuentra la posibilidad de adaptar rápidamente el software a las preferencias y necesidades, tanto de usuarios como de organizaciones. Las herramientas estudiadas y comparadas fueron las siguientes:

**OPENWORKBENCH**



- Requisitos: se ejecuta bajo Microsoft Windows 2000, XP, Vista y 7 y Java Standard Edition
- Idioma: Inglés
- Versión: 1.1.6
- Web de referencia: <http://www.open-workbench.org>

**GANTTPROJECT**




- Requisitos: es multi-plataforma (Windows, Linux y MacOSX)
- Idioma: Multilingüe
- Versión: 2.5.3
- Web de referencia: <http://www.ganttproject.biz/>

**PLANNER**


- Requisitos: Inicialmente Desarrollada Para Linux, dispone de versión para Windows
- Idioma: Multilingüe
- Versión: 0.14.4
- Web de Referencia: <https://live.gnome.org/Planner/>

**TASKJUGGLER**




- Requisitos: Se ejecuta en todos los Linux, Unix, Windows, MacOS y otros sistemas operativos
- Idioma: Inglés
- Versión: 3.2.0
- Web de Referencia: <http://www.taskjuggler.org/>

**CALLIGRA PLAN (antes KPLATO)**



- Requisitos: Se encuentra disponible para su descarga para la mayoría de las distribuciones Linux. Una versión aún en desarrollo para Windows
- Idioma: Inglés
- Versión: 2.4
- Web de Referencia: <http://www.calligra.org/plan/>

**GANTTPV**



- Requisitos: se encuentra disponible para Linux, Unix, Mac y Windows
- Idioma: Inglés
- Versión: 0.11
- Web de Referencia: <http://www.pureviolet.net/ganttppv/>



## Resultados y Objetivos

Habiendo hecho un análisis y comparación de las mismas desde el punto de vista de Software de colaboración, Sistema de Seguimiento, Programación, Gestión de proyectos, Gestión de Recursos, Gestión de Documentos, de las misma se ha transmitido a las empresas de desarrolladoras de software de la provincia de Catamarca, en donde en este momento tres empresas, han decidido utilizar las mismas de acuerdo a nuestro análisis y nos están ayudando en su uso, para ver si las mismas, teniendo en cuenta ya, que la principal función de un gestor es la gestión, sirven o necesitan alguna otra característica.

Del resultado de este trabajo conjunto con las empresas desarrolladoras de software de la Provincia de Catamarca, se espera poder terminar este trabajo con la información necesaria, para poder medir cada software y poder realizar uno con todas las opciones que las mismas necesitan.

## Formación de Recursos Humanos

El equipo de proyecto está formado por una Mg en Ingeniería de Software, 4 Licenciadas en Sistemas de la Información, que han terminado de cursar la Maestría en Ingeniería de Software y dos alumnos de la carrera Ingeniería en Informática. De los cuatro Lic. Una se encuentra desarrollando ya su Tesis de maestría, Dos Lic. Han presentado el plan de Tesis y están esperando su aprobación. Los alumnos que forman parte del equipo están desarrollando su tesis de grado.

## Referencia

[1] Pressman, Roger. Ingeniería del Software. Un enfoque Práctico (Quinta Edición). Editorial McGraw-Hill. Madrid, (2002)

[2] Cavallero Cervantes, O.: Tecnologías de Información y Herramientas para la Administración de Proyectos de Software. Revista Digital Universitaria. Volumen 7 Número 6, ISSN: 1067-6079, (2006).

[3] Cortés Tarrá, A.: TIC como factor de competitividad, (2011)

[http://www.tarratech.net/web2.0/images/stories/publicaciones/Tic\\_como\\_factor\\_de\\_competitividad.pdf](http://www.tarratech.net/web2.0/images/stories/publicaciones/Tic_como_factor_de_competitividad.pdf)

## Definición Formal de una metodología para la generación de sistemas de software orientados a servicios.

Germán Montejano<sup>1,2</sup>; Oscar Testa<sup>2</sup>; Pablo García<sup>2</sup>; Silvia Bast<sup>2</sup>; Oscar Dieste<sup>3</sup>

<sup>1</sup>Departamento de Informática

Universidad Nacional de San Luis

Ejército de los Andes 950 – (5700) San Luis – San Luis – Argentina

Tel.: +54-2652-424027 – Int. 251

gmonte@unsl.edu.ar – web: <http://www.unsl.edu.ar>

<sup>2</sup>Departamento de Matemática

Universidad Nacional de La Pampa

Av. Uruguay 151 – (6300) Santa Rosa – La Pampa – Argentina

Tel.: +54-2954-425166 – Int. 28

[otesta, pablogarcia, silviabast]@exactas.unlpam.edu.ar

<sup>3</sup>Facultad de Informática

Universidad Politécnica de Madrid

Campus Montegancedo – (28660) Boadilla del Monte – Madrid – España

Tel.: +34 91 336 5011

odieste@fi.upm.es

### RESUMEN

El objetivo de este trabajo es realizar aportes a la Ingeniería de Software orientada a servicios. Se analizaron los distintos enfoques y estado del arte de la orientación a servicios junto a la arquitectura SOA y se comenzó a delinear un modelo de desarrollo de sistemas orientados a servicios. Inicialmente, el modelo se planteó en el uso de tecnologías ya probadas como BPM, BPEL y UML siendo SOA la arquitectura que soporta todo el modelo. Durante el primer trimestre del año 2013 se puso a prueba dicha metodología para desarrollar un prototipo de software para el registro y consulta de datos de experimentos en el campo de la Ingeniería de Software Experimental en la Universidad Politécnica de Madrid.

Una vez realizados los ajustes necesarios a la metodología la misma se especificará formalmente a través de RSL. Posteriormente, como un segundo caso de prueba de la mencionada metodología se validará a través de un caso específico: “Una Arquitectura abierta y estándar de servicios Web de GIS”, también especificada formalmente en RSL, trabajo éste, realizado como Tesis de Maestría en Ingeniería de Software<sup>1</sup>.

**Palabras clave:** RAISE, RSL, SOA, Servicios Web, Framework, Infraestructura, Métodos formales, Programación Orientada a Servicios, Computación Orientada a Servicios, Arquitectura Orientada a Servicios.

### CONTEXTO

El presente trabajo se enmarca en el Proyecto de Investigación: Ingeniería de Software, Conceptos, Métodos y Herramientas en un Contexto de “Ingeniería de Software en Evolución” – Facultad de

Ciencias Físico-Matemáticas y Naturales, Universidad Nacional de San Luis y en el Proyecto de Investigación: Definición y Especificación Formal de un Modelo basado en Servicios para la Generación de Sistemas de Software – Facultad de Ciencias Exactas y Naturales, Universidad Nacional de La Pampa. Las líneas aquí presentadas fueron expresadas en una Tesis de Postgrado en la Maestría en Ingeniería de Software auspiciada por el proyecto de la Universidad Nacional de San Luis. Actualmente se está trabajando sobre el proyecto de investigación mencionado, donde se espera poder definir formalmente un Modelo de desarrollo basado en Servicios para la generación de Sistemas de Software.

### 1. INTRODUCCIÓN

La capacidad para responder rápidamente ante los cambios y optimizar los procesos de negocio son factores claves para la competitividad y el crecimiento de las organizaciones. La agilidad de éstas puede verse cuestionada si se apoya en entornos de IT que no pueden responder de forma flexible a los cambios que afectan la actividad del negocio. Liberar el potencial que poseen las aplicaciones y recursos de IT y hacerlo disponible de forma general a toda la organización, facilita la optimización de procesos y mejora la agilidad empresarial. Las organizaciones, por lo tanto, han vislumbrado que el objetivo final es la automatización de los procesos de negocio (desarrollar aplicaciones que darán soporte a todos y cada uno de los pasos de un proceso de negocio, desde su comienzo hasta su finalización). La Arquitectura Orientada a Servicios (SOA, Service Oriented Architecture) es una filosofía de diseño que permite un mejor alineamiento de las Tecnologías de Información (IT) con las necesidades de negocio[7] [12].

<sup>1</sup>Tesis defendida en la Universidad Nacional de San Luis en Octubre de 2009.

Se percibe cada vez con mayor claridad que los procesos de negocio no son constantes y necesitan ser adaptados. Necesitan ser flexibles, por lo tanto, IT también debe ser flexible. [12]

Las empresas necesitan poder interconectar los procesos, personas e información tanto con la propia Organización como -atravesando sus fronteras- con subsidiarias y socios comerciales. La falta de integración e interoperabilidad entre los componentes de IT –sistemas, aplicaciones y datos- hace difícil obtener una respuesta rápida y efectiva ante los cambios que afectan de forma natural a los negocios. La inflexibilidad genera costos, reduce la capacidad de respuesta ante los clientes, compromete el cumplimiento de normativas legales y afecta negativamente en el personal de la empresa. En síntesis, la interoperabilidad y la poca adaptabilidad de los sistemas son algunos de los problemas más importantes a los que las organizaciones deben hacer frente para mantener su competitividad y garantizar su crecimiento[7].

El software administrativo está estrechamente unido a la organización interna, a los procesos y modelos de negocio. Este tipo de software subyace tanto en las dependencias interdepartamentales como en las relaciones exteriores de la empresa. En consecuencia, una arquitectura de software administrativo debe hacer frente a un gran número de requisitos diferentes. Muchos de estos requisitos son contradictorios, mientras que otros no están claros. En casi todos los casos, los requisitos son un blanco en movimiento por el cambio permanente de mercados, la organización de la empresa, y sus objetivos de negocio. Es por esto que el desarrollo de software administrativo se hace realmente muy complejo.

Las aplicaciones administrativas de las organizaciones rara vez contienen una gran cantidad de algoritmos complicados. El código que describe un trozo de lógica de negocio suele ser muy simple. La estructura de un sistema de facturación basado en COBOL es mucho más simple que, por ejemplo, un sistema embebido de un dispositivo específico (que contiene necesidades de cálculos complejos en tiempo real, y necesidades multi-threading). En las aplicaciones empresariales usualmente se encuentran estructuras de datos muy simples, a diferencia, nuevamente, de otros sistemas tales como sistemas de información geográfica (SIG)[8].

Los sistemas administrativos normalmente tienen que interactuar con distintos stakeholders, incluso algunos de ellos pueden ser CEOs de otras compañías, proveedores o usuarios que pueden verse influenciados por entornos políticos (de la propia empresa o del propio gobierno) que hacen realmente complejo contemplar todas estas variables dentro del sistema a desarrollar. En cambio en otros tipos de

software (software de escritorio, embebido, etc), el entorno y el espacio del problema son mucho más pequeños e incluso más acotado[8].

Se puede concluir que los sistemas o el software administrativo es único en varios aspectos y, por lo tanto, requiere el uso de medidas y metodologías únicas para mejorar su desarrollo y mantenimiento posterior[8].

Como resultado de la estrecha relación con la organización interna, los procesos y modelos de negocio, una arquitectura de software para la empresa debe cumplir con requisitos muy diferentes que los de una arquitectura de software para un sistema que es controlado por un pequeño número de expertos altamente calificados de dominio, como enviar un robot a Marte o un motor de juego de video.

Para poder brindar agilidad y eficiencia, una arquitectura de software administrativo debe contemplar características particulares, a saber: simplicidad, flexibilidad y mantenimiento, reusabilidad y, por último, poder desacoplar la funcionalidad y la tecnología.

A través de la utilización de una Arquitectura Orientada a Servicios (SOA) se puede ayudar a lograr los objetivos de diseño para el desarrollo de software administrativo, planteados anteriormente. Para ello se introducirán algunos conceptos de la orientación a servicios.

El término "servicio" ha estado presente en la computación comercial durante mucho tiempo y ha sido utilizado de muchas maneras diferentes. Hoy, por ejemplo, nos encontramos con grandes empresas, tales como IBM, promocionando el concepto de "servicios a la carta". A principios del nuevo siglo, el término "servicios web" se convirtió en muy popular, aunque a menudo se ha utilizado para referirse a conceptos muy diferentes de computación: algunas personas lo utilizan para referirse a la aplicación de servicios prestados a los usuarios humanos a través de Internet, otras personas lo utilizaron para referirse a los módulos de aplicación accesibles para otras aplicaciones a través de Internet, haciendo uso de los protocolos basados en XML[8].

Se puede definir entonces Arquitectura orientada a servicios (SOA) como "una arquitectura débilmente acoplada diseñada para satisfacer las necesidades de negocio de una organización"[9].

SOA es un enfoque arquitectural para la creación de sistemas a partir de servicios autónomos. Esta arquitectura permite la creación de sistemas de software altamente escalables que reflejan el negocio de la organización, brindando además una forma bien definida de cómo se exponen y se invocan los servicios, lo que facilita la integración e interacción de sistemas tanto propios como de terceros.

La Arquitectura SOA establece un marco de diseño para la integración de aplicaciones independientes de manera que desde la red pueda accederse a sus funcionalidades, las cuales se ofrecen como servicios. La forma más habitual de implementarla es mediante Servicios Web, una tecnología basada en estándares e independiente de la plataforma, con la que SOA puede descomponer aplicaciones monolíticas en un conjunto de servicios e implementar esta funcionalidad en forma modular. En este punto es importante brindar una definición de “servicio”, si bien en la literatura existen varias, este trabajo utiliza la siguiente conceptualización:

“Un servicio es una funcionalidad concreta que puede ser descubierta en la red y que describe tanto lo que puede hacer como el modo de interactuar con ella.”

Desde la perspectiva de la empresa, un servicio realiza una tarea concreta: puede corresponder a un proceso de negocio tan sencillo como introducir o extraer un dato como “Código del Cliente”. Pero también los servicios pueden acoplarse dentro de una aplicación completa que proporcione servicios de alto nivel, con un grado de complejidad muy superior –por ejemplo, “introducir datos de un pedido”-, un proceso que, desde que comienza hasta que termina, puede involucrar varias aplicaciones de negocio.

La estrategia de orientación a servicios permite la creación de servicios y aplicaciones compuestas que pueden existir con independencia de las tecnologías subyacentes. En lugar de exigir que todos los datos y lógica de negocio residan en un mismo ordenador, el modelo de servicios facilita el acceso y consumo de los recursos de IT a través de la red, en un modelo distribuido.

Puesto que los servicios están diseñados para ser independientes, autónomos y para interconectarse adecuadamente, pueden combinarse y recombinarse con suma facilidad en aplicaciones complejas que respondan a las necesidades de cada momento en el seno de una organización. Las aplicaciones compuestas (también llamadas “dinámicas”) son las que permiten a las organizaciones mejorar y automatizar sus procesos manuales, disponer de una visión consistente de sus clientes y socios comerciales, y orquestar sus procesos de negocio para que cumplan con las regulaciones legales y políticas internas.

El resultado final es que las organizaciones que adoptan la orientación a servicios pueden crear y reutilizar servicios y aplicaciones, adaptarlos ante los cambios evolutivos que se producen dentro y

fuera de ellas, y con ello adquirir la agilidad necesaria para ganar ventaja competitiva.

La figura 1 muestra un resumen de cómo se puede descomponer un proceso en servicios.

## 2. LÍNEAS DE INVESTIGACIÓN y DESARROLLO

Tal como se menciona en la Introducción, la capacidad para responder rápidamente ante los cambios y optimizar los procesos de negocio son factores claves para la competitividad y el crecimiento de las organizaciones.

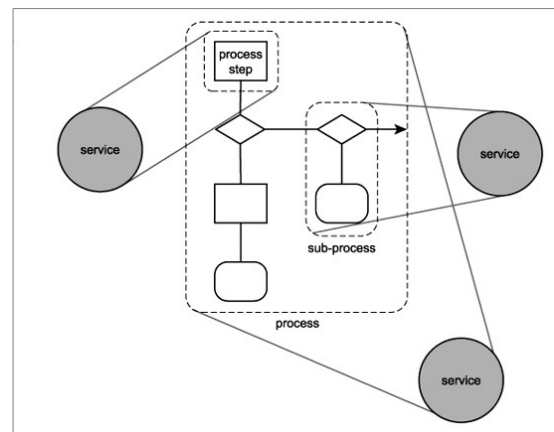


Figura 1. Descomposición de un proceso en servicios.

Con el ánimo de dar respuesta a las necesidades planteadas anteriormente, se genera la arquitectura orientada a servicios, que permite enfocar la solución de los problemas de desarrollo de software para empresas, de una manera más sencilla, rápida y ágil. De esta forma es posible conseguir una adaptación fácil a los cambios organizacionales y estructurales de la empresa. A partir de esta arquitectura, también se han generado nuevos paradigmas de programación, como por ejemplo: Programación orientada a servicios (SOP), Computación Orientada a Servicios (SOC). Además, debe recordarse que la orientación a servicios se basa en metodologías y paradigmas existentes como los Servicios Web, la Programación orientada a objetos o programación orientada a componentes. SOA tiene sus bases también en la Arquitectura basada en modelos (MDA) y UML.

Un enfoque que no debe perderse de vista, que se encuentra explícitamente expresado en la Introducción, es el concepto de BPM, que a su vez ha sido acompañado en estos últimos años por el lenguaje de especificación de procesos BPEL, que es la traducción a servicios de los procesos de negocios. Con todo esto es posible expresar que el modelo de desarrollo de software para generar aplicaciones orientadas a servicios debe estar basado en la arquitectura SOA, utilizar la notación de procesos de negocios BPMN, su traducción BPEL, y



los diagramas y especificaciones de UML. La Figura 2 muestra un detalle de cómo sería una posible arquitectura, que contiene los conceptos mencionados y sirve de base al modelo en el que se está trabajando para el desarrollo de sistemas orientados a servicios.

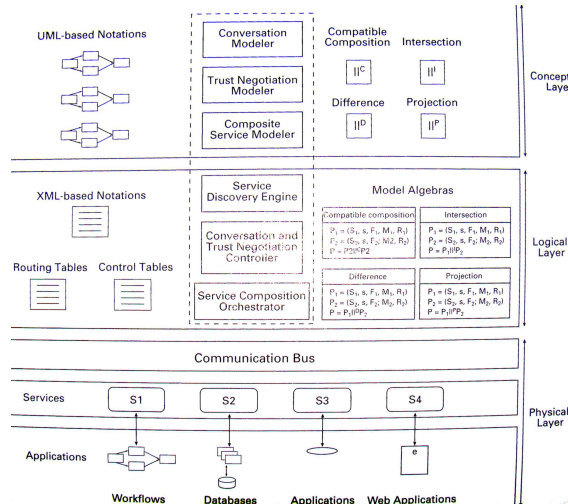


Figura 2. Arquitectura de servicios. [13]

Existe la fuerte convicción de que si las organizaciones tuvieran herramientas que les permitieran desarrollar sistemas en forma rápida y sencilla, siguiendo un ciclo de vida que incluya modelos / técnicas / arquitecturas / paradigmas / metodologías establecidos, podrían tener ventajas competitivas respecto del resto.

Estas herramientas y la forma de combinarlas para lograr el desarrollo rápido y sencillo, surgirán de la especificación formal de un modelo para el desarrollo de sistemas orientados a servicios basados en los conceptos mencionados anteriormente.

La hipótesis más importante que se plantea, por lo tanto, es que si las organizaciones contaran con herramientas que les permitan desarrollar sistemas en forma rápida y sencilla siguiendo modelos establecidos, probados y de utilización directa tendrían ventajas competitivas respecto del resto. Esto conduciría o redundaría en un mejor posicionamiento en el mercado frente a los desafíos que se presentan hoy en día.

Cuando se menciona la construcción de un modelo de desarrollo, no significa estrictamente un modelo nuevo, sino por el contrario, se intentará tomar la mejor estrategia luego de hacer un exhaustivo análisis de las tecnologías existentes. Posteriormente se decidirá si es mejor o no realizar una construcción nueva, mejorar alguna existente, o bien tomar la mejor alternativa ya probada.

Las líneas de investigación se basan principalmente en los siguientes tópicos, a saber:

- Búsqueda minuciosa acerca de las tecnologías / metodologías en relación al ámbito de este proyecto de investigación. Este análisis permitirá realizar un cuadro comparativo que posibilitará evaluar las fortalezas y desventajas de cada una de las tecnologías. De este análisis surgirán las conclusiones acerca de cómo debería ser un modelo teórico para el desarrollo de sistemas enfocados a servicios.
- Definición teórica del modelo para desarrollo de sistemas orientados a servicios a utilizar y luego especificación formal del mismo. La definición del modelo se realizará de manera informal, es decir en un lenguaje ambiguo, pero que sentará las bases para luego poder especificarlo formalmente. En esta definición del modelo se incluirán todos los detalles necesarios para que el pasaje a lenguajes formales sea lo más directo posible. Para la definición del modelo se utilizarán todas las herramientas que puedan facilitar la tarea, como, por ejemplo, diagramas UML.
- Formalización del modelo definido durante el segundo año de trabajo. Para esta tarea se utilizarán lenguajes de especificación formal como son RSL.

El lenguaje de especificación formal a utilizar para la especificación del modelo será RAISE Specification Language (RSL), dado que el mismo es reconocido en la industria del software para especificaciones formales de desarrollos reales.

RSL es un lenguaje formal, basado en el formalismo de la matemática usando conceptos tales como la teoría de conjuntos, lógica de primer orden, lógica de orden superior. Está netamente orientado a construir modelos, ya sea describiendo un dominio de la realidad o describiendo una herramienta a desarrollar y sus requerimientos

Durante el transcurso de los dos primeros años también se harán estudios del estado del arte respecto de lenguajes para especificaciones formales, pudiendo surgir otra alternativa de lenguaje a utilizar como puede ser MAUDE o Alloy (sólo por mencionar algunos). Si bien se asume que se utilizará RSL, se espera evaluar otras alternativas.

### 3. RESULTADOS OBTENIDOS/ESPERADOS

Al momento de escribir este trabajo se está culminando con una prueba de concepto donde se aplica la metodología seleccionada. De esta prueba se espera poder sacar conclusiones que permitan realizar ajustes a la metodología con el fin de

optimizarla. La metodología elegida se basa principalmente en el uso de BPM, como base de todo el proceso, haciendo uso de casos de uso y modelos de UML para determinar y descubrir los distintos servicios, su interrelación y los documentos que que los conectan.

Se espera como próximo paso poder realizar ajustes necesarios para optimizar la metodología y poder de esa forma especificarla formalmente.

#### 4. FORMACIÓN DE RECURSOS HUMANOS

Además de los resultados obtenidos/esperados en el punto 3, se espera como resultado en la formación de recursos humanos, la continuación de esta misma línea de proyecto como tesis doctoral de alguno(s) de los investigadores. También se espera lograr una mayor interrelación con la Universidad de Minas Gerais con la que se cuenta con un convenio con tal objetivo como parte de él. Se espera avanzar también en un convenio de colaboración con la Universidad Politécnica de Madrid para la aplicación de las metodologías aquí presentadas en los proyectos de Ingeniería de Software Empírica. Adicionalmente, se espera que otras tesis de Maestría, así como tesinas de Licenciatura surjan a partir de los logros obtenidos en la presente línea de investigación. Es oportuno mencionar que uno de los docentes investigadores se encuentra actualmente becado por el Ministerio de Educación de la Nación con una beca de Movilidad Docente a Madrid, para realizar tareas de investigación relacionadas con el área de este proyecto. La beca incluye una estadía durante el primer bimestre de 2013 en la Universidad Politécnica de Madrid. España. Se espera con esta estancia establecer vínculos con la mencionada Institución para la realización de futuros intercambios de docencia e investigación.

#### 5. BIBLIOGRAFÍA

1. The RAISE Method Group, "The RAISE Development Method", The Practitioner Series, PrenticeHall, 1995.
2. Clarke, E; Wing, J., "Formal Methods: State of the Art and Future Directions".
3. Ethan, Cerami, "Web Services Essentials – Distributed Applications with XML-RPC, SOAP, UDDI & WSDL", O'Reilly, 1th ed, 2002.
4. Newcomer, Eric, "Understanding Web Services Xml Wsdl Soap And Uddi", Addison Wesley.
5. The RAISE Language Group, "The RAISE Specification Language", The BCS Practitioner Series, PrenticeHall, 1992.
6. Dang, Van Hung, George, Chris, Janowski, Tomasz, Moore, Richard, "Specification Case Studies in RAISE", UNU-IIST, 2002.
7. Microsoft Corporation, "La Arquitectura Orientada a Servicios (SOA) de Microsoft aplicada al mundo real", Whitepaper, Diciembre de 2006.
8. Dirk, Krafzig; Karl, Banke; Dirk, Slama, "Enterprise SOA: Service-Oriented Architecture Best Practices", Prentice Hall PTR, 2004.
9. James P. Lawler; H. Howell-Barber, "Service-Oriented Architecture. SOA Strategy, and Technology", Auerbach Publications, 2008
10. Thomas Erl, "Service-Oriented Architecture: Concepts, Technology, and Design", Prentice Hall PTR, 2005.
11. Testa, Oscar; Montejano, Germán, Riesco, Daniel, "Especificación Formal en RSL de una Infraestructura abierta y estándar de Servicios Web de SIG.", Tesis de Maestría de Ingeniería de Software, Universidad Nacional de San Luis, 2010.
12. Matjaz B. Juric, Kapil Pant, "Business Process Driven SOA using BPMN and BPEL", Packt Publishing, 2008.
13. Dimitrios Georgakopoulos, Michael P. Papazoglou, "Service-Oriented Computing", MIT Press, 2009.

## Criterios de calidad en el desarrollo de aplicaciones web

Ferraro, M. <sup>(1)</sup>; Medina Y, <sup>(1)</sup>; Dapozo, G.<sup>(1)</sup>; Estayno M.<sup>(2)</sup>; Pedrozo Petrazzini, G.<sup>(1)</sup>; Villafañe, A.

<sup>(1)</sup>

(1) Departamento de Informática. Facultad de Ciencias Exactas y Naturales y Agrimensura  
Universidad Nacional del Nordeste

{ferraro, yanina, gndapozo}@exa.unne.edu.ar

gabriel.pedrozopetrazzini@gmail.com, afv0185@hotmail.com

(2) Departamento de Informática. Facultad de Ingeniería. UNLZ  
mestayno@gmail.com

### Resumen

La trazabilidad en la Ingeniería de Software es una práctica de control que ayuda a obtener el producto en el dominio de la solución lo más exacto y fiable posible a las necesidades expresadas por el cliente. En el desarrollo de aplicaciones web, los requerimientos poseen características particulares, las cuales no están contempladas explícitamente en los estándares vigentes. En esta línea de trabajo se analizan los aspectos vinculados a la especificación de requerimientos y la explicitación de mecanismos de trazabilidad en el desarrollo de aplicaciones web, basadas en criterios de calidad definidos por los estándares vigentes. El objetivo principal es generar metodologías y herramientas de software que contribuyan al desarrollo de software para la web, cumpliendo requisitos de calidad.

**Palabras clave:** Ingeniería de Requerimientos. Trazabilidad. Calidad en aplicaciones web. Metodología NDT.

### Contexto

La línea de I/D presentada en este trabajo forma parte del proyecto F007-2009: "Modelos y métricas para la evaluación de la calidad de software", acreditado por la Secretaría General de Ciencia y Técnica de la Universidad Nacional del Nordeste (UNNE), cuya Unidad Ejecutora está integrada por docentes de la Universidad Tecnológica Nacional Regional Resistencia (UTN-FRRe), de la Facultad de Ciencias Exactas de la UNNE y de la Universidad Nacional de Lomas de Zamora (UNLZ).

El objetivo fundamental del proyecto es contribuir a la mejora en la calidad de los productos software mediante modelos y métricas aplicados al producto y al proceso de creación, diseño, desarrollo y mantenimiento de software, como medio para aumentar la competitividad de las pymes de la región NEA en el contexto de la industria del software.

### Introducción

El proyecto en el que se inserta este trabajo se orienta al estudio de la calidad del producto software y del proceso de desarrollo del mismo, teniendo especialmente en cuenta las iniciativas gubernamentales, empresariales y académicas, que impulsan la promoción de la industria del software. Muchas de las organizaciones vinculadas al sector público han implementado diversos esfuerzos de mejora que permitan cumplir con las normativas vigentes, para así mejorar la eficiencia de sus respectivas organizaciones.

La Ingeniería de Requerimientos cumple un papel primordial en el proceso de desarrollo de software, ya que se especializa en la definición del comportamiento del sistema, es decir, de lo que se desea desarrollar o producir. Su objetivo principal es la definición clara, consistente y compacta de las especificaciones correctas que definen el comportamiento del sistema con el fin de minimizar al máximo los problemas que se presentan en el desarrollo de software y que tanto afectan a la calidad del producto final. La captura correcta de los requerimientos contribuye a la mejora de la calidad de software dado que permite definir con precisión las condiciones que éste debe cumplir.

En esta línea de trabajo se analizan los aspectos vinculados a la especificación de requerimientos y la explicitación de mecanismos de trazabilidad en el desarrollo de aplicaciones web, basadas en criterios de calidad definidos por los estándares vigentes.

### Calidad en aplicaciones web

En el desarrollo de aplicaciones web, el requerimiento está inmerso en un proceso de ingeniería más amplio y detallado. La existencia de una importante estructura de navegación obliga a un desarrollo preciso de este aspecto que garantice que el usuario no se "pierda en el espacio navegacional del sistema" [1]. Estas características particulares requieren atención también en la fase de especificación de requerimientos [2].

Desde la perspectiva de desarrollo de aplicaciones web, en la mayoría de los métodos se proporciona poco apoyo a la construcción de un modelo de requisitos para capturar correctamente las necesidades de los usuarios que deben ser soportadas por las aplicaciones web. Sin embargo, existen algunos enfoques que proponen nuevas técnicas o amplían las tradicionales para especificar los requisitos en las aplicaciones web. Ellos son: OOHDM, WSDM, UWE, SOHDM, WebML, OOH, OOWS, W2000 y NDT. En estas metodologías se observa que la mayoría consideran la especificación de requisitos funcionales, y algunas pocas, consideran los datos y los requisitos de navegación. Los requisitos de navegación son considerados principalmente por: OOHDM, WSDM, UWE, WebML, OOH, OOWS, W2000 y NDT (*Navigational Development Techniques*). En particular NDT proporciona mecanismos para describir la información publicada, capacidades de navegación, filtros de información, y las funcionalidades requeridas [3].

#### **Trazabilidad de requerimientos y calidad**

La trazabilidad de requerimientos se define como la habilidad para describir y seguir la vida de un requerimiento en ambos sentidos, hacia sus orígenes o hacia su implementación, a través de todas las especificaciones generadas durante el proceso de desarrollo de software. Para ello el proceso de trazabilidad ha de considerar dos subprocesos: a) configuración de la trazabilidad de acuerdo con las necesidades concretas del proyecto, para así conseguir un resultado positivo respecto del costo-beneficio asociado, y b) especificación de la trazabilidad en el proyecto y la posterior explotación de dicha información [4].

En la actualidad, no existen estándares asociados al proceso de trazabilidad que ayuden a determinar qué tipos de artefactos y de enlaces se han de considerar. Cabe destacar que la trazabilidad se considera como una medida de la calidad del sistema y la madurez del proceso de desarrollo, además es una prescripción de muchas normas o estándares, tales como CMMI (*Capability Maturity Model Integration*), específicamente en el nivel 2, en el Área de Proceso de Gestión de Requerimientos [5].

La carencia de soporte para trazabilidad, es un problema común en la mayoría de las aproximaciones metodológicas que debe de ser solucionado [6]. Según IEEE, un buen Documento de Requerimientos, pese a no ser obligatorio que siga estrictamente la organización y el formato dados en el estándar IEEE 830-1998, sí deberá incluir, de una forma u otra, toda la información presentada en dicho estándar. El mismo no está libre de defectos ni de prejuicios, y por ello ha sido justamente criticado por múltiples

autores y desde múltiples puntos de vista, llegándose a cuestionar incluso si es realmente un estándar en el sentido habitual que tiene el término en otras ingenierías.

Debido a la heterogeneidad de los usuarios de una aplicación web, cualquier método de ingeniería web debe considerar una fase de análisis de requerimientos en el cual se especifiquen las necesidades de los diferentes actores implicados en la misma y que sirva para poder determinar cada una de las características que dicha aplicación debe cumplir para satisfacerlas [1]. Aunque en la actualidad existen varias propuestas para la especificación de requerimientos web [7, 8], la mayoría de ellas sólo proponen un conjunto de guías de diseño informales para la derivación manual de modelos conceptuales a partir de los requerimientos web [6].

Ante esta situación, es necesario conocer el posible impacto derivado del cambio de un requerimiento, es decir, si un requerimiento cambia, por ejemplo, debido al desarrollo gradual de las necesidades del usuario, es necesario saber las partes de los modelos conceptuales de la aplicación web que serán afectadas. Igualmente, si un modelo conceptual es modificado, obligado por el cambio constante en las tecnologías de implementación de las aplicaciones, es indispensable conocer qué requerimientos serán afectados.

#### **Metodología NDT**

La metodología NDT [9], es una propuesta de metodología orientada a la web que surgió con un objetivo muy concreto. A partir de diferentes estudios comparativos [10,11,12], se analizó que si bien la ingeniería web estaba ofreciendo nuevas técnicas y modelos para el desarrollo de software orientado a la Web, la fase de ingeniería de requisitos había quedado poco tratada. NDT comenzó pues a aportar propuestas y técnicas orientadas al tratamiento de requisitos en entornos Web. De esta forma, inicialmente NDT trabaja sólo en dos fases: la ingeniería de requisitos y el análisis.

La propuesta de NDT se define de manera formal sobre un conjunto de metamodelos de requisitos y de metamodelos de análisis. Estos metamodelos, definidos formalmente mediante diagramas de clases, permiten estudiar los artefactos del sistema de una manera abstracta sin entrar en detalles de su representación. Entre los metamodelos se definen una serie de relaciones y restricciones que garantizan la concordancia entre ellos y la calidad de los resultados en un proyecto que use a NDT como metodología de desarrollo. Estas relaciones y restricciones permiten definir en NDT un conjunto de transformaciones, que formalmente se representan mediante QVT [13].



En síntesis, NDT aporta propuestas y técnicas orientadas al tratamiento de requisitos en entornos web y se convierte en una metodología que define formalmente los artefactos necesarios para la fase de requisitos y análisis haciendo un especial hincapié en definir los elementos necesarios para tratar los aspectos propios de sistemas Web, como la navegación o la interfaz de usuario.

### Líneas de investigación y desarrollo

En el marco de la línea de investigación presentada se propone:

- Estudio en profundidad de los conceptos fundamentales de Ingeniería de Requerimientos, Requerimientos web y Trazabilidad.
- Análisis, estudio y discusión de métodos de especificación de requerimientos con presencia de trazabilidad sobre productos software.
- Estudio metodologías de desarrollo web y los estándares ISO 12207 e ISO 15504, vinculados con la calidad de proceso.
- Definición de una propuesta metodológica que cubra los aspectos requeridos en la especificación de software, profundizando en el aspecto de la trazabilidad.
- Diseño y desarrollo de una herramienta de software que implemente la propuesta metodológica realizada.

### Resultados y Objetivos

En esta primera etapa de la investigación se ha obtenido un estado del arte de la Ingeniería de Requerimientos, los aspectos característicos de las aplicaciones web y de las principales técnicas de trazabilidad, en base a estos conocimientos se generó:

- Una propuesta metodológica para la especificación de requerimientos de aplicaciones web basada en, principalmente, un modelo de plantilla de especificación de requerimientos web que considera lo estipulado por el estándar IEEE 830-1998 e incluye, además, las características particulares de los requerimientos web basados en NDT, y elementos trazables y vínculos de trazado para facilitar el rastreo de los requerimientos y el impacto de los cambios. Un objetivo de la misma es que pueda ser utilizada como patrón en diferentes proyectos de desarrollo, contribuyendo de este modo a lograr que el software cumpla con los criterios de calidad definidos en los estándares vigentes. La metodología se describe en [13].
- Para sistematizar la utilización de la plantilla de especificación, se diseñó y desarrolló una herramienta de software que implementa el seguimiento de los proyectos de software, permitiendo el control de los requerimientos.

### Características de la herramienta

Generada como una aplicación web, permite al usuario del sistema, administrar múltiples proyectos de software, sin perder de vista características esenciales entre la solicitud del requerimiento y su correspondiente entrega de producto; permitiendo el seguimiento de los requerimientos en cada una de las fases del desarrollo de software.

El sistema permite consultar y presentar la información referente a la documentación de los distintos requerimientos que componen un proyecto. El mismo será capaz de establecer las relaciones relevantes entre los *workproducts*, de manera que se pueda seguir la vida de un requerimiento en ambas direcciones, hacia delante y hacia atrás, es decir, a través de su origen y especificación, hasta su implementación. Esto permite contar con información sobre vínculos de trazado, para facilitar el rastreo de los requerimientos y el impacto de los cambios que podría originarse sobre los mismos.

A partir de esta herramientas, se avanzará en la validación de la misma aplicando al desarrollo de software en situaciones del mundo real. Para lo cual se llevarán a cabo las siguientes actividades:

- Difusión de la herramienta y sus posibilidades en las empresas de software públicas y/o privadas.
- Aplicación de la herramienta, evaluación, análisis y medición de los resultados obtenidos.
- Implementación de encuestas a fin de obtener información acerca de qué metodologías web se utilizan en los procesos de desarrollo de software en las empresas del medio.
- Identificación de problemas de calidad en los productos software que desarrollan las empresas pymes de la región.
- Propuestas de solución para los problemas de calidad en los productos software de las empresas pymes de la región.

### Formación de recursos humanos

En esta línea de investigación se desarrolla el plan de trabajo de una docente para la obtención del título de Especialista en Ingeniería del Software de la UNLP, se ha desarrollado un Trabajo Final de Aplicación de la carrera Licenciatura en Sistemas de Información de la UNNE y se lleva a cabo el plan de actividades de un becario de pregrado de la SECYT-UNNE, bajo la dirección de docentes integrantes del proyecto.

## Referencias

- [1] ESCALONA, M. and Koch, N., Requirements engineering for Web Applications: a comparative study. *Journal of Web Engineering*, 2004. 2: p. 193-212.
- [2] ESCALONA, M.J. "Metodología para el desarrollo de sistemas de información global: análisis comparativo y propuesta". Department of Language and Computer Science. University of Seville. Seville.2002.
- [3] Valderas, P., and Pelechano, V. 2011. A survey of requirements specification in model-driven development of web applications. *ACM Trans. Web* 5, 2, Article 10 (May 2011), 51 pages.
- [4] DÖMGES R. and Pohl K. Adapting Traceability Environments to Project-Specific Needs. *Communications of ACM*, Vol. 41, No 21, December 1998.
- [5] Object Management Group, Query QVT-Merge Group, Revised submission for MOF2.0 Query/Views/ Transformations RFP, OMG, 2004. <http://www.omg.org/cgi-bin/apps/doc?ad/04-04-01.pdf>.
- [6] AGUILAR, J. A., Garrigos, I., Mazon, J.-N. and Trujillo, J. Web Engineering Approaches For Requirement Analysis - A Systematic Literature Review. in 6th Web Information Systems and Technologies (WEBIST). 2010. Valencia, Spain.
- [7] NICOLÁS, J. and TOVAL, A., On the generation of requirements specifications from software engineering models: A systematic literature review. *Information and Software Technologies*, 2009. 51(9): p. 1291-1307
- [8] MOLINA, F., Pardillo, J. and Toval, A., Modelling web-based systems requirements using WRM. *Web Information Systems Engineering–WISE 2008 Workshops*, 2008. p. 122-131.
- [9] ESCALONA M.J., G. Aragón, "NDT. A model-driven approach for web requirements", *IEEE Transactions on Software Engineering*, vol. 34, n°3, 2008.
- [10] ESCALONA M.J., J. Torres, M. Mejías, J.J. Gutiérrez, D.Villadiego, "The treatment of navigation in Web Engineering", *Advances in Engineering Software*, vol. 38, pp. 267-282, 2007.
- [11] Koch N., "Software Engineering for Adaptive Hypermedia Applications". Ph. Thesis, FAST Reihe Softwaretechnik, vol 12, Uni-Druck Publishing Company, Munich. (Germany), 2001.
- [12] Retschitzegger, W. & Schwinger, W. "Towards Modeling of Data Web Applications - A Requirements Perspective", *American Conference on Information Systems AMCIS 2000*, vol 1, pp. 149-155, USA, 2000.
- [13] Ferraro, M.; Medina, Y.; Dapozo, G.; Estayno, M., "Especificación y trazabilidad de requerimientos en el desarrollo de aplicaciones web". *II Jornadas de Investigación en Ingeniería del NEA y Países Limítrofes*. N° ISBN: 978 – 950 – 42 – 0142 – 7

## Diseño de Algoritmos para Plataformas Paralelas

Marcelo Alaniz, Fabricio Bustos, Verónica Gil-Costa, Virginia Mancini, Cesar Ochoa,  
Marcela Printista

Departamento de Computación  
Facultad de Ciencias Físico Matemáticas y Naturales  
Universidad Nacional de San Luis  
Ejército de los Andes 950, 1° piso. (02652-420823)

### CONTEXTO

La línea de investigación presentada en este trabajo recurre a dos grandes proyectos de investigación de la Universidad Nacional de San Luis, a) el proyecto de sistemas distribuidos y paralelos sobre arquitecturas multi-core; b) el proyecto de recuperación de la información y estructuras de datos.

En el primer proyecto nos enfocamos en el uso adecuado de tecnologías y de la estimación de costos que surgen del desarrollo de nuevos algoritmos paralelos y distribuidos. El diseño de estos algoritmos generalmente se realiza utilizando el modelo de programación paralela BSP [Val90], Multi-BSP[Val11] y librerías de programación asíncronas como MPI o PVM para comunicación inter-nodo y openMP o pthreads para comunicación intra-nodo.

Haciendo uso de las características particulares de las arquitecturas, es posible delinear pautas que formen parte de una metodología de programación que tenga en cuenta: la optimización de algoritmos BSP y su modelo de costo para una aproximación adecuada del tiempo de ejecución total del algoritmo en arquitecturas específicas. El desarrollo de estos temas de investigación también se enmarca dentro del proyecto 13STIC-06 – SEHLOC financiado por STIC-AMSUD.

En el segundo proyecto nos enfocamos en el diseño y desarrollo de nuevas estructuras de indexación y algoritmos de búsqueda para datos. En particular, los algoritmos de búsqueda e indexación son estudiados sobre colecciones de datos u objetos multimediales (sonidos, imágenes, video, etc.).

### RESUMEN

Con el creciente avance de la tecnología y la capacidad de cómputo que proveen las CPU multi-cores, es importante el diseño y desarrollo de las técnicas que exploten las ventajas de estos procesadores multi-cores para acelerar las aplicaciones paralelas que poseen una gran demanda de cómputo paralelo. En particular, para aplicaciones que requieren de un gran poder computacional de los recursos disponibles, es esencial poder desarrollar estrategias y algoritmos que aprovechen el uso adecuado del hardware.

En este trabajo se presentan los objetivos y los desafíos de una línea de investigación que abarca los problemas de uso adecuado de las nuevas arquitecturas de procesadores, y cómo estas nuevas arquitecturas pueden ser utilizadas para mejorar el desarrollo de algoritmos de computación de grafos y cálculo de matrices, utilizando como base formal el modelo de programación paralela BSP; conjuntamente con algoritmos de búsqueda e indexación sobre grandes colecciones de datos como la Web.

**Palabras clave:** *Sistemas de computación Híbridos. Multi-core. BSP. High Performance Computing. Estrategias de búsqueda. Espacios Métricos.*

### 1. INTRODUCCION

Estudios realizados recientemente indican que el incremento promedio en performance de procesadores ha excedido el incremento en la velocidad de reloj [Hennesy07, Hennesy08], lo que muestra que el incremento en el número de transistores en un chip da lugar a mejoras en la arquitectura de un procesador que reducen el tiempo promedio de ejecución de una instrucción. El número de transistores dentro de un chip puede ser usado como un estimador de su complejidad y performance. La ley de

Moore es una observación empírica que establece que el número de transistores de un típico chip se duplica cada 18-24 meses [Rauber10]. Esta observación fue realizada en 1965 por Gordon Moore y se ha mantenido vigente por más de 40 años.

Varias técnicas de diseño de microprocesadores han sido utilizadas para explotar el paralelismo interno de un único procesador. Entre las técnicas más relevantes se pueden mencionar paralelismo a nivel de bit, técnicas de pipelining y multiplicidad de unidades funcionales, entre otras [Culler99]. Estas tres técnicas asumen un único flujo de control secuencial el cual es provisto por el compilador y el cual determina el orden de ejecución en el caso de existir dependencias entre instrucciones. Para un programador, este tipo de técnicas internas tiene la ventaja de permitir ejecución paralela de instrucciones por medio de un lenguaje de programación secuencial. Sin embargo, el grado de paralelismo obtenido por pipelining y múltiples unidades funcionales está limitado.

Un enfoque alternativo para aprovechar el número creciente de transistores en un chip, es ubicar múltiples cores sobre un único chip de procesador [Hager11]. Este enfoque se ha estado utilizando en los procesadores de escritorio desde 2005 y son denominados procesadores multi-core. Estos nuevos procesadores capaces de incrementar el poder de cómputo, hacen resurgir el paradigma de computación paralela olvidado en la década anterior.

Los chips multi-core agregan nuevos niveles en la jerarquía de caché. Cada core posee una cache denominada L1 capaz de almacenar unos pocos Kb. Un segundo nivel de caché denominada L2 es compartida por todos los cores agrupados en el mismo chip y tiene una capacidad de almacenamiento de unos pocos Mb. Un ejemplo es la máquina de Intel Sandy Bridge-E que tiene hasta 8 cores en un único chip, 32 Kb en la cache L1, 256 Kb en la cache L2 (esta cache no es compartida por los cores, sino que es privada a cada uno), y hasta 20Mb de cache L3 que es compartida por todos los cores de un mismo chip.

Las tendencias de la tecnología indican que el número de cores (núcleos) en un chip seguirá creciendo a medida que lo indiquen los planes de trabajo de los fabricantes más importantes. Los procesadores actuales como

el IntelXeon Quad-core o el Intel Core i7 proveen cuatro CPUs físicas y ocho lógicas que pueden ser administradas por el programador mediante librerías como pthreads o OpenMP que posee un alto nivel de abstracción. Existen varias librerías de programación para sistemas multi-core como TBB [Reinders07] que utiliza ciclos para describir el paralelismo de datos y otros como IBM X10[Charles05] y Fortress [Allen07] que se focaliza en el paralelismo de datos pero también provee paralelismo de tareas.

Un sistema multi-core ofrece a todos los cores un acceso rápido a una única memoria compartida, pero la cantidad de memoria disponible es limitada.

El desafío en este tipo de arquitecturas es reducir el tiempo de ejecución de los programas. En particular, se propone estudiar, implementar y comparar la eficiencia y performance de algoritmos de búsqueda de texto (usando índices como el Arreglo de Sufijos[Manber93]) y datos multimediales utilizando diferentes índices métricos, y optimizar estos algoritmos al ser ejecutados en clusters de procesadores que soportan multi-core. Además, caracterizar las distintas arquitecturas de hardware multi-core utilizando benchmark desarrollados para tal propósito, y herramientas como hwlock y contadores de hardware (librería PAPI).

### **1.1 Diseño de Algoritmos Paralelos basados en el modelo BSP y Multi-BSP**

El modelo Bulk-Synchronous Parallel (BSP)[Val90] propone que los algoritmos paralelos sean diseñados y evaluados no sólo por el balance clásico entre el tiempo y el número de procesadores, sino también por la comunicación y la sincronización. Este modelo establece un puente entre los algoritmos paralelos y las arquitecturas de hardware. La sencillez de la programación, la portabilidad y el modelo de costo asociado han estimulado el desarrollo de un gran número de algoritmos BSP paralelos.

Pero, con el paso del tiempo, uno de los supuestos básicos de BSP en relación a la visión de un hardware paralelo estaba cambiando. Hoy en día, la atención se centra en la utilización óptima de las arquitecturas complejas que no sólo son altamente escalables, pero también jerárquicas y no homogéneas.



El problema surge cuando se intenta adaptar estos algoritmos a la variedad de arquitecturas existentes. Y se debe principalmente a que BSP no tiene en cuenta el impacto de esta nueva complejidad. En estos casos, el desarrollador debe tener en cuenta afinidades de hardware cuando se trata de explotar el rendimiento del hardware real. Por ejemplo, dos tareas que cooperan estrechamente probablemente deberían ser colocadas sobre cores que comparten una memoria caché.

Software como Hardware Locality (hwloc, desarrollado por el grupo de investigación Inria Runtime) permite abstraer información de topología y exponerlo de una manera portable. Aunque esta información está disponible para la capa superior tal como MPI o implementaciones OpenMP, sólo se utiliza para ubicar las tareas sobre los cores en forma segura y basado en políticas proporcionadas por el usuario.

Dentro de este proyecto, se propone colaborar en el desarrollo de sistemas de tiempo de ejecución que combinan características de aplicación con información de topología. Un objetivo es ofrecer automáticamente sugerencias de programación que tratan de respetar las afinidades de hardware y software. Además se desea analizar la convergencia de los resultados obtenidos de los algoritmos con la recientemente propuesta multi-BSP[Val11], modelo que considera niveles anidados de cálculos que corresponden a las capas naturales de arquitecturas de hardware hoy en día.

## 1.2 Algoritmos de Búsqueda de texto y sobre Espacios Métricos

La línea de investigación presentada en este trabajo, propone el diseño y desarrollo de algoritmos de búsquedas textuales sobre arquitecturas paralelas, utilizando el índice denominado Arreglo de Sufijos. Un primer avance ha sido publicado en [Ochoa13].

Los arreglos de sufijos (también conocido como PAT array [Manber93]) es un arreglo ordenado de todos los sufijos de una palabra. Si representamos cada sufijo de una palabra  $w$  con el número de la posición en la palabra  $w$  en la que empieza el sufijo, entonces podemos ver al arreglo de sufijos como una permutación de estos números de 1 a  $n$ . Finalmente, la permutación particular que el

arreglo de sufijos representa es la única permutación que ordena lexicográficamente los sufijos. En otras palabras, es un arreglo de enteros o punteros, y por lo tanto sólo está la información del orden lexicográfico de los sufijos, pero no hay información acerca del texto. Por lo que además del mencionado arreglo, es necesario contar con el texto.

Por otro lado, también se propone el diseño de algoritmos de búsqueda de datos multimediales. Los espacios métricos [Samet05] han demostrado ser útiles y prácticos para representar este tipo de datos multimediales. Además, permiten realizar búsquedas por similitud en grandes colecciones de objetos de datos. El objetivo en este tipo de búsquedas es recuperar los objetos más similares a una consulta dada. El grado de similitud entre dos objetos es determinado por una función que depende de la aplicación utilizada y se la denomina función de distancia. Generalmente, esta función es muy costosa de calcular y por este motivo se utilizan distancias pre-calculadas para indexar la base de datos a fin de reducir el número promedio de llamadas a esta función durante el proceso de búsqueda.

Estas aplicaciones complejas requieren soluciones que demandan mayor poder computacional. Para resolver este tipo de problemas, es posible modificar los programas secuenciales para que sean ejecutados en un ambiente distribuido o paralelo. Este tipo de modificaciones permite acelerar la velocidad de ejecución y/o mejorar la cantidad de memoria disponible.

El procesamiento de consultas sobre ambientes distribuidos fue abordado por primera vez en [Papa01]. En este trabajo, los autores presentan resultados analíticos y experimentales que muestran que es posible obtener un sistema escalable sobre este dominio de aplicaciones. El trabajo presentado en [Papa01] es extendido en [Gil09] en el contexto de un índice basado en clustering para colecciones de espacios métricos. Aquí, se estudian varias alternativas para distribuir el índice entre los procesadores, concluyendo que el índice basado en particionado global obtiene una mejor performance que el índice basado en particionado local. El particionado global se refiere a construir un único índice considerando la colección de objetos completa. Luego el índice se distribuye uniformemente entre los procesadores. Sin embargo, una

desventaja del particionado global es su potencial a procesar consultas en forma desbalanceada. Esta situación de desbalance surge cuando varias consultas tienden a seleccionar el mismo conjunto de procesadores. Usualmente, las consultas de los usuarios tienden a ser sesgadas a segmentos particulares de la colección de objetos y este sesgo suele cambiar en forma dinámica y de forma impredecible.

El trabajo presentado en [Marin10] y extendido en [Gil12] propone una solución al problema de desbalance del particionado global para un índice basado en clustering. Sin embargo, el método propuesto requiere el pre-procesamiento con un costo de  $O(n^2)$ , del índice secuencial de tamaño  $n$ , para asignar las diferentes partes del índice. Adicionalmente, esta solución falla cuando las consultas cambian su foco a lo largo del tiempo.

El trabajo presentado en [Bustos11, Bustos11b] presenta la implementación de un índice métrico en un ambiente híbrido de computación paralela que combina el uso de memoria compartida con pasaje de mensajes. Se muestra la construcción del índice métrico utilizando un enfoque de particionado local de datos. Luego se explica el proceso de búsqueda en una arquitectura híbrida, especificando las funciones que desempeñan los distintos nodos y threads durante el procesamiento de consultas.

Posteriormente, el trabajo [Mancini12] extiende el sistema híbrido propuesto en [Bustos11] para analizar diferentes técnicas de particionado global.

## 2. LINEAS DE INVESTIGACION y DESARROLLO

La línea de investigación descrita en las secciones anteriores involucra una serie de desarrollos individuales que en su conjunto logran obtener el objetivo planteado. Para ello, es necesario estudiar formas eficientes de monitorizar y/o estimar las comunicaciones y sincronizaciones en algoritmos BSP y Multi-BSP. Además para optimizar la eficiencia sobre las cotas establecidas y aumentar el speed-up de los algoritmos paralelos, es necesario determinar las características particulares de las aplicaciones.

Por otro lado es necesario estudiar las características inherentes en los índices tanto textuales como métricos para optimizar los

procesos de búsqueda y obtener el mejor rendimiento posible aplicando diferentes técnicas de particionado y actualización.

## 3. RESULTADOS OBTENIDOS ESPERADOS

### Algoritmos Paralelos modelados con BSP y Multi-BSP

Los resultados obtenidos hasta el momento son:

- Descubrimiento de la Arquitectura de Clusters en tiempo de ejecución.
- Affinity de tareas combinado con descubrimiento de Arquitectura.
- Estudio de librerías existentes que implementan las primitivas del modelo de programación BSP.

Los resultados esperados son:

- Diseño e implementación de un benchmark que permita medir el costo de sincronización en sistemas multi-core.
- Implementar metodologías que permitieran hacer una abstracción genérica de las topologías actuales y futuras de los clusters.

### Algoritmos de búsqueda

Los resultados obtenidos hasta el momento son:

- Diseño e implementación algoritmos de búsqueda sobre arquitecturas híbridas.
- Diseño e implementación de un índice métrico basado en un particionado de datos local y global.
- Diseño e implementación de un algoritmo de búsqueda textual para sistemas multi-core.

Los resultados esperados son:

- Extender los trabajos presentados en [Bustos11] y [Mancini12] para estudiar algoritmos eficientes para la actualización dinámica del índice.
- Estudio de técnicas eficientes para optimizar el uso de los diferentes niveles de memoria caché.
- Realizar pre-procesamiento de consultas para la asignación eficiente de las mismas a los threads.

#### 4. FORMACION DE RECURSOS HUMANOS

Actualmente, se cuenta con dos doctores en ciencias de la computación realizando la investigación teórica y dirección de los algoritmos propuestos. También se cuenta con un alumno de doctorado; un alumno de maestría y dos alumnos de grado. Uno de los cuales está próximo a finalizar su tesis.

Mediante este trabajo de investigación se podrán formar profesionales en el área de sistemas distribuidos y paralelismo que puedan modelar, diseñar e implementar algoritmos eficientes que se ejecuten en sistemas de clusters multi-core.

#### 5. BIBLIOGRAFIA

- [Allen07] E. Allen and D. Chase and J. Hallett and V. Luchangco and W. Maessen and S. Ryu and S. Tobin-Hochstadt, S. The Fortress Language Specification, version 1.0beta. 2007.
- [Bustos11] F. Bustos. Sistemas de búsquedas Multimediales en Ambientes Paralelos Híbridos. Tesis de Lic. en Cs. de la Computación. UNSL, Argentina 2011.
- [Bustos11b] F. Bustos, M. Alaniz, V. Gil-Costa and M. Printista. Hybrid Architecture for Metric Space Searches. CACIC 2011, La Plata, Argentina. Pp. 312-323.
- [Charles05] P. Charles and C. Grothoff and V.A. Saraswat and C. Donawa and A. Kielstra and K. Ebcioğlu and von Praun and V. Sarkar. X10: an objectoriented approach to non-uniform cluster computing. In International Conference on Object Oriented Programming, Systems, Languages and Applications. Pg. 519-538. 2005.
- [Gil12] V. Gil-Costa, and M. Marin, "Load Balancing Query Processing in Metric-Space Similarity Search", In 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid 2012), May 13-16, 2012, Canada.
- [Gil09] V. Gil-Costa, M. Marin, and N. Reyes. Parallel query processing on distributed clustering indexes. Journal of Discrete Algorithms (Elsevier), 7:3-17, 2009.
- [Hager11] G. Hager and G. Wellein. Introduction to High Performance Computing for Scientists and Engineers. CRC Press, New York, 2011.
- [Hennesy07] J. L. Hennesy and D. A. Patterson. Computer Architecture - A Quantitative Approach. Morgan Kaufmann, 4th edition, 2007
- [Hennesy08] J. L. Hennesy and D. A. Patterson. Computer Organization & Design - The Hardware/Software Interface. Morgan Kaufmann, 4th edition, 2008.
- [Manber93] Manber, U. and Myers, E. W. 1993. Suffix Arrays: A new method for on-line string searches. SIAM J. Comput. 22, 5, 935-948.
- [Mancini12] "Data Partitioning Evaluation for Multimedia Systems in Hybrid Environments". Virginia Mancini, Fabricio H. Bustos, Graciela Veronica Gil-Costa, Marcela Printista. 1st International Workshop on Soft Computing Techniques in Cluster and Grid Computing Systems SCCG 2012. November 12-14, 2012, University of Victoria, Victoria, Canada
- [Marin10] M. Marin, F. Ferrarotti and V. Gil-Costa, "[Distributing a Metric-Space Search Index onto Processors](#)", In 39th International Conference on Parallel Processing (ICPP 2010), San Diego, California, Sept. 13-16, 2010.
- [Ochoa13] "Suffix Array Performance Analysis for Multi-core Platforms". Cesar Ochoa, Veronica Gil Costa, Marcela Printista. 4th International SuperComputing Conference in Mexico (ISUM).5-8 Marzo 2013
- [Papoa01] N. Papadopoulos and Y. Manolopoulos. Distributed processing of similarity queries. Distrib. Parallel Databases, 9(1):67-92, 2001.
- [Rauber10] Thomas Rauber and Gudula Rünger. Parallel Programming For Multicore and Cluster Systems. Springer, 2010.
- [Reinders 07] J. Reinders. Intel Threading Building Blocks: Outfitting C++ for Multicore Processor Parallelism. O'Reilly (2007)
- [Samet05] H. Samet. "Foundations of Multidimensional and Metric Data Structures" (The Morgan Kaufmann Series in Computer Graphics and Geometric Modeling). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA. 2005.
- [Val90] L. Valiant. A bridging model for parallel computation. Communication of the ACM, Vol 33. pp 103-111. 1990.
- [Val11] L. G. Valiant. A bridging model for multi-core computing. J. Comput. Syst. Sci. 77(1): 154-166, 2011.

# Desarrollo y Sintonización Automática de Aplicaciones Paralelo/Distribuidas

Caymes-Scutari Paola<sup>1,2</sup>, Tardivo María Laura<sup>2</sup>, Méndez-Garabetti Miguel<sup>2</sup>, Cía Flavia Carolina<sup>2</sup>, Vega-Hissi Francisco Javier<sup>2</sup>, Diamante Diego<sup>2</sup>, Bianchini Germán<sup>2</sup>

<sup>1</sup>Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)

<sup>2</sup>Laboratorio de Investigación en Cómputo Paralelo/Distribuido  
Departamento de Ingeniería en Sistemas de Información  
Facultad Regional Mendoza/Universidad Tecnológica Nacional

Rodríguez 273 (M5502AJE) Mendoza

Tel. +54 261 5244579

pcaymesscutari@frm.utn.edu.ar, lauratardivo@dc.exa.unrc.edu.ar,  
miguelmendezgarabetti@gmail.com, ciacarolina@yahoo.com.ar, aehtipicus@gmail.com  
seba\_dmza@hotmail.com, gbianchini@frm.utn.edu.ar

## Resumen

El cómputo paralelo se encuentra en continua expansión dada la potencia y la velocidad con la que pueden obtenerse los resultados. Desafortunadamente, el desarrollo de aplicaciones paralelas constituye una tarea no trivial, dado que involucra una serie de aspectos adicionales a los meramente funcionales que inciden directamente en la eficiencia y en la calidad de los resultados esperados, tanto en lo que respecta a precisión como a tiempo de respuesta. Cuando se trata de un usuario no experto, los efectos negativos tienden a potenciarse dada la falta de experiencia y habilidad para subsanar los problemas. Es por ello que esta línea de investigación aborda el desarrollo de un entorno para el desarrollo automático y la sintonización automática de aplicaciones paralelas que haga transparente el proceso de resolución

del problema y la paralelización de la solución, mediante la instanciación de *problem solvers*. El usuario sólo clasifica y especifica el problema a resolver, mientras que la herramienta encapsula la resolución del problema. Como primer paso se trata la clase de problemas cuya resolución se lleva a cabo mediante algoritmos genéticos [Mic92].

**Palabras clave:** HPC, Cómputo Paralelo, Desarrollo Automático, Sintonización, Problem Solver

## Contexto

La línea de I+D presentada en este artículo es coordinada y desarrollada en el ámbito del LICPaD (Laboratorio de Investigación en Cómputo Paralelo/Distribuido) de la Universidad Tecnológica Nacional – Facultad Regional Mendoza – Departamento de



Ingeniería en Sistemas de Información (UTN-FRM-DISI). Dicha línea está presente en el proyecto de investigación titulado “Desarrollo Automático de Aplicaciones Paralelo/Distribuidas basadas en Algoritmos Genéticos”, el cual cuenta con financiamiento tanto de la UTN como de la FRM.

## Introducción

En los últimos años, los avances tecnológicos han propiciado un acelerado incremento en el rendimiento de los sistemas computacionales, muchos de ellos basados en algún tipo de paralelismo, coadyuvando a la creciente demanda de alto rendimiento para el tratamiento de problemas de gran tamaño y/o complejidad [Buy+99, Sta06]. No obstante, las aplicaciones para entornos paralelo/distribuidos son complejas de diseñar, las herramientas de programación paralelo/distribuida presentan diversas limitaciones, la ejecución de las aplicaciones no siempre hace un uso eficiente de los recursos computacionales involucrados y su optimización o sintonización muchas veces está fuera del alcance incluso del usuario disciplinar. El desarrollo de aplicaciones debe realizarse de una forma específica que permita su ejecución en un sistema paralelo/distribuido, lo que puede resultar una tarea complicada especialmente para usuarios no expertos dado que involucra el uso o aplicación de conceptos puramente informáticos como por ejemplo la concurrencia, los modelos de programación paralelo/distribuida (como Master/Worker o Pipeline [Mat+04]), las comunicaciones y/o diferentes tipos de middleware. Además, el desarrollo de aplicaciones distribuidas conlleva un conjunto de aspectos adicionales a los meramente algorítmicos que deben ser considerados para obtener un adecuado comportamiento, a saber la técnica de

descomposición, la granularidad, el grado de concurrencia, el balanceo de carga y la escalabilidad, entre otros [Gra+03]. A lo largo de los años han surgido diferentes aproximaciones para facilitar la tarea del usuario, como la utilización de librerías que encapsulan primitivas de comunicación (como PVM [Gei+94] o MPI [Gro+96, Sni+96]) o la utilización de esqueletos que implementan patrones de comportamiento paralelo [Bac+95, Kul01, Mac+02, Ser+02]. Cada una de estas opciones ofrece facilidades a diferentes niveles pero a la vez presentan ciertas restricciones y requieren del usuario un cierto grado de conocimiento de los conceptos relacionados con concurrencia y paralelismo para llevar a cabo una correcta implementación del programa. Algunas aproximaciones afrontan de alguna manera el análisis de concurrencia para simplificar y/o asistir la tarea del usuario no-experto, pero aún requieren cierto grado de conocimiento y experiencia (como por ejemplo [Stu+03, You+92, Ter07]). Si bien estas herramientas constituyen un medio para mejorar el diseño, no proveen un mecanismo explícito para diseñar y/o desarrollar programas concurrentes y/o paralelos. Esto se debe a que el paralelismo que pueda lograrse en un programa depende mucho de la naturaleza del problema, lo cual dificulta la generalización cuando de herramientas automáticas se trata. Por otro lado, debido a que el rendimiento constituye un aspecto clave, tanto la calidad de la aplicación creada como el comportamiento de la misma en función del conjunto de datos de entrada y/o del estado del entorno de ejecución deben ser contemplados ya que pueden influir drásticamente en la utilidad del programa. A lo largo de los años se ha utilizado una serie de índices para evaluar el rendimiento de las aplicaciones, tales como el *speedup*, la

escalabilidad, la eficiencia, o el balanceo de carga [Wil+05]. Sin embargo, tales índices proporcionan una ponderación general de alguna característica particular del programa, pero ninguno proporciona información específica ni de utilidad para modificar y en consecuencia mejorar el comportamiento del programa. Es por ello que resulta habitual recurrir al proceso de sintonización. El proceso de sintonización u optimización es un proceso que implica varias fases y que permite buscar y solucionar problemas concretos en un programa para asegurar que no existen cuellos de botella durante la ejecución de la aplicación. En primer lugar, durante la fase de monitorización debe registrarse información acerca del comportamiento de la aplicación. En segundo lugar se ejecuta el análisis de la información. El análisis del rendimiento permite hallar cuellos de botella, deducir sus causas y determinar las acciones necesarias para su eliminación. Finalmente, debe realizarse la sintonización de la aplicación a través de la implementación de los cambios apropiados en el código para solucionar los problemas y mejorar la performance [Buy+99]. Aún cuando las herramientas existentes para el análisis y sintonización de aplicaciones representan un importante recurso de gran ayuda y utilidad para simplificar y asistir de alguna manera la tarea del usuario, existe una problemática asociada a la amplia variedad de opciones, tecnologías y aproximaciones tanto para el desarrollo de aplicaciones como para la sintonización de las mismas. En dicho escenario, el usuario continúa necesitando un alto grado de conocimiento para decidir cuál o cuáles de las herramientas serán las más adecuadas según las características de la aplicación y el entorno de ejecución. Por otro lado, en general cada una de las herramientas utiliza un cierto conjunto

de formalismos propios que el usuario se ve obligado a aprender o modificar dependiendo de las características particulares de cada una de sus aplicaciones y de las herramientas que sea viable utilizar. Aquel usuario que no domina los conceptos de concurrencia y paralelismo queda ajeno a todo potencial de esta tecnología. En este contexto, es clara la necesidad de contar con herramientas que integren el proceso de desarrollo desde un nivel básico juntamente con el proceso de sintonización de aplicaciones paralelo/distribuidas, de modo que la intervención del usuario se vea simplificada.

Por todo lo expuesto, la línea de investigación presentada en este trabajo abarca dos grandes aspectos de la computación paralelo/distribuida: el desarrollo y la sintonización combinados en un mismo entorno. La arquitectura general del entorno se basa en tres módulos para cada *problem solver*. En primer lugar, una interface con el usuario que de soporte para la especificación de la aplicación. El segundo módulo o traductor, es responsable de traducir la información provista por el usuario a fin de instanciar el esqueleto del *problem solver*. Finalmente, el módulo de sintonización permite incorporar capacidades de adaptación de la ejecución a la aplicación. La sintonización de las aplicaciones resultará transparente al usuario, ya que durante la etapa de desarrollo –y de acuerdo a las especificaciones brindadas por el usuario– la aplicación se prepara automáticamente para poder ser sintonizada de acuerdo a las características que presente y al entorno de ejecución.

## Líneas de Investigación y Desarrollo

La línea de investigación presentada en este trabajo posee tres ejes principales: el desarrollo automático, la sintonización automática y los métodos de optimización (o *problem solvers*). En lo que hace al desarrollo automático, involucra los aspectos de reutilización y generación de código, así como también la amigabilidad con el usuario, a fin de ofrecer una herramienta de uso sencillo. Por su parte, la sintonización automática permite dotar a la aplicación paralela generada con la capacidad de adaptar su ejecución a las condiciones del ambiente sin la necesidad de que intervenga el usuario. Ambas componentes (desarrollo y sintonización) dependen del *problem solver* que esté bajo consideración, para lo cual se requiere una familiarización previa con cada *problem solver*, de manera que se identifique su funcionamiento y dependencias, sus oportunidades o posibilidades de paralelización, y sus problemas de rendimiento a ser sintonizados.

## Resultados y Objetivos

En la actualidad se cuenta con la infraestructura general del entorno. Como se ha mencionado anteriormente, el primer *problem solver* abordado ha sido el correspondiente a Algoritmos Genéticos, para el cual ya se cuenta tanto con la interface de usuario como con el módulo traductor, y se están incorporando las propiedades de sintonización automática. Al concluir la validación y depuración de los desarrollos alcanzados hasta el momento, se dará paso al tratamiento, diseño y desarrollo de otros *problem solver* que puedan adicionarse al entorno, de modo de ampliar de forma incremental su espectro de posibilidades para ofrecer soporte a las necesidades de los diferentes tipos de problemas (como búsqueda tabú,

*simulated annealing*, evolución diferencial, etc. [Tal09]).

## Formación de Recursos Humanos

Esta línea de investigación cuenta con la dirección de la Dra. Paola Caymes Scutari y la codirección del Dr. Germán Bianchini, quienes llevan adelante las tareas relacionadas con la planificación, administración y desarrollo del proyecto, así como también la formación de los alumnos de grado y postgrado que integran el LICPaD. En lo que hace a estudiantes de doctorado del LICPaD, esta línea de investigación cuenta con la participación de la Lic. María Laura Tardivo (cuyo plan de tesis doctoral versa específicamente dentro de esta línea de investigación) y del Ing. Miguel Méndez Garabetti, quienes cursan el doctorado en Ciencias de la Computación de la Universidad Nacional de San Luis, y a partir de abril del corriente año se incorporarán al programa de becas de CONICET (tipo I), realizando sus tareas doctorales en el marco del LICPaD. Además, se cuenta con la colaboración de tres estudiantes de grado de la carrera de Ingeniería en Sistemas de Información (Flavia Carolina Cía, Héctor Mazzucotelli y Diego Diamante) y de un alumno de la Universidad Católica (Javier Vega Hissi).

## Referencias

[Bac+95] Bacci B., Danelutto M., Orlando S., Pelagatti S., Vanneschi M., "P3L: A structured high level programming language and its structured support", *Concurrency: Practice and Experience*, 7(3):225-255, 1995.

- [Buy+99] Buyya R. et al, "High Performance Cluster Computing – Architectures and Systems (Volume 1)", Prentice Hall, 1999.
- [Gei+94] Geist A., Beguelin A., Dongarra J., Jiang W., Manchek R., Sunderam V., "PVM: Parallel Virtual Machine, A User's Guide and Tutorial for Network Parallel Computing", MIT Press. Cambridge, MA, 1994.
- [Gra+03] Gramma A., Gupta A., Karypis G., Kumar V., "Introduction to Parallel Computing (2° Ed)", Pearson Addison Wesley. 2003.
- [Gro+96] Groop W., Lusk E., Doss N., Skjellum A., "A high-performance, portable implementation of the MPI message passing interface standard", Parallel Computing, volume 22-6, pp.789-828, septiembre de 1996.
- [Kul01] Kulkarni S., "An intelligent framework for Master-Worker applications in a dynamic metacomputing environment", Computer Science Department. University of Wisconsin – Madison. 2001.
- [Mac+02] MacDonald S., Anvik J., Bromling S., Schaeffer J., Szafron D., Tan K., "From patterns to frameworks to parallel programs", Parallel Computing 28(12):1663-1684.
- [Mat+04] Mattson T., Sanders B., Massingill B., "Patterns for Parallel Programming", Addison-Wesley, 2004.
- [Mic92] Michalewicz Z., "Genetic Algorithms + Data Structures = Evolution Programs", Springer-Verlag, Berlin Heidelberg.
- [Ser+02] Sérot J., Ginhac D., "Skeletons for parallel image processing: an overview of the SKIPPER project", Parallel Computing 28(12):1685-1708, 2002.
- [Sni+96] Snir M., Otto S., Huss-Lederman S., Walker D., Dongarra J., "MPI: The Complete Reference". The MIT Press. Cambridge Massachusetts. London. England. 1996.
- [Sta06] Stallings W., "Organización y Arquitectura de Computadores". Pearson Prentice Hall. 2006.
- [Stu+03] Stuijk S., Ypma J., Basten T., "CAST - A Task-Level Concurrency Analysis Tool", 9th Annual Conference of the Advanced School for Computing and Imaging. ACSD 2003, pg 237-238. IEEE Computer Society.
- [Tal09] Talbi E., "Metaheuristics – From Design to Implementation". Wiley. 2009.
- [Ter07] Tersteeg A., "Application Concurrency Audit Tool: CFinder", <http://softwarecommunity.intel.com/articles/eng/1613.htm>. 2007.
- [Wil+05] Wilkinson B., Allen M., "Parallel Programming - Techniques and Applications Using Networked Workstations and Parallel Computers", Pearson Prentice Hall. Second Edition. 2005.
- [You+92] Young M., Taylor R., Forester K., Levine D., Brodbeck D., "A Concurrency Analysis Tool Suite: Rationale, Design and Preliminary Experience", Technical Report, TR-128-P. Software Engineering Research Center, Purdue University, Oct. 1992.



# Desarrollo de Aplicaciones Paralelo/Distribuidas orientadas a la Predicción de Incendios Forestales

Bianchini Germán<sup>2</sup>, Méndez-Garabetti Miguel<sup>2</sup>, Tardivo María Laura<sup>2</sup>, Caymes-Scutari Paola<sup>1,2</sup>

<sup>1</sup>Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)

<sup>2</sup>Laboratorio de Investigación en Cómputo Paralelo/Distribuido  
Departamento de Ingeniería en Sistemas de Información  
Facultad Regional Mendoza/Universidad Tecnológica Nacional

Rodríguez 273 (M5502AJE) Mendoza

Tel. +54 261 5244579

gbianchini@frm.utn.edu.ar, miguelmendezgarabetti@gmail.com,  
lauratardivo@dc.exa.unrc.edu.ar, pcaymesscutari@frm.utn.edu.ar

## Resumen

La problemática existente a raíz de la falta de exactitud presente en los parámetros de entrada en cualquier modelo científico o físico, puede producir consecuencias dramáticas en la salida del mismo si se trata éste de algún sistema crítico. Además, al citado problema deben sumarse las limitaciones impuestas por los propios modelos, las restricciones que agregan las soluciones numéricas y, por qué no, las provenientes de las propias implementaciones y versiones informáticas. Por tal motivo, resulta de gran interés el desarrollo de métodos y herramientas informáticos que se enfoquen en el tratamiento de la incertidumbre de los valores de entrada para lograr así una predicción lo más confiable posible por parte del modelo en cuestión. En el caso concreto de los incendios forestales, la simulación de la

propagación constituye un desafío desde el punto de vista computacional, dada la complejidad que involucran los modelos, los métodos numéricos y la administración de los recursos. La clase de métodos que aborda nuestra línea de investigación constituye una importante herramienta para la prevención y predicción, dado que provee información acerca del posible comportamiento del fuego y las zonas que corren mayor peligro.

**Palabras clave:** HPC, Cómputo Paralelo, Incendios Forestales, Reducción de Incertidumbre, Algoritmos Evolutivos, Estadística

## Contexto

La presente línea de I+D se enmarca en los siguientes proyectos: Sintonización y Evaluación de Aplicaciones Paralelo/Distribuidas

orientadas a la Predicción y Prevención de Desastres Naturales, financiado por CONICET y FONCyT (ANPCyT); Desarrollo y Sintonización de Aplicaciones Paralelo/Distribuidas orientadas a la Predicción y Prevención de Desastres Naturales, financiado por la UTN. En todos los casos, las actividades se llevan a cabo en el ámbito del LICPaD (Laboratorio de Investigación en Cómputo Paralelo/Distribuido) de la UTN-FRM.

## Introducción

El uso de modelos para representar sistemas físicos se ha tornado una estrategia común de trabajo en múltiples áreas científicas. Generalmente, estos modelos reciben un conjunto de parámetros de entrada representando condiciones particulares y proveen una salida que refleja la evolución del sistema. Además, es común que dichos modelos se encuentren integrados en herramientas de simulación que normalmente son ejecutadas en computadoras [Cho+08, Dal+08, Iva+03, Lea+06, Mos+05, Ods03, Rif+06]. En general, el tamaño del conjunto de datos y la complejidad de las operaciones sobre dichos datos requieren la utilización de sistemas de gran potencia para resolver el problema en el menor tiempo posible. Esta creciente necesidad de contar con sistemas de alto rendimiento ha orientado a los científicos hacia los sistemas paralelo/distribuidos, cuya utilización se ha propiciado gracias a los avances tecnológicos de los últimos años.

Sin embargo los modelos pueden exhibir ciertas limitaciones relacionadas con la gran cantidad de parámetros de entrada que manejan. Dichos parámetros suelen presentar algún tipo de incertidumbre debido a la imposibilidad de medirlos en tiempo real, y por lo tanto deben ser estimados

a partir de medidas indirectas. Además, en muchos casos, los modelos no pueden ser resueltos analíticamente y deben resolverse aplicando métodos numéricos que sólo son una aproximación de la realidad (aun sin considerar las limitaciones que presenta la traducción de estas soluciones cuando son llevadas a cabo mediante computadoras).

El enfoque más prometedor para solventar este problema es el uso de asimilación de datos en tiempo real combinado con algún método computacional para analizar la desviación de la predicción de acuerdo al comportamiento real, lo cual serviría para determinar los valores de los parámetros que reproducen el comportamiento correcto en el momento actual y usar dichos valores para el siguiente paso de simulación.

Algunos de los métodos existentes para la asimilación de datos para optimizar los parámetros de entrada [Abd04, Bev+01, Man+05, Tho+08], en general, operan sobre un gran número de valores de entrada y, por medio de algún tipo de optimización, se enfocan en la búsqueda de un único conjunto de valores que describa el comportamiento previo de la mejor manera. Por lo tanto, es de esperar que el mismo conjunto de valores pueda ser usado en el futuro inmediato. A los sistemas que aplican este tipo de metodología se los conoce como *Data Driven Methods with Unique Solution* o métodos conducidos por datos con una única solución.

Como trabajo antecedente, se encuentra el framework de optimización BBOF (*Black-Box Optimization Framework*) [Abd+06, Abd+02] el cual afronta el problema de la incertidumbre de datos mediante una estructura que permite la aplicación de diferentes métodos heurísticos: *Genetic Algorithms* [Bäc95], *Taboo Search* [Glo+97] y *Simulated Annealing* [Kir+83]. Tal framework aplica sus

métodos de la misma forma: cálculo intensivo, observación y ajuste de valores.

Sin embargo, a pesar de que esta clase de metodología mejora los resultados que pueden obtenerse mediante la simple simulación, los métodos *Data Driven* aquí comentados adolecen de un mismo problema: encuentran un único conjunto de valores, y, como se mencionó anteriormente, para aquellos parámetros que presentan un comportamiento dinámico, el valor hallado no resulta de utilidad para describir correctamente el futuro inmediato del modelo en cuestión.

## Líneas de Investigación y Desarrollo

La línea de I+D que se sigue en el marco de los proyectos mencionados en el apartado **Contexto**, abordan el análisis, diseño y desarrollo de métodos alternativos para tratar con el problema de la Incertidumbre en los Valores de los Parámetros de Entrada de modelos como el de Incendios Forestales. Dichos modelos centran sus bases en conceptos tales como el análisis estadístico, el cómputo paralelo/distribuido y la aplicación de algoritmos evolutivos. Los métodos desarrollados se hallan clasificados dentro de los *Data Driven Methods* (Métodos Conducidos por Datos), pero en una nueva rama denominada *Data Driven Methods with Multiple Overlapped Solution* (Métodos Conducidos por datos con Solución Solapada Múltiple), dado que la solución que aportan no se basa en el resultado de una única solución sino en la conjunción de una multiplicidad de casos estudiados. Su objetivo es hallar un patrón de comportamiento, independientemente de los valores de los parámetros del modelo en cuestión. En estos métodos, cada parámetro es

representado mediante un rango de valores y una cardinalidad. Se generan múltiples escenarios a través de diversas combinaciones de los parámetros de entrada de acuerdo a ciertos rangos, y entonces se evalúa el modelo para cada caso. Los resultados son agregados estadísticamente para determinar la probabilidad general. En el caso particular de los incendios forestales, esta agregación se utiliza para predecir el área quemada en el siguiente paso de simulación.

## Resultados y Objetivos

Como resultado de la línea de I+D se desarrollaron dos métodos. El primero, basado en los conceptos de experimentación factorial, análisis estadístico y cómputo paralelo, denominado *S<sup>2</sup>F<sup>2</sup>M (Statistical System for Forest Fire Management)* [Bia+05], se centra en hallar un patrón del comportamiento del modelo en el cual se aplica, a través de la generación de los posibles escenarios (combinaciones de valores de parámetros de entrada) y la evaluación del conjunto de resultados para ofrecer una tendencia en el comportamiento del modelo, ajustándolo con la observación actual del mismo.

El segundo, denominado *ESS (Evolutionary Statistical System)* [Bia+12], corresponde a la mejora del método previo a través de la incorporación de algoritmos evolutivos para tratar de forma más eficiente la generación de escenarios adecuada para la etapa de análisis estadístico.

En ambos casos, los resultados de la aplicación de ambos métodos sobre casos de incendios forestales reales, ha ofrecido muy buenos resultados, superando los valores alcanzados con la aplicación de otros métodos equivalentes [Bia+10, Bia+12]

Dentro de los objetivos futuros se tiene el desarrollo de diversas variantes

de ESS con aplicación de paralelismo a distintos niveles, ya que el desarrollo actual aplica el esquema paralelo a nivel de función de evaluación (*fitness function*). Actualmente se está trabajando en un esquema de islas [Tal09] en el cual la población de individuos (posibles escenarios) se vea distribuido en diversas subpoblaciones que se administren y puedan evolucionar de forma paralela.

Finalmente, también se están analizando otras heurísticas que permitan la mejora de los métodos actuales e incluso el desarrollo de otras metodologías que permitan resolver el problema de la incertidumbre en los valores de los parámetros incluso de maneras más eficientes.

## Formación de Recursos Humanos

Los proyectos mencionados previamente cuentan con la dirección del Dr. Germán Bianchini y la codirección de la Dra. Paola Caymes Scutari, quienes llevan adelante las tareas relacionadas con la planificación, administración y desarrollo de los mismos, así como también la formación de los alumnos de grado y postgrado que integran el LICPaD. En lo que hace a estudiantes de doctorado del LICPaD, esta línea de investigación cuenta con la participación del Ing. Miguel Méndez Garabetti (cuyo plan de tesis doctoral versa específicamente dentro de esta línea de investigación) y de la Lic. María Laura Tardivo. Ambos cursan el doctorado en Ciencias de la Computación de la Universidad Nacional de San Luis, y a partir de abril del corriente año se incorporarán al programa de becas de CONICET (tipo I), realizando sus tareas doctorales en el marco del LICPaD. Además, se cuenta con la colaboración de tres estudiantes

de grado de la carrera de Ingeniería en Sistemas de Información.

## Referencias

- [Abd+06] Abdalhaq B., Bianchini G., Cortés A., Margalef T., Luque E., “Between Classical and Ideal: Enhancing Wild-land Fire Prediction Using Cluster Computing”, *Journal of Cluster Computing Special Issue on cluster computing in science and engineering*. Vol 9, Num 3. pp. 329-343. 2006.
- [Abd04] Abdalhaq B., “A methodology to enhance the Prediction of Forest Fire Propagation”. Ph. D Thesis. Universitat Autònoma de Barcelona (Spain). June 2004.
- [Abd+02] Abdalhaq B., Cortés A., Margalef T., Luque E., “Evolutionary optimization techniques on computational grids”, LNCS 2329. International Conference on Computational Science 2002. pp. 513–522. 2002.
- [Bäc95] Bäck T., “Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms”, Oxford University. ISBN: 0-19-509971-0. 1995.
- [Bev+01] Beven K. J., Freer J., “Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems”, *Journal of Hydrology* 249. pp.11–49. 2001.
- [Bian+12] Bianchini G., Mendez Garabetti M., Caymes Scutari P., “Evolutionary-Statistical System for Uncertainty Reduction Problems in Wildfires” XVIII Congreso Argentino de Ciencias de la Computación (CACIC), Bahía Blanca, Argentina. XII Workshop de Procesamiento Distribuido y Paralelo (WPDP). ISBN: 978-987-1648-34-4, pp.230-238. 2012.



- [Bia+10] Bianchini G., Denham M., Cortés A., Margalef T., Luque E., "Wildland Fire Growth Prediction Method Based on Multiple Overlapping Solution". *Journal of Computational Science (JOCS)* Elsevier. Vol 1 Issue 4, pp.229-237. 2010.
- [Bian+05] Bianchini G., Cortés A., Margalef T., Luque E., "S2F2M – Statistical System for Forest Fire Management". *LNCS 3514*, pp. 427-434. 2005.
- [Cho+08] Chow T., He W., Chan A., Fong K., Lin Z., Ji J., "Computer modeling and experimental validation of a building-integrated photovoltaic and water heating system". *Applied Thermal Engineering*, Volume 28, Issues 11-12. pp. 1356–1364. 2008.
- [Dal+08] Dalforno C., Mostaccio D., Suppi R., Luque E., "Increasing the scalability and the speedup of a fish school simulator", *LNCS 5073 (PART 2)*, pp. 936–949. 2008.
- [Glo+97] Glover F., Laguna M., "Tabu Search", Kluwer Academic Publishers, Boston. ISBN: 0-79-238187-4. 1997.
- [Iva+03] Ivanovskayaa V., Enjashina A., Sofronova A., Makurina Y., Medvedevab N., Ivanovskii A., "Quantum chemical simulation of the electronic structure and chemical bonding in (6,6), (11,11) and (20,0)-like metal-boron nanotubes". *Journal of Molecular Structure: THEOCHEM*. Volume 625, Issues 1-3. pp. 9–16. 2003.
- [Kir+83] Kirkpatrick S., Gelatt C., Vecchi M., "Optimization by Simulated Annealing" ,*Science*, 220, 4598, 671-680, 1983.
- [Lea+06] Learmount J., Taylor M., Smith G., Morgan C., "A computer model to simulate control of parasitic gastroenteritis in sheep on UK farms". *Veterinary Parasitology*, Volume 142, Issues 3–4. pp. 312-329. 2006.
- [Man+05] Mandel J., Bennethum L. S., Chen M., Coen J. L., Douglas C. C., Franca L. P., Johns C. J., Kim M., Knyazev A. V., Kremens R., Kulkarni V., Qin G., Vodacek A., Wu J., Zhao W., Zornes A., "Towards a Dynamic Data Driven Application System for Wildfire Simulation", *LNCS 3515*, pp. 632-639. 2005.
- [Mos+05] Mostaccio D., Suppi R., Luque E., "Simulation of Ecologic Systems Using MPI". *PVM/MPI 2005*: pp. 449–456. 2005.
- [Ods03] Odstrcil D., "Modeling 3-D solar wind structure".*Advances in Space Research*, Volume 32, Issue 4. pp. 497–506. 2003.
- [Rif+06] Riffat S., Ma X., Wilson R., "Performance simulation and experimental testing of a novel thermoelectric heat pump system". *Applied Thermal Engineering*, Volume 26, Issues 5-6. pp. 494–501. 2006.
- [Tal09] Talbi, El-Ghazali, "Metaheuristics – From Design to Implementation" ISBN 978-0-470-27858-1. Wiley. 2009.
- [Tho+08] Thorndahl S., Beven K.J., Jensen J.B., Schaarup-Jensen K. "Event based uncertainty assessment in urban drainage modelling, applying the GLUE methodology", *Journal of Hydrology* 357 (3-4), pp. 421–437. 2008.

## Construyendo un Sistema de Cómputo Distribuido Multipropósito

Oscar Martín Bianchi<sup>(a,b)</sup>  
[mbianchi@ejercito.mil.ar](mailto:mbianchi@ejercito.mil.ar)

José Ignacio Ariznabarreta Fossati<sup>(b)</sup>  
[ignacioariznabarreta@gmail.com](mailto:ignacioariznabarreta@gmail.com)

Alejandro J. M. Repetto<sup>(a,b)</sup>  
[arepetto@ejercito.mil.ar](mailto:arepetto@ejercito.mil.ar)

<sup>(a)</sup>CIDESO<sup>0</sup>, DIGID<sup>1</sup> - Ejército Argentino

<sup>(b)</sup>EST<sup>2</sup>, IESE<sup>3</sup> - Ejército Argentino

**RESUMEN**

El concepto de sistemas distribuidos, como la tecnología asociada a dichos desarrollos, es un concepto con una popularidad en constante crecimiento, así como un poderoso paradigma.

Dentro del ámbito académico en el que se enmarca este estudio, la computación distribuida prueba ser una herramienta valiosa para la experimentación tanto científica experimental como práctica, en cualquiera de las especialidades de ciencias donde se las desee enmarcar. Esto se debe en gran parte a la posibilidad de obtener grandes capacidades de cómputo con a bajo costo para quienes se plantean el uso del modelo, especialmente si se lo asocia con los sistemas de Computación de Escritorio de Grilla (Desktop Grid Computing).

En este entorno, la incorporación de tecnologías de computación de alto rendimiento a bajo costo, por ejemplo la computación voluntaria, surge como una opción interesante, especialmente por parte de organizaciones que cuentan con grandes cantidades de puestos de trabajo que presentan capacidad de cómputo ociosa.

En el presente trabajo se expondrán los avances realizados hacia la obtención de un sistema de computación distribuida multipropósito, que dé apoyo a los laboratorios de las distintas especialidades de ingeniería de la EST.

**Palabras Clave:** *Computación Distribuida, Multipropósito, Infraestructura como Servicio,*

*Webservices, Computación en Grilla de Escritorio.*

**CONTEXTO**

En las últimas dos décadas, las tecnologías orientadas a la Computación en Grilla (*Grid Computing*), experimentaron un avance exponencial de la mano de proyectos como Globus<sup>a</sup> y BOINC<sup>b</sup>, los cuales pusieron a disposición del público general la posibilidad de crear y gestionar proyectos de alta complejidad técnica.

Específicamente, la arquitectura de Computación en Grilla de Escritorio (*Grid Desktop Computing – GDC*)(1), es una forma particular de computación distribuida que implica la utilización de grandes grillas de computadores personales de forma organizada y sincronizada de manera que funcionen como un único computador, ofreciendo grandes cantidades de recursos con un costo relativamente bajo. Se pueden referenciar aplicaciones de este tipo de arquitectura en áreas críticas como la medicina (proyecto GENOMA), el estudio del espacio (con el proyecto SETI), el criptoanálisis, la toma de decisiones, el análisis de imágenes por software, entre otros.

En lo que hace al ámbito de aplicación dentro del sistema el Ejército Argentino, el Centro de Investigación y Desarrollo de Software del Ejército Argentino (CIDESO), implementó estos GDC en diversos proyectos, aplicándolo sobre todo a Sistemas de Simulación Constructiva Distribuidos. Estos sistemas

<sup>a</sup><http://www.globus.org/toolkit/>

<sup>b</sup><http://boinc.berkeley.edu/>

<sup>0</sup> CIDESO: Centro de Investigación y Desarrollo de Software

<sup>1</sup> DIGID: Dirección General de Investigación y Desarrollo

<sup>2</sup> EST: Escuela Superior Técnica - Facultad de Ingeniería del Ejército Argentino

<sup>3</sup> IESE: Instituto de Enseñanza Superior del Ejército - Universidad del Ejército Argentino

usualmente requieren un alto poder de cómputo y, por ser de simulación constructiva, una respuesta en cuasi tiempo real. La implementación de GDC al problema logró solucionar los problemas asociados, proveyendo una arquitectura adaptativa al problema del crecimiento del cómputo junto con el crecimiento en el número de participantes de los juegos (2)(3). Con relación a esto último, la mayoría de las fuerzas armadas y de seguridad alrededor del mundo están incorporando este tipo de tecnologías en asuntos críticos como las ejecuciones de modelos de propagación de epidemias, incendios, el uso de sensores distribuidos para facilitar la asistencia durante catástrofes o el control de zonas de alto riesgo, entre otras aplicaciones.

Desde otro punto de vista, en lo que hace al ámbito académico particularmente, la Escuela Superior Técnica del Ejército Argentino (EST), posee experiencias previas en el uso de estos conceptos en sus laboratorios de investigación. La primera incursión en el uso de GDC estuvo relacionada a aplicaciones asociadas con la criptografía, orientada al desarrollo científico y evaluación de algoritmos para la seguridad de la información, dentro del CriptoLab.

Entre las investigaciones llevadas a cabo por el CriptoLab se puede destacar, por ejemplo, la de identificar y evaluar algoritmos de factorización de grandes números – esencial en la seguridad de la información actual –, la cual requería de gran poder de cálculo, difícil de obtener incluso con equipos de alto rendimiento pero disponible a través de la utilización de computación paralela y/o distribuida. Por eso, y gracias a un estrecho trabajo en conjunto entre el CIDESO y los laboratorios de la EST, surgió la posibilidad de realizar investigaciones en conjunto, aportando las experticias de cada uno, logrando una sinergia en la investigación.

Así nace el proyecto CODEC, de cuyos adelantos y propuestas futuras se ocupara el presente documento.

## INTRODUCCIÓN

Luego de las experiencias obtenidas durante el año 2011 en el desarrollo del proyecto CoDiSe (4), se llevó adelante durante el año 2012 el proyecto CODEC (5), el cual, a diferencia del prototipo experimental, que solo estaba orientado a solucionar una problemática específica (validación de algoritmos asociados a la seguridad informática), está concebido para brindar apoyo a un abanico variado de necesidades de cómputo asociados a las distintas ramas de la ingeniería. Dichas necesidades varían enormemente tanto en su alcance como en su disciplina, yendo del contexto de la educacional civil, al de la aplicación militar, incluyendo a su vez proyectos mixtos entre las dos áreas de interés. Específicamente, se pretende que el proyecto CODEC aporte en los próximos años soluciones concretas a los laboratorios de las carreras de Ingeniería Química, Informática y Mecánica, realizando para cada una de ellas experimentos, pruebas y evaluaciones que de otra forma serían impracticables debido a su costo en tiempo y recursos.

Así mismo, en el marco de los organismos de Defensa y Seguridad, las aplicaciones de sistemas del tipo de CD y GDC sufrieron un importante incremento en los últimos años, específicamente en sus usos asociados en el apoyo a la comunidad y el control de crisis (a través de modelos de simulación predictivos de alta complejidad de cálculo). Además del uso exclusivo para un sector, podemos mencionar el caso de una utilización mixta, en donde estas tecnologías son utilizadas por laboratorios y universidades ajenas al ámbito castrense con el fin de realizar tareas de testing y certificaciones sobre los equipos y medios de comunicaciones utilizados por las distintas fuerzas; entre otros.

## LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

CODEC nos acerca en su propuesta la utilización de la infraestructura de cómputo ociosa disponible de la universidad, para aportar soluciones a problemáticas surgidas de la necesidad de realizar complejas experimentaciones, que requieren poderosos equipos de cómputo, para complementar el dictado de las clases, las prácticas profesionales, o las investigaciones de cualquier tipo.

Por lo mencionado, es que el sistema plantea dos grandes líneas de investigación, las cuales incluyen:

- Proveer a la universidad de una herramienta de cómputo multipropósito para sus investigaciones a un bajo coste.
- Utilización y aprovechamiento de la infraestructura informática ociosa disponible a través de las políticas incluidas en el VC.

De esta forma, las líneas de investigación que se llevaran a cabo tomando como base el prototipo realizado serán:

- Sistemas de cómputo distribuido multipropósito.
- Explotación de recursos computacionales ociosos en grandes organizaciones.
- Inclusión de BOINC en la nube (a través de webservices)
- Ejecución distribuida con foco multidisciplinario.

Este último apartado merece una mención especial, ya que las aplicaciones específicas de las distintas disciplinas son índole muy variado y de gran importancia a nivel académico. Entre ellas podemos mencionar:

- Ingeniería Química: Para dar apoyo a experimentos asociados al cálculo de densidad de compuestos, basados en el modelo DFT (*Density Functiona lTheory*).
- Ingeniería Informática: Como se demostró con el proyecto CoDiSe, el uso de un sistema con las características del

planteado aplicado a la seguridad informática (tanto para el cálculo teórico, así como para la experimentación práctica) es de gran utilidad y provee un importante recurso.

- Ingeniería Mecánica: Cálculos como los asociados a la resistencia y estabilidad de los distintos componentes y materiales que forman parte de un motor o mecanismo sensible, así como la ejecución de modelos de simulación que permitan detectar y prevenir fallas desde las fases de desarrollo asociadas al diseño.

## RESULTADOS ESPERADOS

Los objetivos propuestos para el próximo ciclo cubren desde la interacción del usuario para con el sistema hasta las cualidades que presente dicho sistema para con la arquitectura disponible.

Un objetivo es la consolidación de la interfaz proveedora de servicios del Sistema de Cómputo Distribuido con el fin de que la interacción usuario(no especializado)-computador pueda realizarse de forma más sencilla y que el objetivo principal de acercar una solución tratando de abstraer a los usuarios de la complejidad del sistema. El concepto subyacente es el de IaaS (*Infrastructure as a Service*). Esto se analizará mediante la estrategia de prototipos evolutivos, recurriendo a los docentes especialistas en cada rama de la ingeniería que queramos avanzar, con el fin de poder determinar el grado de aceptación que pueda tener dicha interfaz.

Para lograr el mayor nivel de usabilidad posible, se dividirá el problema en tres aspectos: acceso, envío de trabajos y respuesta de resultados. El primero se refiere al control de la seguridad básica del sistema en conjunto con un interface tipo panel de control. El envío de trabajos abarca la estandarización de las interfaces de BOINC reduciendo la complejidad a través de unmeta-lenguaje de *scripting*. Por último, la entrega de resultados también se abordará



como un esquema similar, estandarizando las interfaces de modo de unificar la visión del lado del usuario final.

Además de los aspectos amenos al usuario y la funcionalidad para con respecto a este, se propone un objetivo importante para garantizar el funcionamiento de los nodos (aquellos computadores cuya función será ceder los recursos) la generación de métricas de rendimiento del sistema, con el fin de poder evaluar el funcionamiento del sistema.

## FORMACIÓN DE RECURSOS HUMANOS

Desde el punto de vista de su concepción, CODEC es fruto del trabajo conjunto entre dos centros de investigación enmarcados en el seno del Ejército Argentino, con la particularidad que uno de dichos centros es parte funcional de la EST. Esto pone al proyecto en un ámbito privilegiado para la formación de recursos humanos.

Por un lado la EST, específicamente la cátedra de Lenguajes de Programación I y el CriptoLab, de esa misma institución, tiene amplia experiencia tanto en la investigación básica en el terreno de computación distribuida como en el de la criptografía respectivamente. Así mismo, en los laboratorios de la institución se dispone del material y los recursos humanos específicos tanta para la investigación en estas áreas como para dar apoyo a los alumnos a cargo del proyecto.

Por otro lado, el CIDESO tiene experiencia en la formación de recursos humanos en el terreno de la investigación aplicada en sistemas de información de diversa índole, incluyendo sistemas de simulación para el adiestramiento, sistemas de información geográfica, sistemas de visualización, sistemas inteligentes, sistemas móviles, sistemas de comunicación de alta complejidad y sistemas de cómputo de alto rendimiento.

Ambos laboratorios, a través del dictado de materias de grado en Ingeniería Informática, aportan recursos humanos a la Escuela Superior Técnica del Ejército. Es así que

investigadores de los laboratorios son jefes de cátedra en la EST y, de manera análoga, alumnos de la escuela aportan sus análisis a los laboratorios a través de trabajos prácticos de laboratorio, prácticas profesionales supervisadas o tesis y tesinas de grado y posgrado.

En particular, el prototipo experimental del cual nace la idea de CODEC fue implementado por los alumnos de tercer año de Ingeniería Informática y Electrónica como trabajo práctico de laboratorio de la materia Lenguajes de Programación I, y tutorado por investigadores del CIDESO y el CriptoLab, que a la vez son docentes de dicha cátedra.

Para el próximo paso, se pretende continuar con esta interacción fluida entre los centros de investigación y el alumnado, formado profesionales con conocimientos de campo en el terreno de la computación de alto rendimiento y un conocimiento acabado sobre temas criptográficos.

Además, al expandirse el sistema para ser aplicado en cualquier problema que requiera altos niveles de cómputo, se pretende incorporar alumnos y docentes de otras cátedras, de cualquiera de las ingenierías que se dictan en la Facultad.

Por otro lado, se continuará con la formación de profesionales en investigación utilizando estas nuevas tecnologías en los dos laboratorios participantes a través de la utilización de la GDC.

Así, pues, se formarán recursos humanos de todos los niveles, grado, posgrado o investigadores activos, incorporando más alumnos a los laboratorios y, potencialmente, becarios que se dediquen de modo formal (no sólo académico) a la profundización de los modelos propuestos.

## Referencias

1. **Berstis, Viktors.** *Fundamentals of Grid Computing*. s.l. : Redbooks Paper, IBM Corp, 2002.
2. *Hybrid Architecture for Constructive Interactive Simulation: Evaluation and Outcomes.* **Repetto, Alejandro J. M.**

Orlando, FL : s.n., 2010. /ITSEC'10, Interservice/Industry Training, Simulation and Education Conference.

3. *Grid Desktop Computing for Constructive Battlefield Simulation*. **Repetto, Alejandro J. M.** San Salvador de Jujuy : Red UNSI, 2009. XV Congreso Argentino de Ciencias de la Computación (CACIC 2009). 978-897-24068-4-1.

4. **Schroeder, Ralph.** *The Usability of Collaborative Virtual Environments and Methods for the Analysis of Interaction*. [ed.] University of Oxford Oxford Internet Institute. Oxford : Massachusetts Institute of Technology, 2006.

5. **Craig, Donald.** *Advantages of Simulation*. St. John's : Memorial University, Faculty of Science.

6. **Moriello, Sergio A.** *Inteligencias Sintéticas*. Buenos Aires, Argentina : Alsina, 2001.

7. *Interoperability-Ready, Training-Focused Architecture for Command and Control Systems*. **Repetto, Alejandro Juan Manuel**. Orlando, FL : s.n., 2011. Interservice/Industry Training, Simulation and Education Conference.

8. *Framework de Interoperabilidad para Sistemas de Comando y Control*. **Repetto, Alejandro Juan Manuel**. El Calafate, Santa Cruz : WICC'10, 2010. ISBN 978-950-34-0652-6.

9. **Ajulo, Adeshina B.** *Grid Computing An Advancement of E-Science To Computing And Beyond*. s.l. : Department of Computer Science, Federal University of Technology.

10. *Computación Distribuida para Cálculo Científico*. **Bianchi, Oscar Martín and Repetto, Alejandro**. Posadas : XIV Workshop de Investigadores en Ciencias de la Computación, 2012.

11. **Bianchi, Oscar Martín and Repetto, Alejandro.** *Sistema de Computacion Distribuida para Seguridad Informatica*. Posadas, Misiones : Workshop de Investigadores en Ciencias de la Computación 2012, 2012.

12. **Computerized Bussiness Solutions.** *Centralized vs Distributed Computing: White Paper*. 2007.

13. **Berman F, G. C. Fox and A. J. G. Hey.** *Grid Computing: Making the global infrastructure a reality*. s.l. : Wiley publishers, 2003.

14. **David P. Anderson, Eric Korpela, Rom Walton.** *High-Performance Task Distribution for Volunteer Computing*. Berkeley : Space Sciences Laboratory, University of California.

15. **Rodrigo Neves, Nuno Mestre, Francisco Machado, Joao Lopez.** *Parallel and Distributed Computing BOINC Grid Implementation*.

16. *The Challenge of Volunteer Computing*. **Christensen, Carl, Aina, Tolu and Stainforth, David**. Oxford, United Kingdom : Clarendon Laboratory, University of Oxford, 2009.

## Búsquedas por Rango sobre Plataformas GPU en Espacios Métricos

Osiris SOFIA<sup>1</sup>, Jacobo SALVADOR<sup>1</sup>, Eder DOS SANTOS<sup>1</sup>  
Roberto URIBE PAREDES<sup>2</sup>

1: Unidad Académica Río Gallegos, Universidad Nacional de La Patagonia Austral.  
2: Departamento de Ingeniería en Computación, Universidad de Magallanes, Chile.

[sistemasuarg@gmail.com](mailto:sistemasuarg@gmail.com)

### RESUMEN

La búsqueda por similitud consiste en recuperar todos aquellos objetos dentro de una base de datos que sean parecidos o relevantes a una determinada consulta. Actualmente es un tema de gran interés para la comunidad científica debido a sus múltiples campos de aplicación, como reconocimiento de patrones, recuperación de la información, bases de datos multimedia, entre otros. La búsqueda por similitud o en proximidad se modela matemáticamente a través de un espacio métrico, en el cual un objeto es representado como una caja negra, donde la única información disponible es la función de distancia de este objeto a los otros.

En general, el cálculo de la distancia es costoso, por ello el objetivo es reducir la cantidad de evaluaciones de distancia necesarias para resolver la consulta. Para esto se han desarrollado numerosas estructuras métricas, que realizan un preprocesamiento de los datos a fin de disminuir las evaluaciones de distancia al momento de la búsqueda.

En la actualidad, la necesidad de procesar grandes volúmenes de datos hace poco factible la utilización de una estructura en aplicaciones reales si ésta no considera la utilización de sistemas de altas prestaciones en entornos de procesamiento paralelo. Existen una serie de tecnologías para realizar implementaciones paralelas, siendo una de las más nuevas, las plataformas basadas en GPU / Multi-GPU, que son interesantes debido a la cantidad de procesadores y los bajos costes involucrados.

**Palabras clave:** Búsquedas por similitud, Espacios Métricos, Paralelismo, GPU, CUDA, Bases de Datos.

### CONTEXTO

Este trabajo está enmarcado en el proyecto 29/A274-1 de la Unidad Académica Río Gallegos de la Universidad Nacional de la Patagonia Austral.

### 1. INTRODUCCION

La búsqueda de objetos similares sobre un gran conjunto de datos se ha convertido en una línea de investigación de gran interés. Aplicaciones de estas técnicas pueden ser encontradas en reconocimiento

de voz e imagen, en problemas de minería de datos, detección de plagios y muchas otras.

En general, se utilizan diversas estructuras de datos con vistas a mejorar la eficiencia en términos de los cálculos de distancia, comparados con la búsqueda secuencial (algoritmo *fuerza bruta*) dentro la base de datos.

Por otro lado, se busca reducir los costos en términos de tiempo mediante el procesamiento en paralelo de grandes volúmenes de datos [12]. Existe una variada gama de plataformas paralelas, siendo una de las más recientes aquella basada en tarjetas gráficas GPU (*Graphic Processor Units*). En este sentido, estas nuevas tecnologías permiten un alto nivel de paralelismo a un muy bajo coste.

#### 1.1 Búsqueda por Similitud

La similitud se modeliza en muchos casos a través de un espacio métrico, y la búsqueda de objetos más similares a través de una búsqueda por rango o de  $k$  vecinos más cercanos. Un espacio métrico es un conjunto  $X$  con una función de distancia  $d : X^2 \rightarrow \mathfrak{R}$ , tal que  $\forall x, y, z \in X$ , se deben cumplir las propiedades de: positividad ( $d(x, y) \geq 0$  and  $d(x, y) = 0$  ssi  $x = y$ ),

simetría ( $d(x, y) = d(y, x)$ ) y desigualdad triangular ( $d(x, y) + d(y, z) \geq d(x, z)$ ).

Sobre un espacio métrico  $(X, d)$  y un conjunto de datos finito  $Y \subseteq X$ , se pueden realizar una serie de consultas. La consulta básica es la *consulta por rango*. Sea una consulta  $x \in X$ , y un rango  $r \in \mathfrak{R}$ . La consulta de rango alrededor de  $x$  con rango  $r$  es el conjunto de puntos  $y \in Y$ , tal que  $d(x, y) \leq r$ . Un segundo tipo de consulta, que puede construirse usando la consulta por rango, es los *k vecinos más cercanos*. Sea una consulta  $x \in X$  y un entero  $k$ . Los  $k$  vecinos más cercanos a  $x$  son un subconjunto  $A$  de objetos de  $Y$ , donde  $|A| = k$  y no existe un objeto  $y \in A$  que  $d(x, y)$  sea menor a la distancia de algún objeto de  $A$  a  $x$ .

#### 1.2 Estructuras Métricas Basadas en Pivotes

Existe una serie de estructuras de datos métricas e índices cuyo objetivo es reducir las evaluaciones

de distancias durante la búsqueda, y con ello disminuir el tiempo de procesamiento. Algunas de las estructuras basan la búsqueda en pivotes y otras en *clustering*.

Los algoritmos basados en *clustering* dividen el espacio en áreas o planos, donde cada área tiene un centro y se almacena alguna información sobre el área que permita descartarla completa solo con comparar la consulta con su centro. Ejemplos de estos son BST, GHT, M-Tree, GNAT, EGNAT y muchos otros [1].

En los métodos basados en pivotes, se selecciona un conjunto de pivotes y se precálculan las distancias entre los pivotes y todos los elementos de la base de datos. Los pivotes sirven para descartar objetos durante la búsqueda utilizando la desigualdad triangular. Algunas estructuras de este tipo son: LAESA [2], FQT y sus variantes [3], Spaghettis y sus variantes [4], FQA [5], SSS-Index [6] y otros [10].

Para la construcción de una estructura basada en pivotes, se selecciona un conjunto de pivotes  $p_1, \dots, p_k$ , los cuales pueden o no pertenecer a la base de datos a indexar.

Existen estructuras de tipo árbol y de tipo arreglo, estas últimas son más adecuadas para ser implementadas en plataformas basadas en GPU [7].

Una estructura métrica genérica (*Generic Metric Structure: GMS*) puede considerarse como una tabla de distancias entre los pivotes y todos los elementos de la base de datos. Es decir, cada celda almacena la distancia  $d(y_i, p_j)$ , siendo  $y_i$  un elemento de la base de datos.

Durante la búsqueda por rango sobre esta estructura dada una consulta  $q$  y un rango  $r$ , el algoritmo sería:

1. Para cada consulta  $q$ , se calcula la distancia entre  $q$  y todos los pivotes  $p_1, \dots, p_k$ . Con esto se obtienen  $k$  intervalos de la forma  $[a_1, b_1], \dots, [a_k, b_k]$ , donde  $a_i = d(p_i, q) - r$  y  $b_i = d(p_i, q) + r$ .
2. Los objetos, representados en la estructura por sus distancias a los pivotes, son candidatos a la consulta  $q$  si todas sus distancias están dentro de todos los intervalos.
3. Para cada candidato  $y$ , se calcula la distancia  $d(q, y)$ , y si  $d(q, y) \leq r$ , entonces el objeto  $y$  es solución a la consulta  $q$ .

La Figura 1 representa una estructura de datos métrica genérica (GMS) construida con 4 pivotes.

	1	2	3	4	link	
0	1	6	5	1		Objeto 1
8	7	5	6	2		Objeto 2
6	5	0	7	3		Objeto 3
5	6	7	0	4		Objeto 4
15	14	13	14	5		Objeto 5
10	9	9	7	6		Objeto 6
9	9	7	6	7		Objeto 7
7	8	7	7	8		Objeto 8
5	4	6	6	9		Objeto 9
8	7	7	8	10		Objeto 10
1	0	5	7	11		Objeto 11
2	2	8	6	12		Objeto 12
8	7	6	8	13		Objeto 13
8	9	6	9	14		Objeto 14
6	7	6	7	15		Objeto 15
11	2	10	10	16		Objeto 16
2	2	6	6	17		Objeto 17

**Figura 1. Búsqueda sobre una estructura métrica genérica (GMS) de 4 pivotes. Para cada consulta  $q$  con distancias a los pivotes  $d(q, p_i) = \{8, 7, 4, 6\}$  y un rango de búsqueda  $r = 2$ , define los siguientes intervalos  $\{(6, 10), (5, 9), (2, 6), (4, 8)\}$ . Las celdas marcadas en gris oscuro son aquellas que están dentro del intervalo de búsqueda. Las celdas tachadas con líneas son los objetos candidatos (2, 13 y 15), que serán evaluados directamente con la consulta.**

### 1.3 Unidades de Procesamiento Gráfico

Actualmente las unidades de procesamiento gráfico (GPU) disponen de un alto número de *cores*<sup>1</sup> con un alto ancho de banda. Este tipo de dispositivos permite aumentar la capacidad de procesamiento respecto de las CPU [8] y arquitecturas *multi-core* [19]. Una tendencia denominada *Computación de Propósito General sobre GPU* o GPGPU ha orientado la utilización de GPU sobre nuevos tipos de aplicaciones.

En general, las GPU utilizan todos sus recursos en tener la mayor cantidad de elementos de procesamiento y gran ancho de banda con memoria, lo que posibilita una potencia de cálculo superior al de las CPUs. Así, uno de los objetivos de las GPU es aumentar la capacidad de procesamiento explotando el paralelismo a nivel de datos para problemas paralelos o que se puedan paralelizar.

Existen nuevos lenguajes o extensiones para los lenguajes de alto nivel propuestos por algunos fabricantes; un ejemplo es CUDA [11]. El modelo de Programación CUDA de NVIDIA considera a la

<sup>1</sup>En la actualidad, se encuentran disponibles modelos de GPU con más de 2600 *cores*.



GPU como un dispositivo capaz de ejecutar un alto número de hilos en paralelo sobre la GPU (*device*) y su DRAM (*device memory*). CUDA incluye herramientas de desarrollo de software C/C++ y un mecanismo de abstracción que oculta los detalles de hardware al desarrollador.

En CUDA los hilos son organizados en bloques del mismo tamaño y los cálculos son distribuidos en una malla o *grid* de bloques. Estos hilos ejecutan el código de la GPU, denominado *kernel*. Un *kernel* CUDA ejecuta un código secuencial en un gran número de hilos en paralelo. Los hilos en un bloque pueden trabajar juntos eficientemente, intercambiando datos localmente en una memoria compartida y sincronizando a baja latencia en la ejecución mediante barreras de sincronización. Por el contrario, los hilos ubicados en diferentes bloques pueden compartir datos, solamente accediendo a la memoria global de la GPU o *device memory*, que es una memoria de más alta latencia.

#### 1.4 GPU y espacios métricos

Existe una importante cantidad de índices o estructuras métricas para búsquedas sobre espacios métricos. Sin embargo, existe una serie de características poco comunes en estas estructuras que hacen difícil su implementación directa para aplicaciones reales. La primera tiene relación con capacidades dinámicas, la mayoría de estas estructuras deben ser reconstruidas si existen nuevos objetos que indexar o si hubiese que eliminar objetos de la base de datos.

Otra característica, aún menos frecuente, está relacionada con estructuras que permitan la manipulación eficiente de los datos en memoria secundaria, donde deben ser considerados parámetros de costo adicional como son el número de accesos a disco y el tamaño del índice entre otros. En si, dichas estructuras han sido diseñadas como prototipos, por lo que su utilización en aplicaciones reales no ha sido probada, mas aún considerando que en aplicaciones reales los objetos son de gran tamaño y/o la base de datos es considerablemente grande, donde es posible pensar que sólo un bajo porcentaje de datos podría entrar en memoria principal. En general, las estructuras métricas no consideran la jerarquía de memoria, por ello, resulta relevante considerar este aspecto para poder obtener un mayor rendimiento y eficiencia ante la posibilidad de utilizar nuevas tecnologías, como es el caso de las GPU que poseen distintos niveles de memoria dentro de su sistema.

Otra característica es la relación con el procesamiento masivo de datos. Resulta poco factible la utilización de una estructura en aplicaciones reales si ésta no considera su implementación en entornos paralelos. En la paralelización de estructuras métricas deben

considerarse, entre otros aspectos, la distribución adecuada de la base de datos sobre el entorno, por ejemplo en un cluster de PCs; la paralelización de los métodos de búsqueda; eficiencia en la comunicación entre los procesadores; etc.

Desde el punto de vista de las GPU, esencialmente se han abordado soluciones para las consultas kNN sin utilizar estructuras de datos, es decir, se utilizan para paralelizar búsquedas exhaustivas (fuerza bruta) [13, 14, 15]. Todos estos trabajos abordan el problema desde el supuesto de que los elementos de la base de datos poseen una gran dimensión (con una gran dimensión intrínseca), lo cual implica poder descartar muy pocos objetos usando desigualdad triangular, y en donde la indexación es ineficiente frente a la búsqueda secuencial.

La paralelización de estructuras métricas sobre GPU y en clusters de multicores son áreas de investigación poco exploradas. En este contexto, se conocen dos grupos de investigación que están desarrollando trabajos en torno a estructuras métricas sobre plataformas basadas en GPU. El primero de estos grupos, de la Universidad Complutense de Madrid, ha enfocado sus estudios a dos estructura métricas, la *Lista de Clusters* y *SSS-Index*, y ha presentado diversos trabajos en propuestas para kNN y consultas por rango [15, 16]. En el segundo grupo, de la Universidad de Castilla La Mancha, participa el profesor Roberto Uribe-Paredes, las líneas de investigación abordadas por este grupo están orientadas a desarrollar y potenciar, sobre un ambiente híbrido, la estructura genérica presentada en la sección 1.2 [7, 10, 17, 18].

## 2. LINEAS DE INVESTIGACION y DESARROLLO

De acuerdo a los antecedentes disponibles y los trabajos previos realizados, se han planteado las siguientes líneas de investigación:

- Implementación de distintas soluciones sobre espacios disímiles, tales como histogramas de colores, bases de datos de palabras, vectores de coordenadas, vectores de gauss, imágenes NASA y otros.

- Forma óptima de distribución de los datos, tanto de la base de datos como de las consultas, en ambientes de GPU y/o Multi-GPU, en las memorias administradas por las CPUs y las memorias dedicadas de las GPU.

- Escalabilidad de las implementaciones de acuerdo al tamaño de la Base de Datos, y consecuentemente el almacenamiento y procesamiento de BD y consultas en memoria secundaria.

### 3. RESULTADOS OBTENIDOS / ESPERADOS

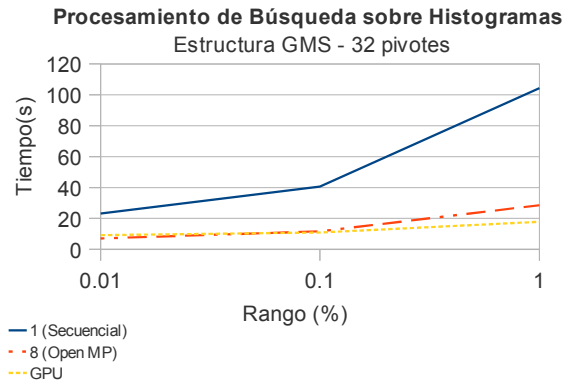
Basados en las experiencias de los integrantes del grupo de investigación, se realizarán experimentos para determinar espacios y dimensiones donde es realmente conveniente utilizar estructuras métricas sobre plataformas basadas en GPU. Se espera confirmar o refutar la hipótesis de la conveniencia de utilización de fuerza bruta o a lo sumo estructuras genéricas basadas en pivotes en ambientes GPU, extendiendo las pruebas de laboratorio a un amplio conjunto de ambientes y espacios.

Como evaluación experimental, actualmente se está trabajando con espacios métricos en un entorno *multicore* y un entorno GPU utilizando la estructura GMS basada en pivotes.

Para los casos de estudio se han considerado dos conjuntos de datos: el primero es un diccionario con palabras en español con 86.061 elementos. La función de distancia utilizada es la *distancia de edición*, definida como el mínimo número de inserciones, eliminaciones o sustituciones de caracteres necesarios para que una palabra sea convertida en otra. Para este espacio, los rangos de búsqueda son discretos y con rangos entre  $r = 1$  y  $r = 4$ . El segundo espacio corresponde a un conjunto de 112.682 histogramas de colores (vectores de dimensión 112) de una base de datos de imágenes. Para este espacio se seleccionó la distancia euclidiana. Los radios utilizados son aquellos que permiten recuperar el 0.01%, 0.1% y 1% de la base de datos. Para ambos espacios se construye la estructura con el 90% de los datos, dejando el restante 10% para consultas.

Se han elegido estas condiciones para la realización de las pruebas ya que son las típicamente usadas para este tipo de experimentos. Para cada rutina, se ejecuta la aplicación 4 veces y se obtiene un promedio.

La plataforma utilizada corresponde a un Intel® Core™ i7-2600 CPU @3.40GHz de 4 núcleos y soporte a Hyper-Threading, 08GB de memoria principal y dos tarjetas Nvidia EVGA DDR5 de 384 cores CUDA y 1 GB de memoria global cada, usando CUDA SDK v3.2 [14]. La codificación se ha realizado utilizando el lenguaje C (gcc 4.3.4) y librerías OpenMP. La Figura 2 nos permite tener una visión inicial respecto de la utilización de una plataforma multicore y una plataforma GPU en comparación con su procesamiento secuencial, sobre el espacio de histogramas mencionado anteriormente.



**Figura 2. Tiempos de ejecución absolutos para las versiones secuencial (1 core), multi-core de 8 núcleos (4 núcleos con *Hyper-Threading*) y para la versión GPU. Histogramas de colores de 112.682 vectores.**

Como era de esperar, puede apreciarse en la Figura 2 la enorme diferencia al utilizar una plataforma *multicore* o *manycore* sobre una secuencial, observándose una disminución considerable del tiempo de procesamiento. Asimismo, se puede observar un rendimiento levemente mejor de la plataforma GPU a medida en que se aumenta el radio de búsqueda. Similar comportamiento se ha observado en los *tests* realizados sobre la base de datos de palabras. Ello en virtud de que la cantidad de evaluaciones sobre objetos candidatos aumenta notoriamente al aumentar el rango o radio de búsqueda. La cantidad de hilos disponibles para resolver las consultas compensa el tiempo aplicado en la alocaión de los recursos en la memoria de la GPU.

Por otra parte, debido al avance tecnológico y la reducción de los costos, se pretende integrar en un solo equipo múltiples CPU y múltiples GPU [9, 16, 17], de manera que resulte posible extender el paralelismo a todos estos ambientes en forma simultánea. Ello crea la posibilidad de evaluar el rendimiento de las distintas estructuras para ambientes de múltiples CPU / GPU. Si bien se han realizado pruebas de laboratorio falta aún definir la mejor estrategia de distribución de datos tanto para el procesamiento como en el almacenamiento en memoria volátil de la CPU / GPU.

Si bien hasta el momento las pruebas de laboratorio se han realizado con bases de datos de importante tamaño, queda aún por validar las conclusiones obtenidas en cuanto a la escalabilidad de las mismas. Las pruebas de escalabilidad llevarán necesariamente a investigar sobre la distribución de los datos, consultas y estructuras métricas entre la memoria principal y la secundaria, debido a que en algún momento no será posible almacenar la base de datos completa en memoria volátil.

#### 4. FORMACION DE RECURSOS HUMANOS

El investigador Roberto Uribe-Paredes, realizó su tesis de Magister en el área de espacios métricos. Actualmente desarrolla su tesis de doctorado en la misma área desarrollando estructuras sobre entornos basados en GPU, en la Universidad de Castilla La Mancha, España, donde espera rendir su examen el año corriente. Este investigador ha apoyado al grupo en la fase inicial realizando cursos a distancia sobre espacios métricos y sobre tecnologías basadas en GPU. Además, a través de este investigador se están desarrollando lazos de trabajo con los grupos de investigación de la Universidad de Magallanes, Chile y de Castilla La Mancha, España.

Un investigador (Dos Santos) se encuentra cursando la “Especialización en Management Tecnológico” de la Universidad Nacional de la Patagonia Austral, y actualmente desarrolla su tesis titulada “Proyecto de Educación Tecnológica en el Programa de Educación a Distancia de la UNPA”.

Un investigador (Salvador) ha obtenido recientemente su título de doctorado en Ingeniería, mención procesamiento de señales e imágenes en la Universidad tecnológica Nacional, regional Buenos Aires, con la tesis titulada “Estudio del comportamiento de la capa de ozono y la radiación UV en la Patagonia Austral y su proyección a la comunidad”.

#### 5. AGRADECIMIENTOS

El presente trabajo fue parcialmente financiado por la Universidad Nacional de la Patagonia Austral, Santa Cruz, Argentina.

#### 6. BIBLIOGRAFIA

- [1] Edgar Chávez, Gonzalo Navarro, Ricardo Baeza-Yates, and José L. Marroquín. “Searching in metric spaces”. In *ACM Computing Surveys*, 2001, pp. 33(3):273–321.
- [2] L. Micó, J. Oncina, and E. Vidal. “A new version of the nearest-neighbor approximating and eliminating search (aes) with linear preprocessing-time and memory requirements”. *Pattern Recognition Letters*, vol. 15, pp. 9–17, 1994.
- [3] R. Baeza-Yates, W. Cunto, U. Manber, and S. Wu. “Proximity matching using fixedqueries trees”. In *5th Combinatorial Pattern Matching (CPM'94)*, 1994, LNCS 807, pp. 198–212.
- [4] E. Chávez, J. Marroquín, and R. Baeza-Yates. “Spaghettis: An array based algorithm for similarity queries in metric spaces”. In *6th International Symposium on String Processing and Information Retrieval (SPIRE'99)*. IEEE CS Press, 1999, pp. 38–46.
- [5] E. Chávez, J. Marroquín, and G. Navarro. “Fixed queries array: A fast and economical data structure for proximity searching”. *Multimedia Tools and Applications*, vol. 14, no. 2, pp. 113–135, 2001.
- [6] Oscar Pedreira and Nieves R. Brisaboa. “Spatial selection of sparse pivots for similarity search in metric spaces”. In *33rd Conference on Current Trends in Theory and Practice of Computer Science (SOFSEM 2007)*, Harrachov, Czech Republic, 2007, vol. 4362 of LNCS, pp. 434–445, Springer.
- [7] R. Uribe-Paredes, P. Valero-Lara, E. Arias, J.L. Sanchez, and D. Cazorla. “Similarity search implementations for multi-core and many-core processors”. In *International Conference on High Performance Computing and Simulation (HPCS)*, 2011, pp. 656–663.
- [8] Wu-Feng and Dinesh Manocha. “High-performance computing using accelerators,” *Parallel Computing*, vol. 33, pp. 645–647, 2007.
- [9] Roberto Uribe-Paredes, Enrique Arias, Diego Cazorla, José Luis Sánchez. “Una estructura Métrica Genérica para Búsquedas por Rango sobre una Plataforma Multi-GPU”. *XVII Jornadas de Ingeniería del Software y Bases de Datos (JISBD2012)*. Sept. 2012, Almería, España.
- [10] Roberto Uribe-Paredes, Diego Cazorla, José L. Sánchez, and Enrique Arias. “A comparative study of different metric structures: Thinking on gpu implementations”. In *International Conference of Computational Statistics and Data Engineering (ICCSDE'12)*, London, England, July 2012.
- [11] NVIDIA CUDA C Programming Guide, Version 4.0, NVIDIA, 2011, <http://developer.nvidia.com/object/gpucomputing.html>
- [12] Ananth Grama, George Karypis, Vipin Kumar, and Anshul Gupta. *Introduction to Parallel Computing (2nd Edition)*. Addison Wesley, 2 edition, 2003.
- [13] Quansheng Kuang and Lei Zhao. “A practical GPU based kNN algorithm”. *International Symposium on Computer Science and Computational Technology (ISCSCT)*, pp. 151–155, 2009.
- [14] Vincent Garcia, Eric Debreuve, and Michel Barlaud. “Fast k nearest neighbor search using GPU”. *Computer Vision and Pattern Recognition Workshop*, vol. 0, pp. 1–6, 2008.
- [15] R.J. Barrientos, J.I. Gómez, C. Tenllado, M. Prieto, and M. Marin. “kNN query processing in metric spaces using gpus”. in *17th International European Conference on Parallel and Distributed Computing (Euro-Par 2011)*, Bordeaux, France, 2011, vol. 6852 of LNCS, pp. 380–392, Springer.
- [16] R.J. Barrientos, J.I. Gómez, C. Tenllado, M. Prieto, and M. Marin. “Range query processing in a multi-GPU environment”. In *10th IEEE International Symposium on Parallel and Distributed Processing with Applications (ISPA 2012)*, Madrid, Spain, July 2012.
- [17] Roberto Uribe-Paredes, Enrique Arias, José L. Sánchez, and Diego Cazorla. “Improving the Performance for the Range Search on Metric Spaces using a Multi-GPU Platform”. To appear: *23rd International Conference on Database and Expert Systems Applications (DEXA 2012)*. Vienna, Austria, Sept. 2012.
- [18] Roberto Uribe-Paredes, Diego Cazorla, Enrique Arias and José L. Sánchez. “Acelerando la Búsqueda por Rango con un Sistema Híbrido de Memoria Compartida”. To appear: *Actas XXIII Jornadas de Paralelismo (JP2012)*. Sept. 2012, El España.
- [19] Roberto Uribe-Paredes, Pedro Valero-Lara, Enrique Arias, José L. Sánchez and Diego Cazorla. “Similarity search implementations for multi-core and many-core processors”. In: *2011 International Conference on High Performance Computing and Simulation (HPCS)*, pp. 656–663 (July 2011). Istanbul, Turkey.

# Aspectos energéticos de sistemas de computación

Javier Balladini, Belén Casanova, Marina Morán, Federico Uribe, Armando De Giusti<sup>1</sup>

Facultad de Informática, Universidad Nacional del Comahue

Buenos Aires 1400, Neuquén Capital, Tel. 0299 - 4490300

{javier.balladini, mb.casanova.retamal, fede.uribe83}@gmail.com

<sup>1</sup> Instituto de Investigación en Informática LIDI, Universidad Nacional de La Plata

50 y 120 - 2do Piso, La Plata, Tel/Fax (54-221) 4227707

degjusti@lidi.info.unlp.edu.ar

## Resumen

En los últimos años, las métricas de potencia y consumo energético han dirigido el diseño del hardware de la computación, incluyendo CPUs multi-core y aceleradoras como por ejemplo las GPUs. El diseño con consciencia de la potencia y energía también ha dominado las grandes instalaciones de supercomputadoras y sistemas de cloud computing. La alta demanda energética tiene serias consecuencias financieras, medioambientales, y en muchos casos también sociales. Además de cambiar el diseño tradicional del hardware, enfocado exclusivamente a aumentar la velocidad de procesamiento, es necesario desarrollar software para gestionar los recursos hardware y permitir adaptarse a determinados niveles de potencia y consumo energético.

Esta situación nos ha motivado a realizar una colaboración entre tres universidades para estudiar diferentes temas relacionados a la computación ecológica. Nuestro trabajo se centra en el diseño del software y no del hardware. Las líneas de interés actuales son: caracterización energética, predicción de potencia, diseño de algoritmos de escalado dinámico de frecuencia y tensión, y estrategias energéticas en cloud computing.

**Palabras claves:** computación ecológica, eficiencia energética, computación de altas prestaciones

## 1. Contexto

La investigación aquí presentada surge como una colaboración entre la Facultad de Informática de la Universidad Nacional del Comahue, el Instituto de Investigación en Informática LIDI de la Universidad Nacional de La Plata, y el grupo de investigación HPC4EAS (High Performance Computing for Efficient Applications and Si-

mulation) de la Universidad Autónoma de Barcelona (España). La colaboración se centra en el estudio del consumo energético de los sistemas de cómputo paralelo al ejecutar aplicaciones científicas computacionalmente complejas.

## 2. Introducción

De acuerdo con la definición de San Murugesan [9], se define el campo de la computación ecológica como el estudio y práctica del diseño, fabricación, uso y disposición de las computadoras, servidores y subsistemas relacionados como monitores, impresoras, almacenamiento y sistemas de redes y comunicación eficientes y efectivos con un impacto mínimo o nulo en el ambiente. El autor identifica cuatro formas en las cuales cree que los efectos de las computadoras en el ambiente deberían ser considerados: diseño ecológico, fabricación ecológica, uso ecológico y eliminación ecológica.

El *diseño* de los sistemas de cómputo ha tenido, por décadas, el único objetivo de incrementar la velocidad de procesamiento. Podría decirse que el requerimiento de hacer computadoras cada vez más rápidas vino principalmente de la computación de altas prestaciones, cuyo fin es acelerar la ejecución de aplicaciones computacionalmente complejas, normalmente científicas. Para estas aplicaciones se diseñaban supercomputadoras con la única intención de aumentar la cantidad de operaciones de coma flotante por segundo (FLOPS, *F*loating-*p*oint *O*perations per *S*econd). Esto se ve reflejado en la lista del TOP500 [1], que utiliza la métrica FLOPS para determinar el orden de clasificación de las supercomputadoras. Sólo importaban las prestaciones y, principalmente para el dueño de la supercomputadora, la relación precio/prestaciones. Pero esto ha cambiado. En la actualidad, el diseño de los sistemas de cómputo tiene como principal restricción el consumo ener-



gético, cambio que se ha producido primordialmente por la influencia de los siguientes factores. Las supercomputadoras se volvieron tan grandes que llegan a consumir tanta electricidad como una ciudad<sup>1</sup>, no solo por su consumo específico sino también por la energía requerida para refrigerar el sistema y así asegurar su correcto funcionamiento, que se traduce en un enorme **costo económico** (y, posiblemente, ecológico y social<sup>2</sup>); se estima que cada 4 años el gasto en energía es equivalente al costo de adquisición de la supercomputadora. Otro factor tiene su origen en la **revolución de los dispositivos móviles**, de quienes cada vez se espera que realicen más procesamiento con iguales capacidades energéticas que obtienen de sus pequeñas baterías. El factor restante es conocido como **power wall**, y se refiere a una limitante física impuesta por la capacidad de refrigeración que puede alcanzarse utilizando técnicas convencionales (otras técnicas implicarían un aumento excesivo del costo).

Para estudiar la evolución de los procesadores, y observar el cambio que ha significado la consideración energética en el diseño de los mismos, utilizaremos la figura 1. Hasta el año 2005 no solo se ha doblado el número de transistores cada 18-24 meses para incrementar las prestaciones de los procesadores (ley de Moore), sino que también se ha doblado el consumo energético[7]. A partir del 2005 el número de transistores continuó incrementándose, pero el consumo energético se mantuvo constante. Este logro está basado en el desarrollo de procesadores con múltiples cores (nucleos) que trabajan a una menor frecuencia de reloj (menor cantidad de ciclos de reloj por unidad de tiempo). A modo de ejemplo, en el cuadro 1 se muestra una tabla comparativa de la línea de procesadores de Intel para servidores "Sandy Bridge-EP" (32 nm) Efficient Performance. Se puede observar la influencia de la velocidad del procesador en la potencia energética demandada, en tal caso, el Xeon E5-4650 tiene una frecuencia de reloj un 12,5 % mayor a la del Xeon E5-4640 pero demanda un 36,8 % más de potencia eléctrica<sup>3</sup>.

Enfocándonos en el dominio de nuestro inte-

<sup>1</sup>En 2002, el Dr. Eric Schmidt, CEO de Google, dijo "lo que más importa a los diseñadores de computadoras de Google no es la velocidad sino el consumo energético, porque los centros de datos pueden consumir tanta electricidad como una ciudad" [8].

<sup>2</sup>La falta de explotación de las energías renovables y limpias hacen que la producción energética afecte en lo ecológico y social. La mayor cantidad (y más grandes) de las supercomputadoras del mundo están en EEUU, donde la mitad de la energía eléctrica en éste país se produce con carbón [2], impactando fuertemente en el medio ambiente, y la salud y riesgo de vida de las personas, por causa de la extracción del mineral y la contaminación por combustión

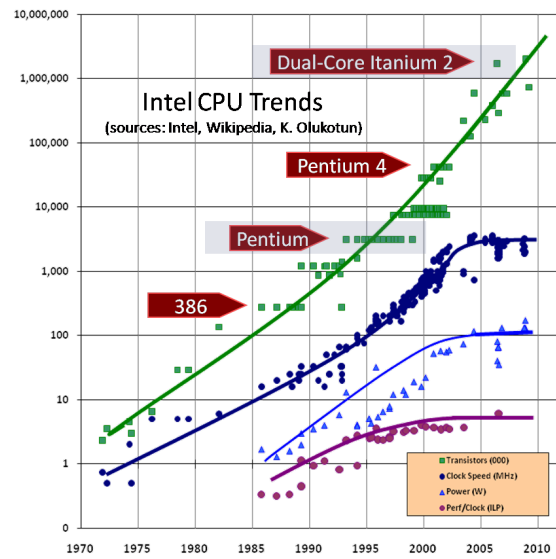


Figura 1: Evolución de los procesadores. Fuente: Intel

Model Number	Cores	Frequency	L2 cache	L3 cache	TDP	Socket	I/O bus
Xeon E5-4640	8	2.4 GHz	8 × 256 KB	20 MB	95 W	LGA 2011	2 × 8.0 GT/s QPI
Xeon E5-4650	8	2.7 GHz	8 × 256 KB	20 MB	130 W	LGA 2011	2 × 8.0 GT/s QPI

Cuadro 1: Características de procesadores Sandy Bridge-EP

res, la computación de altas prestaciones, podríamos decir que el gran hito que determinó la nueva era de la supercomputación ecológica data de 2007. En ese año, una nueva entidad denominada Green500, surgida como alternativa o complemento del TOP500, publica su primera lista que clasifica a las supercomputadoras de mayor eficiencia energética del mundo. Esta clasificación considera el rendimiento por W (FLOPS/W) de cada supercomputadora al ejecutar un cierto benchmark, donde el *diseño* y el *uso ecológico* son los factores que tienen una relación directa con el resultado alcanzado.

### 3. Línea de investigación

Nuestro trabajo no pretende desarrollar nuevas tecnologías hardware que tengan una mayor eficiencia energética (*diseño ecológico*), sino gestionar mediante software el hardware existente para reducir el consumo energético (*uso ecológico*). Nuestra investigación se enfoca en las siguientes temáticas energéticas aplicadas a sistemas de

del carbón, entre otros.

<sup>3</sup>La métrica utilizada en la tabla comparativa se denomina Thermal Design Power (TDP), que representa la máxima potencia que un sistema de refrigeración requiere disipar al ejecutar una "aplicación real".

computación de altas prestaciones:

1. Caracterización energética de los sistemas. La caracterización energética puede ser utilizada para comparar sistemas, para conocer el comportamiento energético del sistema bajo estudio y así diseñar aplicaciones con conocimiento de la energía, y ser un soporte para la construcción de modelos de predicción de potencia. Nos interesan plataformas basadas en CPUs (*Central Processing Unit*) y GPUs (*Graphics Processing Unit*).
2. Predicción de potencia. La predicción de la potencia es necesaria debido a que, en general, los sistemas no disponen de dispositivos internos de medición, en tiempo real, de la potencia demandada<sup>4</sup>; dispositivos externos son imprácticos para sistemas con gran número de nodos y procesadores por nodo. El objetivo de la predicción de potencia es permitir el desarrollo de algoritmos DVFS (ver siguiente punto) que no podrían ser propuestos sin esta información.
3. Diseño de algoritmos de escalado dinámico de frecuencia y tensión (DVFS, *Dynamic Voltage and Frequency Scaling*). La reducción de la tensión suministrada a un circuito reduce el consumo energético, sin embargo, incrementa el retardo de las compuertas lógicas que fuerza a reducir la frecuencia de reloj para que el circuito continúe trabajando correctamente. Basándose en este concepto, los algoritmos DVFS intentan reducir el consumo energético realizando cambios dinámicos de la frecuencia de reloj de los cores. Nos interesan diversos tipos de algoritmos:
  - a) Algoritmos de mejora de la eficiencia energética para sistemas de memoria compartida.
  - b) Algoritmos para aplicaciones SPMD (*Single Program, Multiple Data*) en *clusters* con procesadores multi-core.
  - c) Otros algoritmos: relación energía-rendimiento, limitación de potencia.
4. Estrategias energéticas en cloud computing. El cloud computing es un paradigma, relativamente nuevo, de cómputo distribuido que se presenta como una evolución natural del concepto de *clusters* y *grids*. Proporciona grandes conjuntos de recursos virtuales

<sup>4</sup>Hay excepciones, por ejemplo, la arquitectura Sandy Bridge de Intel cuenta con contadores de rendimiento que entregan mediciones de potencia.

(hardware, plataformas de desarrollo, almacenamiento y/o aplicaciones), fácilmente accesibles. Considerando sus grandes dimensiones, y la tendencia natural a que la plataforma escale para ampliar los servicios provistos, nos interesa estudiar las posibles alternativas que permitan reducir su alta demanda energética.

## 4. Resultados y objetivos

En referencia a la temática de caracterización energética de los sistemas, se ha estudiado la influencia de los modelos de programación paralela (OpenMP y MPI) y escalado de frecuencias de CPUs en sistemas de computación de altas prestaciones de memoria compartida [4]. También hemos estudiado los factores influyentes en el consumo energético de los sistemas de cómputo de altas prestaciones basados en CPUs y GPUs [5]. Posteriormente, hemos propuesto una metodología para caracterizar la potencia de sistemas de computación de altas prestaciones de memoria compartida [3]. La metodología involucra la búsqueda de factores influyentes en la potencia del sistema, realizando un análisis de sensibilidad de las propiedades de la carga de trabajo y parámetros del sistema en el comportamiento de potencia. La carga de trabajo considera aspectos de cómputo y comunicaciones de las aplicaciones. Esta metodología es similar a trabajos previos, pero nosotros proponemos un esquema de mayor profundidad que puede ayudar a mejorar la caracterización del sistema. También realizamos estudios sobre el impacto del sistema de Entrada/Salida en la eficiencia energética, que publicamos en [10].

Hemos propuesto un algoritmo de mejora de la eficiencia energética para sistemas de memoria compartida [6], basado en la aceleración de los relojes de los cores en momentos de bajo paralelismo. Actualmente, estamos realizando el testing de una implementación de ese algoritmo con la característica de que ejecuta de forma transparente al usuario y aplicación (no requiere instrumentación del código fuente). Asimismo, se está avanzando en la búsqueda de nuevas oportunidades orientadas a mejorar aún más la eficiencia energética.

Se está trabajando en la construcción de modelos para predicción de potencia de sistemas basados en CPUs que luego nos permitirá ampliar el espectro de propuestas de algoritmos DVFS. Estos son los casos de algoritmos cuyos objetivos podrían estar orientados a obtener la mejor relación energía-rendimiento, limitar la potencia demandada (útil para sistemas alimentados por

fuentes energéticas limitadas y de potencia variable -planta solar o eólica), entre otros.

Recientemente, se ha comenzado a trabajar en el desarrollo de un algoritmo DVFS para aplicaciones SPMD en clusters con procesadores multi-core. En cuanto al diseño de estrategias energéticas para implantar en entornos de cloud computing, es una temática de gran interés que estamos comenzando a estudiar.

## 5. Formación de recursos humanos

Los estudios aquí expuestos tienen como objetivo formar recursos humanos a nivel de grado y postgrado. En la Universidad Nacional del Comahue, hemos presentado una tesis de Licenciatura en Ciencias de la Computación en el tema de "Computación de Altas Prestaciones Ecológica con GPUs", y se está finalizando otra en el tema "Algoritmos de mejora de la eficiencia energética para sistemas de memoria compartida".

Es de especial interés promover, a docentes-investigadores y alumnos que trabajan en esta línea de investigación, la realización de estudios de maestría y doctorado en las instituciones participantes de la colaboración, la Universidad Nacional de La Plata y la Universidad Autónoma de Barcelona.

## Referencias

- [1] Sitio web del top500: <http://www.top500.org/> (accedido en octubre de 2010).
- [2] U.s. geological survey, <http://energy.usgs.gov/coal.html> (accedido en octubre de 2010).
- [3] Javier Balladini, Enzo Rucci, Armando Eduardo De Giusti, Marcelo Naiouf, Remo Suppi, Dolores Rexachs del Rosario, and Emilio Luque Fadón. Power characterisation of shared-memory hpc systems. In *XVIII Congreso Argentino de Ciencias de la Computación*, 2012.
- [4] Javier Balladini, Remo Suppi, Dolores Rexachs, and Emilio Luque. Impact of parallel programming models and cpus clock frequency on energy consumption of hpc systems. *Computer Systems and Applications, ACS/IEEE International Conference on*, 0:16–21, 2011.
- [5] Javier Balladini, Federico Uribe, Remo Suppi, Dolores Rexachs, and Emilio Luque. Factores influyentes en el consumo energético de los sistemas de computación de altas prestaciones basados en cpus y gpus. *XVII Congreso Argentino de Ciencias de la Computación (CACTC 2011)*, pages 271–281, 2011.
- [6] Belén Casanova, Javier Balladini, Armando Eduardo De Giusti, Remo Suppi, Dolores Rexachs del Rosario, and Emilio Luque Fadón. Mejora de la eficiencia energética en sistemas de computación de altas prestaciones. In *XVIII Congreso Argentino de Ciencias de la Computación*, 2012.
- [7] Wu-Chun Feng. The importance of being low power in high-performance computing. *Cyberinfrastructure Technology Watch Quarterly (CT-Watch Quarterly)*, 1(3):11–20, August 2005.
- [8] John Markoff and Steve Lohr. Intel's huge bet turns iffy. *New York Times*, September 29, 2002.
- [9] San Murugesan. Harnessing green it: Principles and practices. *IT Professional*, 10:24–33, January 2008.
- [10] Javier Panadero, Sandra Méndez, Dolores Rexachs, Javier Balladini, Ismael Rodríguez, Adrián Pousa, Remo Suppi, and Emilio Luque. El impacto del sistema de entrada/salida en la eficiencia energética. In *XVII Congreso Argentino de Ciencias de la Computación*, 2011.

# Elaboración de Estrategias Paralelas para Búsquedas por Similitud en Espacios Métricos

Norma Beatriz Perez, Mario Berón

Facultad de Ciencias Físico Matemáticas y Naturales - Universidad Nacional de San Luis  
Ejército de los Andes, 950 CP: D5700HHW - San Luis - Argentina  
email: {nbperez, mberon}@unsl.edu.ar

Fernando Magno Quintão Pereira

Departamento de Ciência da Computação - Universidade Federal de Minas Gerais  
Av. Antônio Carlos, 6627 - Prédio do ICEx - Pampulha - CEP: 31270-010  
Belo Horizonte Minas Gerais - Brasil  
email: fernando@dcc.ufmg.br

## Resumen

Realizar una búsqueda secuencial de elementos específicos sobre grandes volúmenes de información no es una aproximación apropiada. Esto se debe a que los tiempos de respuesta obtenidos no se adecuan a la exigencia de los usuarios. Una forma de abordar la problemática previamente mencionada consiste en aprovechar los avances de la tecnología, en lo que a software y hardware se refiere, para elaborar e implementar nuevos algoritmos que realicen búsquedas más eficientes, como lo son las búsquedas por similitud paralelas. Llevar a cabo la tarea antes mencionada no es fácil debido a que implica: i) El estudio de diferentes estructuras de datos que puedan ser paralelizables, ii) La selección de un modelo de computación paralela que posea una infraestructura que facilite la tarea del programador y iii) La elaboración de algoritmos que utilicen eficientemente las estructura de datos elegidas y que aprovechen al máximo las capacidades del modelo de computación paralela seleccionado y su infraestructura asociada. En este artículo se describe una línea de investigación que aborda las temáticas previamente mencionadas y cuyo principal objetivo es: La elaboración de algoritmos de búsqueda por similitud paralelos, basados en estructuras de indexación eficientes, cuya eficiencia escala en forma lineal respecto del número de procesadores disponibles en redes de tamaño moderado.

**Palabras clave:** Watershed, Filter-Stream, Modelos de Computación paralela, D-Index, Búsquedas por Similitud.

## 1. Contexto

La línea de investigación descrita en este trabajo se encuentra enmarcada en el contexto de dos proyectos de investigación. El primero, denominado "*Ingeniería del Software: Aspectos de Alta Complejidad Sensibilidad en el Ejercicio de la Profesión de Ingeniero de Software*", se desarrolla en la Universidad Nacional de San Luis. Dicho proyecto, es reconocido por el programa de incentivos y es la continuación de diferentes proyectos de investigación de gran éxito a nivel nacional e internacional. El segundo, llamado "*Proyecto de co-tutela CAFP-BA 004/08*" se lleva a cabo en la Universidad Nacional de San Luis - Argentina y en la Universidade Federal de Minas Gerais, Belo Horizonte - Brasil. Este proyecto fue aprobado por la Secretaría de Políticas Universitarias (SPU) dependiente del Ministerio de Ciencia, Tecnología e Innovación Productiva de la Nación (MinCyT) [1]. Ambos entes soportan económicamente la realización de diferentes misiones de investigación desde Argentina a Brasil y viceversa.

## 2. Introducción

En los últimos años, la recuperación de la información sobre grandes conjuntos de datos tales como: registros de datos científicos, aplicaciones multimedia, bioinformática, etc., se han vuelto un problema de gran interés. Los tipos de datos que componen estas aplicaciones son, generalmente, complejos por lo que el proceso de manipulación de ellos no es una tarea simple. En muchos de esos casos las búsquedas exactas, que son la manera típica de buscar en bases de datos tradicionales, dejan de ser aplicables. Esto ha motivado el surgimiento de diversas



técnicas y métodos que permiten construir estructuras de indexación. Estas estructuras permiten recuperar la información de manera más eficiente y en menos tiempo. Este tipo de métodos se ubican en una categoría denominada *búsquedas por similitud* [2]. En esta categoría se ubican todos aquellos métodos que consisten en buscar aquellos objetos en un conjunto de datos que sean similares a un objeto de consulta dado.

Las búsquedas por similitud pueden ser descritas a través de la definición formal de los *Espacios Métricos* [3, 4]. Un Espacio Métrico está compuesto por un par ordenado formado por un universo de objetos de un conjunto finito y de una función de distancia que satisface las propiedades de: *positividad*, *simetría* y *desigualdad triangular*. En donde, los objetos del universo serán comparados directamente utilizando la función de distancia, que indica el grado de *similitud* entre dos objetos.

Se ha comprobado que la implementación de algoritmos secuenciales de búsqueda por similitud no es apropiada debido a que los mismos no escalan bien para conjuntos de datos grandes como por ejemplo: registros de bases de datos médicas, conjuntos de datos disponibilizados en redes sociales *on-line* y servicios Web. Por esta razón, las investigaciones recientes se han centrado en tres aspectos fundamentales:

1. La búsqueda de estructuras de datos que sean eficientes e inherentemente paralelas.
2. La selección un modelo de computación paralela simple de entender y fácil de utilizar.
3. La elaboración de algoritmos paralelos de búsqueda por similitud eficientes.

Las tres temáticas antes mencionadas son el foco de estudio de la línea de investigación descrita en el presente trabajo.

El artículo está organizado como sigue. En la sección 3 se describe la línea de investigación. La sección 4 presenta los resultados obtenidos a partir del trabajo realizado por el equipo de investigación. Finalmente la sección 5 explica sucintamente los resultados obtenidos y esperados con respecto a la formación de recursos humanos.

### 3. Línea de investigación y desarrollo

La línea de investigación descrita en este artículo consta de tres temáticas principales las cuales se describen en las siguientes subsecciones.

#### 3.1. Paralelización de Estructuras de Datos para Búsquedas por Similitud

En el estado del arte se pudo observar la existencia de un amplio conjunto de estructuras de datos [5] especializadas que permiten resolver búsquedas por similitud en los espacios métricos. Este conjunto se clasifica en dos grupos basados en:

1. Clustering o Particiones Compactas: consiste en particionar el conjunto de datos en áreas. Cada área tiene un centro, y se intentan descartar áreas completas sólo comparando la consulta con el centro del área. Son ejemplos de este grupo las siguientes estructuras: M-trees (MT) [6], Spatial Approximation Trees (SAT) [3].
2. Pivotes: almacenan las distancias que han sido precalculadas de cada objeto en el conjunto de datos a un conjunto de pivotes. Luego esas distancias se utilizan durante la búsqueda para descartar objetos del conjunto resultado. A este grupo pertenecen estructuras tales como: Approximating Eliminating Search Algorithm (AESA) [7], Fixed Queries trees (FQT) [8].

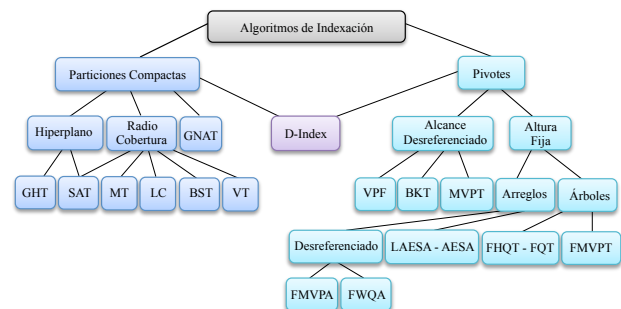


Figura 1: Taxonomía de los Algoritmos.

La Figura 1 ilustra una clasificación de las estructuras de datos para búsquedas por similitud en los espacios métricos existentes en la literatura de acuerdo a sus características generales.

El D-Index secuencial [9] es considerando uno de los métodos más rápidos de acceso disponible para resolver búsquedas por similitud [10] debido a que posee las siguientes características:

- Representa uno de los avances más recientes en el escenario de los espacios métricos
- Une la familia de algoritmos basadas en clustering y pivotes [5].
- Se basa en: i) La definición de una función que divide a los objetos, y ii) El establecimiento de una jerarquía de buckets que almacenan estos objetos.

- Resulta adecuado en espacios de alta dimensionalidad, es decir, espacios en donde el problema de búsqueda por similitud es inherentemente difícil [10].
- Las características estructurales lo hacen adecuado para ser paralelizado.
- La performance depende de varios parámetros, y la configuración óptima de estos sigue siendo un problema abierto [11].

Por todas las características mencionadas previamente, el D-Index ha sido seleccionado, en primera instancia, como estructura soporte para los diferentes tipos de algoritmos que han sido propuestos por los integrantes del grupo de investigación.

### 3.2. Modelos de Computación Paralela

No sólo se requiere seleccionar la estructura de datos apropiada para la implementación de búsquedas por similitud sino que también es necesario determinar el modelo y la infraestructura de computación paralela adecuada a ser utilizada como soporte en la implementación de los algoritmos antes mencionados. Existe un gran número de modelos [12, 13] e infraestructuras [14, 15]. Estos han excedido la cantidad de arquitecturas diferentes y por lo general, la mayoría de ellos, son inadecuados o ineficientes cuando se desea paralelizar un índice.

Generalmente, los programadores al paralelizar sus índices deben tener conocimientos detallados de las componentes del modelo a utilizar. Esta particularidad es imprescindible para evitar, en los algoritmos paralelos, incrementar considerablemente el número de líneas de código y su complejidad. Por lo tanto, el uso y diseño de modelos e infraestructuras no es una tarea fácil e involucra un proceso altamente creativo.

Luego del estudio de varios modelos de computación paralela y sus infraestructuras asociadas se pudo observar que Watershed [16], una infraestructura asociada al modelo *Filter-Stream* [17], ofrece un marco de trabajo simple y unificado que posee las siguientes características: i) Permite el diseño y programación de aplicaciones *on-line* y *off-line* distribuidas, y ii) Posee un mayor poder cómputo al posibilitar resolver consultas en tiempos aceptables.

Una aplicación en Watershed está compuesta por una cadena de elementos (filtros) de procesamiento. Los filtros se comunican por medio de flujos continuos de datos. Cada filtro representa una etapa del procesamiento realizado por la aplicación. Un flujo

de datos está compuesto por resultados intermedios de alguna etapa de procesamiento que son las entradas para etapas posteriores. La Figura 2 ilustra el procesamiento que realiza Watershed a través de filtros que intervienen en la aplicación.

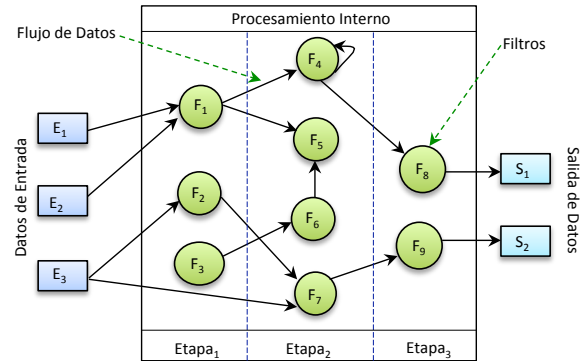


Figura 2: Vista del Programador de la Infraestructura Watershed para una Aplicación Arbitraria.

El interés de trabajar con Watershed como soporte para la implementación de los algoritmos paralelos se basa en las siguientes características:

- Permite que las aplicaciones puedan ejecutarse en modo flujo (donde no existe el concepto de finalización) o modo por lotes (donde los filtros finalizan tan pronto como sea procesado su último mensaje).
- Provee un mecanismo de adición y remoción de filtros en tiempo de ejecución.
- Permite la realización de alteraciones en la topología de las aplicaciones en tiempo de ejecución.
- Permite la reutilización de filtros en ejecución.

Como es de suponer, las características mencionadas en los ítems precedentes guiaron al grupo de investigación a seleccionar al modelo *Filter-Stream* y a la infraestructura Watershed como soporte para la implementación de los algoritmos propuestos.

### 3.3. Algoritmos de Búsqueda por Similitud Paralelos

Los algoritmos desarrollados por el equipo de investigación usan el Modelo de Computación Paralela *Filter-Stream*, la infraestructura Watershed y están basados en la estructura de datos D-Index. Dichos algoritmos constan de dos partes claramente definidas. Cada una de ellas se describe a continuación:

**Construcción del Índice:** esta parte se relaciona con la construcción de la estructura de datos que almacena los índices para facilitar la búsqueda, en este caso dicha estructura es el D-Index. Esta tarea puede ser llevada a cabo usando tres aproximaciones, a saber:

- **Algoritmo Naive\_build:** El conjunto de datos se replica en cada uno de los procesadores que forman parte del cluster y que intervienen en la búsqueda. Sobre cada procesador se construye un D-Index, obteniendo como resultado una replica del índice de todo el conjunto de datos en cada procesador.
- **Algoritmo Local\_build:** El conjunto de datos es dividido de manera uniforme en la cantidad de procesadores que forman parte del cluster y que intervienen en la búsqueda. Sobre cada procesador se construye un D-Index utilizando la partición del conjunto de datos correspondiente. Es relevante destacar que cada procesador tiene asociado una partición diferente del conjunto de datos.
- **Algoritmo Global\_build:** Se construye un único D-Index con el conjunto de datos completo. Luego el D-Index es dividido por niveles, donde cada nivel es asignado a un procesador diferente. Se emplea la política *round-robin* para la asignación de los niveles.

**Búsqueda:** esta parte se centra en la estrategia de búsqueda propiamente dicha. La misma puede ser llevada a cabo como sigue:

- **Algoritmo Naive\_search:** Las consultas son enviadas a los procesadores de manera alternada, utilizando una política tipo *round-robin*. De esta manera las consultas se reparten de manera equitativa entre los procesadores. Cada procesador realiza el proceso de búsqueda sobre su D-index local de la manera usual.
- **Algoritmo Local\_search:** Las consultas son enviadas a todos los procesadores, utilizando una política tipo *broadcast*. De esta manera la misma consulta es realizada simultáneamente en una partición diferente del conjunto de datos.
- **Algoritmo Global\_search:** Las consultas son enviadas a todo los procesadores, utilizando una política tipo *broadcast*. De es-

ta manera la misma consulta es enviada a cada procesador. Cada procesador realiza el proceso de búsqueda sobre el nivel del D-index asociado al procesador.

En los casos de los algoritmos de construcción del D-Index descritos, la infraestructura Watershed realiza este proceso a través de un filtro. Donde: i) El dato de entrada del filtro es un conjunto de datos arbitrario, ii) El Procesamiento interno realiza la construcción del D-Index de acuerdo al enfoque seleccionado y iii) Los datos de salida del filtro constan de un mensaje enviado al filtro receptor para indicar que puede continuar con la siguiente etapa de procesamiento de la aplicación.

En los casos de los algoritmos de búsqueda del D-Index descritos, la infraestructura Watershed realiza este proceso a través de un filtro. Donde: i) Los datos de entrada del filtro constan de: a) El D-Index construido de acuerdo a uno de los enfoques descritos anteriormente y b) Un conjunto de datos de consultas, ii) El procesamiento interno del filtro realiza la búsqueda en el D-Index de acuerdo al enfoque seleccionado y iii) Los datos de salida del filtro, es decir los resultados de la búsqueda, son enviados al filtro receptor. Dicho filtro se encarga de mostrar los resultados de acuerdo algún criterio establecido. Finalmente un mensaje es enviado para indicar que puede continuar con la siguiente etapa de procesamiento de la aplicación en caso de existir.

#### 4. Resultados

En este proyecto, se ha logrado: (a) Analizar diferentes estructuras utilizadas para las búsquedas por similitud; (b) Seleccionar al D-Index como estructura de datos óptima para la realización de búsquedas por similitud; (c) Estudiar diferentes Modelos de Computación Paralela y sus infraestructuras asociadas; (d) Explorar el modelo Filter-Stream y la infraestructura Watershed; (e) Implementar algoritmos paralelos basados en D-Index sobre Watershed; (d) Implementar los mismos algoritmos paralelos bajo otro modelo de programación paralela para su posterior análisis y comparación con la versión implementada en Watershed; (d) Realizar experimentos utilizando dos conjuntos de datos para determinar su desempeño.

Durante este proyecto, se espera: (a) Continuar explorando exhaustivamente la infraestructura Watershed; (b) Obtener una mejora sustancial en el desempeño de los algoritmos basados en los enfoques paralelos propuestos; (c) Realizar estudios comparativos con otros modelos de computación

paralela existentes en la literatura; (d) Realizar experimentos que permitan explorar el amplio rango de aplicación de los algoritmos propuestos.

## 5. Formación de Recursos Humanos

Los trabajos elaborados en la presente línea de investigación forman parte del desarrollo de tesis para optar a los grados de Magister en Ingeniería del Software (UNSL) y Mestre em Ciência da Computação (UFMG). Es importante mencionar que tanto el equipo de la Argentina como el de Brasil se encuentran dedicados a la captura de alumnos de grado y posgrado para la realización de estudios de investigación relacionados con las temáticas presentadas en este trabajo. Dichos estudios pretenden fortalecer la relación entre UNSL y la UFMG.

## Referencias

- [1] "Ministerio de ciencia, tecnología e innovación productiva," 2007. [Online]. Available: <http://www.mincyt.gov.ar/>
- [2] P. Zezula, G. Amato, V. Dohnal, and M. Batko, *Similarity Search: The Metric Space Approach*, 1st ed. Springer Publishing Company, Incorporated, 2010.
- [3] G. Navarro, "Searching in metric spaces by spatial approximation," in *SPIRE*. IEEE, 1999, pp. 141–148.
- [4] E. Chávez, G. Navarro, R. A. Baeza-Yates, and J. L. Marroquín, "Searching in metric spaces," *ACM Comput. Surveys.*, vol. 33, no. 3, pp. 273–321, 2001.
- [5] E. Chávez, J. L. Marroquín, and G. Navarro, "Fixed queries array: A fast and economical data structure for proximity searching," *Multimedia Tools Appl.*, vol. 14, no. 2, pp. 113–135, Jun. 2001.
- [6] P. Ciaccia, M. Patella, and P. Zezula, "M-tree: An efficient access method for similarity search in metric spaces," in *VLDB*. ACM, 1997, pp. 426–435.
- [7] E. V. Ruiz, "An algorithm for finding nearest neighbours in (approximately) constant average time," *Pattern Rec.*, vol. 4, pp. 145–157, 1986.
- [8] R. Baeza-Yates, W. Cunto, and S. Wu, "Proximity matching using fixed-queries trees." In Proc. 5th Combinatorial Pattern Matching (CPM'94), LNCS 807, 1994, pp. 198–212.
- [9] V. Dohnal, C. Gennaro, P. Savino, and P. Zezula, "D-Index: Distance searching index for metric data sets," *Multimedia Tools and Applications*, vol. 21, pp. 9–33, 2003.
- [10] A. J. Müller-Molina and T. Shinohara, "On the configuration of the similarity search data structure d-index for high dimensional objects," in *Proceedings of the 2010 international conference on Computational Science and Its Applications - Volume Part III*, ser. ICCSA'10. Berlin, Heidelberg: Springer-Verlag, 2010, pp. 443–457.
- [11] V. Dohnal, C. Gennaro, P. Savino, and P. Zezula, "D-index: Distance searching index for metric data sets," *Multimedia Tools Appl.*, vol. 21, no. 1, pp. 9–33, 2003.
- [12] L. G. Valiant, "A bridging model for parallel computation," *Commun. ACM*, vol. 33, no. 8, pp. 103–111, 1990.
- [13] J. Dean and S. Ghemawat, "Mapreduce: simplified data processing on large clusters," *Commun. ACM*, vol. 51, no. 1, pp. 107–113, Jan. 2008.
- [14] M. Beynon, C. Chang, U. Catalyurek, T. Kurc, A. Sussman, H. Andrade, R. Ferreira, and J. Saltz, "Processing large-scale multi-dimensional data in parallel and distributed environments," *Parallel Comput*, vol. 28, no. 5, pp. 827–859, May 2002.
- [15] D. Fireman, G. Teodoro, A. Cardoso, and R. Ferreira, "A reconfigurable run-time system for filter-stream applications," in *Proceedings of the 2008 20th International Symposium on Computer Architecture and High Performance Computing*, ser. SBAC-PAD '08. Washington, DC, USA: IEEE Computer Society, 2008, pp. 149–156.
- [16] T. L. Alves de Souza Ramos, R. S. Oliveira, A. P. de Carvalho, R. A. C. Ferreira, and W. Meira Jr., "Watershed: A high performance distributed stream processing system," in *Proceedings of the 2011 23rd International Symposium on Computer Architecture and High Performance Computing*, ser. SBAC-PAD '11. Washington, DC, USA: IEEE Computer Society, 2011, pp. 191–198.
- [17] D. A. Adams, "A computation model with data flow sequencing," Ph.D. dissertation, Stanford University, 1969.



# SIMULACIÓN DE APLICACIONES PARALELAS Y MECANISMOS DE TOLERANCIA A FALLOS

Nilda M. Pérez Otero, Abigail R. N. Verazay, C. Marcelo Pérez Ibarra.

Grupo de Ingeniería de Software (GIS). Facultad de Ingeniería. Universidad Nacional de Jujuy

Ítalo Palanca 10. San Salvador de Jujuy. Jujuy. CP 4600

TE 54+388+4221591 CEL 54+388+154145307

[nilperez@fi.unju.edu.ar](mailto:nilperez@fi.unju.edu.ar), [abigailrn@gmail.com](mailto:abigailrn@gmail.com), [cmperezi@gmail.com](mailto:cmperezi@gmail.com)

## Resumen

Durante los últimos años, los sistemas de cómputo de altas prestaciones hicieron posible el tratamiento de grandes volúmenes de datos a altísimas velocidades de procesamiento para aplicaciones comerciales y científicas. La evolución de las arquitecturas paralelas y el desarrollo de los mecanismos de tolerancia a fallos permiten que la ejecución de las aplicaciones se realice de manera confiable. Teniendo en cuenta que la configuración de recursos, el tipo de aplicación y los mecanismos de tolerancia a fallos influye en el comportamiento de un sistema paralelo, su estudio se lleva a cabo a través de herramientas de simulación que permiten representar el sistema. En virtud de esto, el Grupo de Ingeniería de Software desarrolló un simulador de clusters que permite parametrizar la configuración física de un *cluster*, y las características de cómputo y comunicación de aplicaciones M/W y SPMD. En este trabajo se presentan la línea de investigación abordada.

## Contexto

La línea de investigación aquí presentada está inserta en los proyectos *Extensión de CluSim: Simulación de la arquitectura tolerante a fallos RADIC* (D/0120) y *Extensión Funcional de CluSim: Simulación de aplicaciones paralelas SPMD, Pipeline y Divide/Conquer* (D/0049). Ambos proyectos se encuentran acreditados y financiados por la Secretaria de Ciencia y Técnica y Estudios Regionales de la Universidad Nacional de Jujuy (SECTER-

UNJu), además, el proyecto (D/0120) se encuentra bajo el Programa de Incentivos.

## Introducción

En el mundo actual el procesamiento de gran cantidad de datos en el menor tiempo posible es una necesidad que se presenta tanto en el ámbito comercial como científico. Durante los últimos años, los sistemas de cómputo de altas prestaciones (HPC) contribuyeron gradualmente a suplir esta necesidad, no obstante los costos y complejidad asociados a ellos. Así fue que se desarrollaron técnicas y herramientas para la planificación y mantenimiento de tales sistemas a fin de disminuir el impacto negativo inherente a su implementación. Además, el avance tecnológico que acompañó a los sistemas de altas prestaciones permitió que las primeras arquitecturas, de enormes presupuestos, evolucionaran en alternativas tan accesibles como los *clusters* de computadoras [Catalán, 2003].

El uso de *clusters* se popularizó no sólo por su capacidad de procesamiento y menor costo sino por la confiabilidad y disponibilidad que alcanzaron mediante el empleo de mecanismos adicionales. La importancia de estos mecanismos se evidencia en los informes del Laboratorio Nacional de Los Álamos (LANL) que describen más de 20.000 eventos que causaron detenciones en la ejecución de aplicaciones considerando un período de 10 años y una veintena de *clusters* [Cappello, 2009]. Un estudio realizado por Schroeder y Gibson indica que algunos sistemas presentan más de mil fallos por año [Schroeder & Gibson, 2007] lo que

implica que si pueden ocurrir 3 fallas por día, las aplicaciones que utilizan todos los nodos de cómputo y demoran más de 8 horas, tienen pocas posibilidades de finalizar su ejecución correctamente.

Con el objetivo de mitigar los efectos de estos eventos, es importante contar con mecanismos que aseguren la ejecución exitosa de las aplicaciones en un entorno paralelo. Así, los llamados sistemas de tolerancia a fallos permiten la detección y recuperación de fallos, además de tareas de mantenimiento preventivo tales como el reemplazo de máquinas susceptibles a fallos sin interrupciones al sistema. [Santos *et al.*, 2008].

Es cada vez más frecuente realizar los estudios de las características de los sistemas HPC mediante modelos de simulación orientados al diseño y modelado de prestaciones [Denzel *et al.*, 2008], la exploración de arquitecturas o aplicaciones [Hammond *et al.*, 2009; Minkenberg & Rodriguez, 2009] o las herramientas de predicción de tráfico [Tikir *et al.*, 2009]. En la literatura existen muchos trabajos centrados en simular grandes redes y aplicaciones de HPC. Algunos de los trabajos orientados a la arquitectura de sistemas de altas prestaciones son:

- PARSEC: un entorno de simulación de eventos discretos que simula entidades y constructores de mensajes de comunicación entre entidades [Bagrodia *et al.*, 1998].
- SIMLAB: un entorno de simulación para SANs implementado en C++ que modela discos rígidos, nodos de enrutamiento e interfaces de redes. Se utiliza para el desarrollo y verificación de algoritmos distribuidos para la red PRES TO SAN [Berenbrink *et al.*, 2001].
- SIMCAN: un entorno de simulación para grandes redes complejas de almacenamiento que simula estas redes y sus subsistemas subyacentes (I/O, Networking, etc.) [Nuñez *et al.*, 2008].

Otros trabajos de simulación están centrados en las aplicaciones MPI. Entre ellos se encuentran:

- Un prototipo de un simulador híbrido que emula la ejecución de una aplicación paralela en modo stand-alone y la simulación de una red que se utiliza para el paso de mensajes de MPI. Este prototipo necesita que la aplicación simulada se ejecute en el hardware que se desea simular [Riesen, 2006].
- MPI-SIM: un simulador paralelo diseñado para predecir el rendimiento de aplicaciones MPI y MPI-IO ya existentes. MPI-SIM predice el rendimiento de estos programas en función de características de arquitectura (número de procesadores, latencias en la comunicación, algoritmos de caché, etc.) [Bagrodia *et al.*, 2008].

En algunos trabajos, ha sido útil la utilización de entornos de simulación de redes de propósito general, como OMNeT++ ([www.omnetpp.org](http://www.omnetpp.org)) y OPNET Modeler ([www.opnet.com](http://www.opnet.com)). Bajo estos entornos, es posible crear diferentes configuraciones de redes, con diferentes tipos de nodos, switches, topologías, protocolos, etc.

También es posible encontrar entornos de simulación de redes de propósito general que permiten crear diferentes configuraciones de redes, con diferentes tipos de nodos, switches, topologías, protocolos, etc. Ejemplos de éstos son OMNeT++ ([www.omnetpp.org](http://www.omnetpp.org)) y OPNET Modeler ([www.opnet.com](http://www.opnet.com)).

Considerando estos antecedentes, desde el año 2010, el Grupo de Ingeniería de Software (GIS) desarrolló CluSim, un simulador de clusters basado en OMNeT++ que permite parametrizar la configuración de recursos de un cluster, los patrones de cómputo y comunicación de la aplicación paralela y el tamaño del problema, de modo que sea posible evaluar y predecir el impacto en el rendimiento del sistema considerando diferentes configuraciones (número de nodos del

cluster). Este simulador, presentado en los trabajos [Pérez Ibarra *et al.*, 2010], [Valdiviezo *et al.*, 2010], [Lasserre *et al.*, 2011] y [García *et al.*, 2011], pretende servir como herramienta de soporte a la toma de decisiones para la selección de las configuraciones más adecuadas para un cluster que ejecuta un determinado tipo de aplicación paralela.

A fin de extender la funcionalidad de CluSim, el GIS definió las etapas de un proceso de modelado que permite la formulación de un modelo genérico de aplicaciones paralelas bajo distintos paradigmas de programación, proceso que ya fue utilizado para caracterizar aplicaciones Master/Worker y SPMD (Single Program Multiple Data) [Lasserre *et al.*, 2012; García *et al.*, 2012].

## Líneas de Investigación y Desarrollo

CluSim es un simulador de clusters homogéneos, desarrollado utilizando el framework de simulación OMNeT++, que permite parametrizar la configuración de recursos de un cluster, los patrones de cómputo y comunicación de la aplicación paralela y el tamaño del problema, de modo que sea posible evaluar y predecir el impacto de diferentes configuraciones en el rendimiento de un sistema. A partir de 2010 y a través de sucesivas extensiones se fueron incorporando nuevas funcionalidades, y a la fecha CluSim simula aplicaciones Master/Worker y SPMD.

En forma paralela a CluSim el GIS está desarrollando un modelo genérico para caracterización de aplicaciones paralelas que fue probado con los paradigmas Master/Worker y SPMD.

A partir de estas dos líneas: simulador y modelo genérico, estos proyectos pretenden:

- Continuar agregando funcionalidad a CluSim, por un lado para que permita la simulación de aplicaciones Pipeline y Divide/Conquer y por otro, para que permita simular sistemas HPC con distintas configuraciones de tolerancia a fallos.

- Ampliar el modelo genérico hacia los dos paradigmas restantes, Pipeline y Divide/Conquer.

## Objetivos y Resultados

### Objetivos

Los proyectos aquí presentados pretenden extender la funcionalidad básica del simulador CluSim reformulando su modelo de simulación para incluir no sólo distintos tipos de aplicaciones paralelas sino también los aspectos más relevantes de un mecanismo de tolerancia a fallos. Con esto en mente, se establecieron los siguientes objetivos:

- Caracterizar aplicaciones SPMD, Pipeline y Divide/Conquer.
- Identificar los parámetros característicos de aplicaciones SPMD, Pipeline y Divide/Conquer.
- Formular modelos de simulación para aplicaciones SPMD, Pipeline y Divide/Conquer.
- Integrar los modelos de simulación formulados a CluSim.
- Formular un modelo de simulación para tolerancia a fallos.
- Integrar el modelo de simulación de tolerancia a fallos a CluSim.

### Resultados Obtenidos

En 2012 se obtuvieron los siguientes resultados:

- Definición de las etapas de un proceso de modelado
- Reformulación del modelo de simulación de CluSim para aplicaciones Master/Worker
- Caracterización de aplicaciones SPMD
- Incorporación de aplicaciones SPMD al modelo de simulación de CluSim.

### Resultados Esperados

Para 2013 se prevé obtener los siguientes resultados:

- Caracterización de aplicaciones Pipeline y Divide/Conquer

- Incorporación de aplicaciones Pipeline y Divide/Conquer al modelo de simulación de CluSim
- Incorporación de mecanismos de tolerancia a fallos al modelo de simulación de CluSim

### Formación de Recursos Humanos

En el marco de este proyecto trabajan 6 docentes y 8 alumnos pertenecientes a las carreras Informáticas de la Facultad de Ingeniería de la UNJu y del Departamento Académico San Salvador (DASS) de la Universidad Católica de Santiago del Estero. Entre los docentes se cuenta con un Magíster que está finalizando los estudios de Doctorado y 3 que están realizando carreras de posgrado.

En la actualidad, se desarrollan 5 tesinas de grado y una beca CIN afines a la temática del proyecto y los alumnos restantes, menos avanzados en la carrera, fortalecen su formación en investigación. Como resultado del trabajo realizado, se han finalizado satisfactoriamente una tesina de grado y una beca CIN.

### Referencias

- [Bagrodia *et al.*, 2008] Bagrodia, R., E. Deelman, and T. Phan. Parallel simulation of large-scale parallel applications. *Int. J. High Perform. Comput. Appl.*, 15(1):3–12, 2001.
- [Bagrodia *et al.*, 2008] Bagrodia, R., E. Deelman, and T. Phan. Parallel simulation of large-scale parallel applications. *Int. J. High Perform. Comput. Appl.*, 15(1):3–12, 2001.
- [Berenbrink *et al.*, 2001] Berenbrink, P., A. Brinkmann, and C. Scheideler. Simlab - a simulation environment for storage area networks. In *Workshop on Parallel and Distributed Processing (PDP)*, pages 227–234, 2001.
- [Cappello, 2009] Cappello, F. (2009). Fault Tolerance in Petascale/ Exascale Systems: Current Knowledge, Challenges and Research Opportunities. *International Journal of High Performance Computing Applications*, 23(3):212–226.
- [Catalán, 2003] Catalán i Coit, M. “Nuevo Modelado de Computación Paralela con Clusters Linux”. VI Congreso HIPALinux. miKeL a.k.a.mc2. Septiembre 2003.
- [Denzel *et al.*, 2008] Denzel, W. E.; Li, J.; Walker, P. & Jin, Y. A framework for end-to-end simulation of high-performance computing systems. In *Simutools '08: Proceedings of the 1st international conference on Simulation tools and techniques for communications, networks and systems & workshops*, pp. 1--10, ICST, Brussels, Belgium, Belgium. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).
- [García *et al.*, 2011] García, A.; Pérez Otero, N. M.; Pérez Ibarra, C. M. y C. M. Lasserre. Simulación de Clusters: Integración de INET a CluSim. XVII Congreso Argentino de Ciencias de la Computación (CACIC 2011). ISBN 978-950-34-0756-1. Universidad Nacional de la Plata. Buenos Aires. pp. 367-373. Octubre 2011.
- [García *et al.*, 2012] García, A.; Lasserre, C. M.; Verazay, A. R. N.; Pérez Otero, N. M.; Pérez Ibarra, M.; Martínez, J. G. y S. A Nolasco. Modelado de Aplicaciones SPMD. VIII Jornadas de Ciencia y Tecnología de Facultades de Ingeniería del NOA. Ed. Grupo Loza Impresiones S.R.L. ISSN 1853-7871 San Miguel de Tucumán. Septiembre 2012.
- [Hammond *et al.*, 2009] Hammond, S. D.; Mudalige, G. R.; Smith, J. A.; Jarvis, S. A.; Herdman, J. A. & Vadgama, A. Warpp: a toolkit for simulating highperformance parallel scientific codes. In *Simutools '09: Proceedings of the 2nd International Conference on Simulation Tools and Techniques*, pp. 1--10, ICST, Brussels, Belgium, Belgium. ICST (Institute for Computer Sciences, Social-



- Informatics and Telecommunications Engineering).
- [Lasserre *et al.*, 2012] Lasserre, C. M.; García, A.; Pérez Otero, N. M.; Verazay, A. R. N.; Pérez Ibarra, M.; Nolasco S. A. y J. G. Martínez. Hacia un Modelo Genérico de Aplicaciones Paralelas. XVIII Congreso Argentino de Ciencias de la Computación (CACIC 2012). pp. 296-305. ISBN 978-987-1648-34-4. Universidad Nacional del Sur. Bahía Blanca. Octubre 2012.
- [Lasserre *et al.*, 2011] Lasserre, C. M.; Pérez Ibarra, C. M.; Valdiviezo, L. M.; Verazay, A. R. N.; Quispe, G. L.; Nolasco, S. A.; Chosco, V. H. y N. M. Pérez Otero. Adaptación de CluSim a clusters heterogéneos. Investigaciones en Facultades de Ingeniería del NOA. Tomo 2 – 2011. ISSN 1853-7871. Editorial Científica Universitaria. San Fernando del Valle de Catamarca. pp 1053-1060. Octubre 2011.
- [Minkenberg & Rodriguez, 2009] Minkenberg, C. & Rodriguez, G. Tracedriven co-simulation of high-performance computing systems using OMNeT++. In Simutools '09: Proceedings of the 2nd International Conference on Simulation Tools and Techniques, pp. 1--8, ICST, Brussels, Belgium, Belgium. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).
- [Nuñez *et al.*, 2008] Nuñez, A., J. Fernandez, J. D. Garcia, L. Prada, and J. Carretero. Simcan: a simulator framework for computer architectures and storage networks. In Simutools '08: Proceedings of the 1st international conference on Simulation tools and techniques for communications, networks and systems & workshops, pages 1–8, ICST, Brussels, Belgium, Belgium, 2008. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).
- [Pérez Ibarra *et al.*, 2010] Pérez Ibarra, C. M., L. M. Valdiviezo, N. M. Pérez Otero, H. P. Liberatori, D. Rexachs, E. Luque y C. M. Lasserre. CLUSIM: Simulador de Clusters para Aplicaciones de Cómputo de Altas prestaciones basado en OMNeT++. XVI Congreso Argentino de Ciencias de la Computación. Buenos Aires (Morón). Octubre 2010.
- [Riesen, 2006] Riesen, Rolf. “Supercomputer Simulation Design Through Simulation”. Cray User Group (CUG). ISBN: 1-4244-0327-8. Barcelona. pp. 1–9. 2006.
- [Santos *et al.*, 2008] Santos, G.; Duarte, A.; Rexachs, D. & Luque, E. Providing non-stop service for message-passing based parallel applications with RADIC. In Luque, E.; Margalef, T. & Benitez, D., editores, Euro-Par, volume 5168 of Lecture Notes in Computer Science, pp. 58–67. Springer.
- [Schroeder & Gibson, 2007] Schroeder, B. & Gibson, G. A. Understanding failures in petascale computers. Journal of Physics: Conference Series, 78:012022 (11pp).
- [Tikir *et al.*, 2009] Tikir, M. M.; Laurenzano, M. A.; Carrington, L. & Snaveley, A. PSINS: An open source event tracer and execution simulator for MPI applications. In Euro-Par '09: Proceedings of the 15th International Euro-Par Conference on Parallel Processing, pp. 135--148, Berlin, Heidelberg. Springer-Verlag.
- [Valdiviezo *et al.*, 2010] Valdiviezo, L. M.; Pérez Otero, N. M.; Pérez Ibarra, C. M. y C. M. Lasserre. Caracterización de una aplicación paralela con distintas configuraciones en CluSim. Investigaciones en Facultades de Ingeniería del NOA – 2010. ISSN 3367-5072. Ed. EdiUNJu. S. S. de Jujuy. pp. 499-504. Noviembre 2010.

# Escalabilidad y Paralelización mediante el uso de Hadoop Distributed File System

Damián P. Barry<sup>1</sup>, Carlos E. Buckle<sup>1</sup>, Rodrigo Jaramillo<sup>1</sup>,  
Ignacio Real<sup>1</sup>, Fernando G. Tinetti<sup>2</sup>

<sup>1</sup>Depto. de Informática, Fac.de Ingeniería, Universidad Nacional de la Patagonia San Juan Bosco.  
Puerto Madryn, Argentina  
+54 280-4472885 – Int. 117.

[damian\\_barry@unpata.edu.ar](mailto:damian_barry@unpata.edu.ar), [cbuckle@unpata.edu.ar](mailto:cbuckle@unpata.edu.ar), [rodrigojaramillo@gmail.com](mailto:rodrigojaramillo@gmail.com), [nachoreal@hotmail.com](mailto:nachoreal@hotmail.com)

<sup>2</sup>III-LIDI, Facultad de Informática - Universidad Nacional de La Plata  
<sup>2</sup>Investigador Comisión de Investigaciones Científicas de la Prov. de Bs. As.  
La Plata, Argentina

## Resumen

En el marco del proyecto de investigación TECNICAS DE RECUPERACION DE INFORMACION EN GRANDES VOLUMENES DE DATOS HETEROGENEOS CON BASES DE DATOS NOSQL, el presente trabajo se orienta a evaluar configuraciones de clusters utilizando Hadoop Distributed File System (HDFS) para comprobar las capacidades de disponibilidad, escalabilidad y paralelización en la recuperación de información.

Dicha evaluación permitirá establecer las capacidades necesarias con las que debería contar un File System Distribuido, tanto desde la perspectiva de almacenamiento y técnicas de indexación, como de distribución de las consultas, paralelización, escalabilidad y rendimiento en ambientes heterogéneos.

Para ello se diseñarán arquitecturas tanto centralizadas como distribuidas, y se realizarán las correspondientes verificaciones, estableciendo los porcentajes de mejora de rendimiento para cada arquitectura.

Palabras Clave: Recuperación de Información, HDFS, MapReduce, cluster, Paralelización, Escalabilidad, Rendimiento.

## Contexto

Esta presentación corresponde al proyecto de investigación “Técnicas de recuperación de información en grandes volúmenes de datos heterogéneos con bases de datos No-Sql” desarrollado por docentes y alumnos de la Facultad de Ingeniería de la Universidad Nacional de La Patagonia San Juan Bosco (UNPSJB) Sede Puerto Madryn. El proyecto es financiado por la Secretaría de Ciencia y Técnica de dicha Universidad y se vincula con el III-LIDI Facultad de Informática de la Universidad Nacional de La Plata a través del director quien se desempeña como investigador en dicho laboratorio.

## Introducción

La popularidad de las redes sociales, los sistemas de gestión de contenidos (CMS, Content Management Systems) como portales en general y como plataformas de colaboración, el crecimiento en la producción de información dentro de las organizaciones ya sea por producción de los sistemas o por la digitalización de información existente han generado un crecimiento exponencial en los volúmenes de información producidos, pasando de hablar de gigabyte de información a hablar con total normalidad del orden de los petabyte<sup>[13][14][15]</sup>.

Esta situación y las aplicaciones en cuestión suelen requerir mayor cantidad de recursos que los que están disponibles en una sola unidad de cómputo. El desafío se encuentra por lo tanto en producir infraestructura de cómputo que permita re-utilizar las unidades existentes, normalmente heterogéneas entre sí, de forma colaborativa, alcanzando capacidades de disponibilidad y escalabilidad acordes al incremento en el procesamiento de la información<sup>[1][5]</sup>.

La re-utilización de equipamiento de bajo costo de forma colaborativa que permita abordar esta problemática, requiere de técnicas de sistemas distribuidos, donde cada servidor cuente con capacidad de almacenamiento y cómputo local que permita que procesos y acceso a datos puedan ser distribuidos y balanceados en el cluster heterogéneo<sup>[8]</sup>.

Indagando distintas alternativas para solucionar la distribución de información y de las búsquedas en un ambiente heterogéneo y escalable se definieron un conjunto de propiedades deseables que debiera cumplir una solución<sup>[1][2][3][16]</sup>: Rendimiento, Tolerancia a Fallas y Ejecución en ambientes heterogéneos. Por otra parte, las soluciones NoSQL para administrar grandes volúmenes de información se basan normalmente en la conformación de un sistema de nodos heterogéneos.

Existen diversas técnicas que permiten configurar ambientes heterogéneos y/o mixtos.

En nuestro caso particular de DFS, hemos concentrado el esfuerzo del presente trabajo en el Framework programado en Java llamado Hadoop para almacenar y procesar grandes volúmenes de datos a través de grupos de unidades de cómputo y almacenamiento, llamado Hadoop Distributed File System (HDFS).

El HDFS además de ser un sistema de archivos distribuido, escalable y portable, resuelve el manejo de disponibilidad y confiabilidad replicando los datos en múltiples servidores<sup>[8]</sup>.

Dentro de algunas de las características a destacar de HDFS, podemos mencionar:

### Replicación con balanceo de carga

La arquitectura debe garantizar un conjunto de nodos con la información replicada y consistente en todos los nodos. En este caso el motor de búsqueda cuenta con un pooling de nodos de datos en los cuales buscar la información. La arquitectura misma no paraleliza las búsquedas, simplemente distribuye la carga entre los nodos. Como los nodos son independientes y auto-suficientes son capaces de responder consistentemente a la consulta realizada ya que la responsabilidad de recuperación está en el nodo de datos y la responsabilidad de distribución de carga en el balanceador. La desventaja de este método es que ni resuelve el problema espacial de la información ni paraleliza la búsqueda<sup>[3]</sup>.

### Scatter and Gather

El método realiza un broadcast de la búsqueda de información requerida en sus nodos conocidos, realizando una dispersión de la misma. Cada nodo (independiente de los demás) tiene la capacidad de elaborar una respuesta con la información que contiene dicho nodo. Todas las respuestas se concentran en el nodo que realizó la dispersión y éste es responsable de consolidar las mismas en una única (y consistente) respuesta a la petición. Una ventaja adicional del método es que a su vez los nodos de datos pueden ser dispersores en nuevos nodos (conocidos por él). Conformando de esta forma una red de nodos independientes que contienen información.

Las ventajas en este caso son el particionamiento de la información y la paralelización de las búsquedas. La desventaja es una sobrecarga en la distribución de la información, especialmente si se desea realizar con alguna lógica de segmentación en particular: Geográfica, tipo de contenido, atributos ontológicos, etc. En este último caso se requiere conocimiento e información sobre los contenidos (datos) a ser almacenados, siendo en algunos casos relativamente compleja su resolución, especialmente ante la aplicación de reglas ontológicas sobre los contenidos<sup>[2][3][7]</sup>.

Para este caso es interesante poder aplicar la técnica de Map / Reduce que es una buena técnica para procesar gran volumen de datos en paralelo. El modelo provee un

mecanismo de particionamiento de información que permite distribuir “inteligentemente” de acuerdo a reglas pre-definidas los datos en distintos nodos auto-contenidos.

Map /reduce es una técnica que implica paralelizar los datos clasificados por claves. Esto permite además paralelizar el procesamiento en las búsquedas, donde la clave de la técnica se basa en la inteligencia de separación de los datos. A su vez una ventaja adicional radica en el ahorro de espacio en el resultado de las claves compartidas al reducirlas dentro de un documento<sup>[8][9][17][18]</sup>.

### Líneas de Investigación y Desarrollo

Las principales líneas de investigación se podrían resumir en la lista siguiente:

- Determinar la factibilidad y aplicabilidad de los métodos teóricos en los entornos prácticos estudiados, especialmente en lo referido a la distribución de información e information retrieval.
- Proponer mejoras o nuevas técnicas y/o reformulaciones a las técnicas existentes para el manejo de recursos, en lo que se refiere a las técnicas de distribución y recuperación.
- Implementar y validar las técnicas y métodos propuestos sobre plataformas de desarrollo concretas.
- Seleccionar material bibliográfico y generar una base de conocimiento sobre las técnicas y métodos empleados en los esquemas de particionamiento, replicación, distribución e indexado en la infraestructura apache Hadoop.

### Resultados y Objetivos

- Diseñar distintas arquitecturas del cluster con HDFS variando la cantidad de nodos del cluster y analizar los tiempos de respuestas.
- Investigar y seleccionar uno o varios métodos de particionamiento y replicación sobre Apache Hadoop.
- Verificar el esquema de escalabilidad y desempeño del cluster.
- Evaluar las distintas opciones de parametrización de la plataforma y evaluar distintas arquitecturas de configuración en Hadoop para una optimización.
- Definir métricas que permitan obtener conclusiones relevantes respecto a las técnicas y métodos implementados.
- Definir y desarrollar uno o varios métodos de pruebas de stress sistematizadas para someter a comparación las distintas configuraciones y parametrizaciones.

- Armar un banco de pruebas que permita comprobar las distintas implementaciones y métodos utilizados en el DFS seleccionado.

En la actualidad se ha logrado configurar un cluster HDFS de 8 equipos heterogéneos con un funcionamiento estable. Para ello fue necesario superar los inconvenientes en el armado de este tipo de infraestructura desde el punto de vista espacial en nuestra Unidad Académica como de conseguir los recursos necesarios para poder llevarla adelante.

El cluster hadoop cuenta en la actualidad con la siguiente configuración de equipos inter-conectados mediante un switch de 1Gb:

- PC1: Procesador: AMD Turion(tm) X2 Dual-Core Mobile RM-72 - Memoria: 4 GB - Disco: 320.1 GB. NameNode y DataNode.
- PC2: Procesador: Intel(R) Core(TM) i5-2400 CPU @ 3.10GHz - Memoria: 10 GB- Disco: 500.1 GB. DataNode.
- PC 3: Procesador: Intel(R) Core(TM) i7-2600 CPU @ 3.40GHz - Memoria: 16 GB - Disco: 500.1 GB. DataNode.
- PC 4: Procesador: Intel(R) Core(TM) i7-2600 CPU @ 3.40GHz - Memoria: 8 GB - Disco: 1000.0 GB. DataNode.
- PC 5: Procesador: Intel(R) Celeron(R) D CPU 3.20GHz - Memoria: 685 MB - Disco: 80.0 GB. DataNode.
- PC 6: Procesador: Intel(R) Celeron(R) D CPU 3.20GHz - Memoria: 877 MB - Disco: 80.0 GB
- PC 7: Procesador: Intel(R) Celeron(R) D CPU 3.20GHz - Memoria: 685 MB - Disco: 80.0 GB. DataNode.
- PC 8: Procesador: Intel(R) Core(TM)2 Duo CPU E4400 @ 2.00GHz - Memoria: 1 GB - Disco: 160.0 GB. DataNode.

Sobre el cluster descrito se está comenzando a diseñar las pruebas de carga y stress que permitan determinar conclusiones sobre las distintas opciones de configuración del cluster Hadoop, respecto de la disponibilidad y la escalabilidad.

Las pruebas de carga se realizarán utilizando las herramientas siege y sar de linux. Los resultados serán publicados oportunamente.

### Formación de Recursos Humanos

En lo referido a Formación de Recursos Humanos este proyecto propone las siguientes metas:

- Consolidar mediante el proyecto, un grupo de investigación de la Universidad Nacional de la Patagonia San Juan Bosco sede Puerto Madryn, sobre la disciplina Bases de Datos NoSql. Este grupo se integra actualmente de 4 profesores, 1

JTP y 1 Auxiliar, además participan del mismo 4 alumnos del ciclo superior de los cuales 2 alumnos están realizando el presente trabajo de Tesis.

- Fomentar, incentivar y difundir las tareas de investigación.
- Fomentar la realización de las Tesinas de Grado en el marco del proyecto de investigación. En particular el presente trabajo es el resultado de la ejecución del trabajo final de carrera de 2 de los autores
- Mejorar la formación de recursos humanos altamente calificados, con capacidades de investigación y desarrollo. Especialmente orientados a la inserción laboral altamente calificada en nuestro medio productivo.
- Contribuir a la creación en un futuro Centro o Instituto en investigación informática.
- Interactuar con otros grupos de investigación de las sedes de la universidad y de otras universidades, en tareas conjuntas de investigación y desarrollo, como también en la formación de recursos humanos.
- Incrementar el número de proyectos acreditados y de trabajos publicados por la universidad y la sede.

### Referencias

- [1] M.T. Özsu & P. Valduriez. "Principles of Distributed Database Systems, 2nd edition". Prentice-Hall, 1999. Sitio web: <http://softbase.uwaterloo.ca/~tozsu/ddbook/notes.html>
- [2] David Taniar & Clement H. C. Leung & Wenny Rahayu & Sushant Goel. "High Performance Parallel Database Processing and Grid Databases". John Wiley & Sons, 2008.
- [3] P. Valduriez. "Data Management and Parallel Processing". Chapman and Hall, 1992.
- [4] M.Cohn, "Succeeding with Agile: Software Development Using Scrum", Pearson Education, 2010.
- [5] Ahmed K. Elmagarmid & Marek Rusinkiewicz & Amit Sheth. "Management of Heterogeneous and Autonomous Database Systems". Morgan Kaufmann Publishers, 1999.
- [6] Kristina Chodorow & Michael Dirolf. "MongoDB: The Definitive Guide". O'Reilly, 2010.
- [7] Satnam Alag. "Collective Intelligence in Action". Manning Publication, 2009.
- [8] Jason Venner. "Pro Hadoop". Apress, 2009.
- [9] Tom White. "Hadoop: The Definitive Guide", O'Reilly, 2011.



- [10] Michael McCandless & Erik Hatcher. "Lucene in Action, Second Edition: Covers Apache Lucene 3.0". Manning Publication, 2010.
- [11] David Smiley & Eric Pugh. "Solr 1.4 Enterprise Search Server". Packt Publishing, 2009.
- [12] Erik Hatcher, Otis Gospodnetić. "Lucene in Action", 2nd. ed, Manning Publications Co. 2004.
- [13] WhiteHouse.gov Goes Drupal,  
<http://personaldemocracy.com/node/15131>
- [14] Thoughts on the Whitehouse.gov switch to Drupal,  
<http://radar.oreilly.com/2009/10/whitehouse-switch-drupal-opensource.html>
- [15] Cal Henderson: "Building Scalable Web Sites", O'Reilly Media, 2006
- [16] Ricky Ho: Scalable System Design Patterns, Pragmatic Programming Techniques.  
<http://horicky.blogspot.com/2010/10/scalable-system-design-patterns.html>
- [17] Azza Abouzeid, Kamil BajdaPawlikowski, Daniel Abadi1, Avi Silberschatz, Alexander Rasin: HadoopDB: An Architectural Hybrid of MapReduce and DBMS Technologies for Analytical Workloads.
- [18] Jeffrey Dean and Sanjay Ghemawat: MapReduce: Simplified Data Processing on Large Clusters. Google Inc.
- [19] Michael Stonebraker: The Case for Shared Nothing. University of California, Berkeley, Ca.

## Algoritmos Paralelos y Distribuidos para Cómputo de Altas Prestaciones. Fundamentos, Aplicaciones y Evaluación de rendimiento

Marcelo Naiouf<sup>(1)</sup>, Armando De Giusti<sup>(1)(2)</sup>, Laura De Giusti<sup>(1)</sup>, Franco Chichizola<sup>(1)</sup>, Victoria Sanz<sup>(1)(2)</sup>,  
Fabiana Leibovich<sup>(1)</sup>, Enzo Rucci<sup>(1)</sup>, Silvana Gallo<sup>(1)</sup>, Erica Montes de Oca<sup>(1)</sup>

<sup>(1)</sup> Instituto de Investigación en Informática LIDI (III-LIDI) - Facultad de Informática – UNLP <sup>(2)</sup> CONICET  
{mnaouf, degiusti, ldgiusti, francoch, vsanz, fleibovich, erucci, sgallo, emontesdeoca}@lidi.info.unlp.edu.ar,

### CONTEXTO

Se presenta una línea de Investigación que es parte del Proyecto 11/F011 “Procesamiento paralelo y distribuido. Fundamentos y aplicaciones en Sistemas Inteligentes y Tratamiento de imágenes y video” del III-LIDI acreditado por el Ministerio de Educación, y de proyectos acreditados y subsidiados por la Facultad de Informática de la UNLP.

Existe cooperación con Universidades de Argentina, Latinoamérica y Europa a través de proyectos acreditados por AECID, CyTeD, y OEI, y becas de Telefónica de Argentina. Asimismo, el Instituto forma parte del Sistema Nacional de Cómputo de Alto Desempeño (SNCAD).

### RESUMEN

El eje central de la línea de investigación lo constituye el estudio de los temas de procesamiento paralelo y distribuido para cómputo de altas prestaciones, en lo referente a los fundamentos y a las aplicaciones. Incluye problemas de software asociados con la construcción, evaluación y optimización de algoritmos concurrentes, paralelos y distribuidos sobre arquitecturas multiprocesador.

Los temas de interés abarcan aspectos de fundamentos como diseño y desarrollo de algoritmos paralelos en diferentes arquitecturas multiprocesador y plataformas de software, paradigmas paralelos, modelos de representación de aplicaciones, mapeo o asignación de procesos a procesadores, métricas, escalabilidad, balance de carga, evaluación de performance. Las arquitecturas pueden ser homogéneas o heterogéneas.

Las áreas de experimentación se enfocan principalmente a la concepción de aplicaciones paralelas numéricas y no numéricas sobre grandes volúmenes de datos y/o de cómputo intensivo, con el fin de obtener soluciones de alto rendimiento.

Se han incorporado temáticas como el uso de GPU en el desarrollo de soluciones, y el análisis del consumo y la eficiencia energética en algoritmos paralelos.

Este proyecto se coordina con otros dos en curso en el III-LIDI, relacionados con Arquitecturas Distribuidas y Paralelas y Sistemas de Software Distribuido. Existe colaboración con el grupo CAOS del Departamento de Arquitectura de Computadores y Sistemas Operativos de la Univ. Autónoma de Barcelona en la dirección de tesis de postgrado.

**Palabras clave:** *Sistemas paralelos. Algoritmos paralelos y distribuidos. Clusters. Multicore. GPU. Balance de carga. Evaluación de performance. Consumo energético*

### 1. INTRODUCCION

Debido al interés por el desarrollo de soluciones a problemas con creciente demanda computacional y de almacenamiento, el procesamiento paralelo y distribuido se ha convertido en un área clave dentro de la Ciencia de la Computación, produciendo transformaciones en las líneas de I/D [BEN06][BIS08][RAU10][SHA08][BEC08].

Más allá de las mejoras en la evolución de las arquitecturas físicas, el desafío está centrado en cómo aprovechar sus prestaciones. En este sentido, interesa realizar I/D en la especificación, transformación, optimización y evaluación de algoritmos distribuidos y paralelos, En esta línea de I/D la mayor importancia está en los algoritmos paralelos y en los métodos utilizados para su construcción y análisis [QIU08].

Uno de los cambios de mayor impacto ha sido el uso de manera masiva de procesadores con más de un núcleo (multinúcleo o multicore). Esto ha producido plataformas distribuidas híbridas (memoria compartida y distribuida), llevando a la necesidad de desarrollar sistemas operativos, lenguajes y algoritmos que las usen adecuadamente. También ha surgido la utilización de arquitecturas many-core como las placas gráficas de uso general como máquinas paralelas de memoria compartida, lo que constituye una plataforma con un paradigma de programación propio asociado.

La creación de algoritmos paralelos en arquitecturas multiprocesador, o la paralelización de un algoritmo secuencial, no es un proceso directo. El costo puede ser alto en términos del esfuerzo de programación [SHA08], y el manejo de la concurrencia adquiere un rol central en el desarrollo. Los pasos básicos para diseñar aplicaciones paralelas incluyen particionamiento, comunicación, aglomeración y mapeo de procesos a procesadores. Si bien en las primeras etapas el diseñador puede abstraerse de la máquina sobre la cual ejecutará el algoritmo, para obtener buen rendimiento en general debe tenerse en cuenta la plataforma de destino, dando lugar al concepto de *sistema paralelo* como combinación de hardware y software.

En las máquinas multiprocesador, se deben identificar las capacidades de procesamiento, interconexión, sincronización y escalabilidad [PAR09]. La caracterización y estudio de rendimiento del sistema de comunicaciones es de interés para la predicción y optimización de performance, así como la homogeneidad o heterogeneidad de los procesadores.

Muchos problemas algorítmicos se vieron fuertemente impactados por el surgimiento de las máquinas multicore (que integran dos o más núcleos computacionales dentro de un mismo chip), y la

tendencia creciente al uso de clusters de multicores. A partir de incorporar varios chips multicore dentro de un nodo y conectar múltiples nodos vía red, se puede crear una arquitectura NUMA, de modo que los cores en un chip compartan memoria principal, y puedan acceder remotamente a la memoria dedicada de otro chip, aunque ese acceso sea más costoso. Surgen varios niveles de comunicación: Intra CMP (2 cores del mismo chip), Inter CMP (2 cores que radican en distintos chips pero en el mismo nodo), e Inter Nodo (2 cores de 2 nodos distintos).

Esto impacta sobre el desarrollo de algoritmos que aprovechen adecuadamente las arquitecturas, y motiva el estudio de performance en sistemas híbridos [CHA07][SID07]. Además, es necesario estudiar la utilización de diferentes lenguajes ya que aún no se cuenta con un standard, aunque puede mencionarse el uso de MPI, OpenMP y Pthreads.

Para algunos problemas ha crecido la utilización de arquitecturas many-core como las placas gráficas de uso general (GPGPU, o *general purpose graphic processing unit*) como máquinas paralelas de memoria compartida [PIC11][KIR10]. Esto se debe a la gran cantidad de núcleos de procesamiento disponibles, buena performance y costo accesible. Dentro de los lenguajes asociados pueden mencionarse CUDA y OpenCL [LUE08][NOT09][NVI08].

### Métricas de evaluación del rendimiento y balance de carga

La diversidad de opciones vuelve complejo el análisis de performance de los Sistemas Paralelos, ya que los ejes sobre los cuales pueden compararse dos sistemas son varios. Existe un gran número de métricas para evaluar el rendimiento, siendo las tradicionales tiempo de ejecución, speedup, eficiencia.

La *escalabilidad* permite capturar características de un algoritmo paralelo y la arquitectura en que se lo implementa. Posibilita testear la performance de un programa sobre pocos procesadores y predecirla en un número mayor, así como caracterizar la cantidad de paralelismo inherente en un algoritmo.

En arquitecturas distribuidas, los problemas que caracterizan el análisis de los algoritmos paralelos aparecen potenciados por las dificultades propias de la interconexión en una red en general no dedicada. Esto se torna más complejo aún si cada nodo puede ser un multicore con varios niveles de memoria.

El uso de procesadores con múltiples núcleos conlleva cambios en la forma de desarrollar aplicaciones y software, y evaluar su rendimiento. La cantidad de threads disponibles en estos sistemas también es importante, ya que su creación y administración requiere del uso de recursos como memoria; además los threads deben ser cuidadosamente planificados (scheduling) e incorporados en la pila de ejecución. En este sentido, el desarrollo de técnicas de scheduling eficientes es un tema de interés.

El objetivo primario del cómputo paralelo es reducir el tiempo de ejecución haciendo uso eficiente de los

recursos. El *balance de carga* es un aspecto central y consiste en, dado un conjunto de tareas que comprenden un algoritmo y un conjunto de procesadores, encontrar el mapeo (asignación) de tareas a procesadores tal que cada una tenga una cantidad de trabajo que demande aproximadamente el mismo tiempo. Esto es más complejo si los procesadores (y comunicaciones) son heterogéneos. Dado que el problema general de mapping es NP-completo, pueden usarse enfoques que brindan soluciones subóptimas aceptables [OLI08]. Las técnicas de planificación a nivel micro (dentro de cada procesador) y macro (en un cluster) deben ser capaces de obtener buen balance de carga. Existen técnicas estáticas y dinámicas cuyo uso depende del conocimiento que se tenga sobre las tareas que componen la aplicación [LIU07][DUM08].

### Evaluación de performance. Aplicaciones

Es importante referirse a un algoritmo paralelo mencionando el modelo de computación para el que fue diseñado. Uno de los objetivos en la definición del modelo es la posibilidad de predicción de performance que brinde el mismo, teniendo en cuenta conceptos tales como comunicación, sincronización y arquitectura física. El desarrollo de nuevos modelos requiere caracterizar el contexto de comunicaciones entre los procesadores y la asociación entre los algoritmos, el paradigma de cómputo paralelo elegido y la arquitectura de soporte.

En la evaluación de performance de distintas clases de aplicaciones sobre las arquitecturas, interesa estudiar la influencia de las estrategias de distribución de procesos y datos, y la carga (estática o dinámica) asignada a cada procesador sobre el speedup, eficiencia y escalabilidad. Un aspecto de interés que se ha sumado como métrica es el del consumo energético requerido [FEN05].

Entre las aplicaciones de interés se encuentran las numéricas y no numéricas con alta demanda de cómputo.

## 2. LINEAS DE INVESTIGACION y DESARROLLO

- Paralelización de algoritmos secuenciales. Diseño y optimización de algoritmos.
- Comparación de lenguajes y bibliotecas para procesamiento paralelo y distribuido.
- Arquitecturas multicore y many-core. Multithreading en multicore. Multiprocesadores distribuidos.
- Estudio de complejidad de algoritmos paralelos, en particular considerando multicore y heterogeneidad.
- Modelos y paradigmas de computación paralela.
- Programación sobre modelo híbrido (pasaje de mensajes y memoria compartida) en cluster de multicores.
- Modelos de representación y predicción de performance de algoritmos paralelos.
- Mapping y scheduling de aplicaciones paralelas sobre distintas arquitecturas multiprocesador.

- Análisis de los problemas de migración y asignación de procesos y datos a procesadores.
- Balance de carga estático y dinámico. Técnicas.
- Evaluación de performance. Speedup, eficiencia, escalabilidad.
- Análisis de consumo y eficiencia energética en algoritmos paralelos.
- Implementación de soluciones sobre diferentes modelos de arquitecturas homogéneas y heterogéneas.

### 3. RESULTADOS OBTENIDOS/ESPERADOS

- Formar RRHH en los temas del Subproyecto, incluyendo tesis de postgrado y tesinas de grado.
- Desarrollar y optimizar algoritmos paralelos sobre diferentes modelos de arquitectura. En particular, en aplicaciones numéricas y no numéricas de cómputo intensivo y tratamiento de grandes volúmenes de datos
- Utilizar arquitecturas híbridas que combinan memoria compartida y pasaje de mensajes, evaluando performance para distintos modelos de comunicación.
- Estudiar y desarrollar modelos de representación de aplicaciones paralelas y distribuidas y los algoritmos de mapeo (estático y dinámico).
- Desarrollar algoritmos paralelos sobre GPU. Para problemas regulares y con alta demanda de cómputo, comparar los resultados con otras plataformas.
- Evaluar la performance (eficiencia, rendimiento, speedup, escalabilidad) de las soluciones propuestas.
- Mejorar y adecuar las técnicas disponibles para el balance de carga (estático y dinámico).
- Estudiar y proponer las adecuaciones necesarias para los modelos de predicción y evaluación de performance con diferentes paradigmas de interacción entre procesos, y distintas arquitecturas de soporte.
- Estudiar y comparar los lenguajes sobre las plataformas mencionados.
- Adicionalmente, estudiar el impacto producido por los modelos de programación, lenguajes y algoritmos sobre el consumo y la eficiencia energética.
- Iniciar la investigación sobre la paralelización de aplicaciones en plataformas que combinan multicore y GPU, o que disponen de más de una GPU en una misma máquina.

En este marco, pueden mencionarse los siguientes resultados:

- Para la experimentación se utilizaron y analizaron diferentes arquitecturas homogéneas o heterogéneas, incluyendo multicores, cluster de multicores con 128 núcleos y GPU.
- Se experimentó la paralelización en arquitecturas híbridas, con el objetivo de estudiar el impacto del mapeo de datos y procesos, así como de los lenguajes y librerías utilizadas.
- Respecto de los aplicaciones estudiadas y algoritmos implementados, se trabajó fundamentalmente con los siguientes problemas:

➤ **Best-first search paralelo sobre multicore y cluster de multicore:** El algoritmo de búsqueda A\* (variante de Best-First Search) es utilizado como base para resolver problemas combinatorios y de planificación, donde se requiere encontrar una

secuencia de acciones que minimicen una función objetivo para transformar una configuración inicial (problema a resolver) en una configuración final (solución). El alto requerimiento de memoria y cómputo causados por el crecimiento exponencial o factorial del grafo generado dinámicamente hacen imprescindible su paralelización, que permite beneficiarse de: (a) la gran cantidad de RAM y potencia de cómputo que provee un *cluster*, (b) la potencia de cómputo que proveen los procesadores *multicore*, (c) ambas características en caso de *cluster de multicore*. Tomando como caso de estudio el problema del N-Puzzle, se implementó el A\* secuencial para resolverlo, y dos versiones del paralelo Hash Distributed A\* (HDA\*[KIS12] [BUR10]) que realiza el balance de carga de los nodos generados del grafo mediante una función de Hash: (1) una versión utiliza MPI, haciendo posible su ejecución tanto sobre arquitecturas con memoria distribuida y memoria compartida, y (2) otra versión utilizando Pthreads, para su ejecución sobre un multiprocesador con memoria compartida, que elimina ciertas ineficiencias respecto a (1) cuando la ejecución se realiza sobre memoria compartida, como son las replicaciones de datos entre procesos de estructuras comunes utilizadas y la serialización de datos para la comunicación de nodos entre procesos. Interesa realizar un análisis comparativo del rendimiento alcanzado por HDA\*-MPI y HDA\*-Pthreads sobre un multicore y el desarrollo de una versión de HDA\* híbrida (Pthreads + MPI) para obtener mayor eficiencia al utilizar un cluster de multicore con respecto a la versión HDA\*-MPI.

➤ **Diagonalización de matrices por el método de Jacobi sobre arquitecturas multicore.** El método de Jacobi para diagonalizar matrices simétricas tiene aplicaciones en áreas como biometría, visión artificial, procesamiento digital de señales, entre otros. El incremento en el volumen de datos de entrada provoca un aumento significativo en el tiempo de cómputo. La combinación de librerías de álgebra lineal optimizadas para la arquitectura subyacente, junto con la potencia que brinda un multicore y herramientas de programación paralela para dicha arquitectura permiten reducir el tiempo de ejecución. Se abordó el análisis del problema, y se estudiaron distintas implementaciones del algoritmo secuencial y optimizaciones posibles, la adaptación para usar una implementación de la API BLAS (Basic Linear Algebra Subprograms) optimizada para la arquitectura, y se implementó un algoritmo paralelo utilizando OpenMP. Se realizó un estudio de los tiempos de ejecución de las soluciones secuenciales, observando un mejor rendimiento para los algoritmos que hacen uso de librerías optimizadas para álgebra lineal (ATLAS). Se analizó el rendimiento (speedup, eficiencia) obtenido por el algoritmo paralelo propuesto sobre un multicore, a medida que se incrementa el volumen de datos de entrada (tamaño de la matriz) y al aumentar la cantidad de threads /cores, demostrando la escalabilidad del sistema paralelo



propuesto [SAN12]. Como líneas de trabajo futuro se plantea la migración del algoritmo paralelo para ejecutar sobre GPU, el estudio de la escalabilidad sobre dicha arquitectura, y el análisis del consumo energético de los algoritmos paralelos propuestos.

➤ **Aplicaciones con paralelismo de datos.** Se avanzó en la paralelización en cluster de multicores, tomando como caso de estudio una aplicación base en muchos problemas como la multiplicación de matrices. Se estudió la mejora introducida por las soluciones analizando escalabilidad en dos sentidos, al incrementar tanto el tamaño del problema como la cantidad de núcleos en la experimentación. Se implementaron diferentes soluciones híbridas (pasaje de mensajes y memoria compartida) con dos enfoques: comparación de librerías de programación paralela (MPI + Pthreads vs MPI + OpenMP), y comparación de performance alcanzable por algoritmos que aprovechan eficientemente el uso de la cache L1, de dos maneras diferentes. Para este último análisis se implementó una solución que divide la matriz resultante en bloques para ser procesados al mismo tiempo que una segunda solución en la que las matrices de entrada son divididas en filas de bloques. Al aumentar el tamaño del problema, la segunda solución alcanza mejor performance dado que aprovecha la localidad espacial y temporal de la cache L1 más eficientemente [LEI12a][LEI12b]

➤ **Análisis de Secuencias de ADN:** el análisis de secuencias de ADN tiene múltiples aplicaciones, como la búsqueda de semejanzas entre dos de ellas. El gran tamaño que pueden alcanzar las secuencias (hasta  $10^9$  nucleótidos) y la complejidad computacional para compararlas por medio del algoritmo Smith-Waterman (orden  $N^2$ ) hacen necesaria la paralelización [RUC11]. Se implementaron soluciones paralelas empleando un esquema de pipeline y con diferentes modelos de comunicación (memoria compartida, pasaje de mensajes y combinación de ambos). Los algoritmos se ejecutaron sobre un cluster de multicores homogéneo. Se realizó un análisis de los rendimientos obtenidos contemplando las métricas speedup y eficiencia [RUC12a]

➤ **Construcción de árboles filogenéticos:** filogenia es la relación entre los diferentes conjuntos de especies del planeta, la cual puede representarse mediante un árbol. Su tarea consiste en inferir dicho árbol a partir de las observaciones realizadas sobre los organismos existentes. Los avances en las tecnologías de secuenciación, que permiten obtener conjuntos de datos cada vez más grandes, y la complejidad computacional para construir los árboles por medio del método Neighbor-Joining (orden  $N^3$ ) hacen necesaria su paralelización [STU88]. Se desarrollaron soluciones paralelas empleando un esquema maestro-esclavo y con diferentes modelos de comunicación (memoria compartida, pasaje de mensajes y combinación de ambos). Los algoritmos

se ejecutaron sobre un cluster de multicores homogéneo, combinando memoria compartida y distribuida. Se estudió el desempeño individual de cada solución (teniendo en cuenta speedup y eficiencia) y se realizó un análisis comparativo de los mismos [RUC12b] [RUC13].

➤ **Simulación de eventos discretos:** los problemas de simulación implican sistemas con grandes necesidades de cómputo; en particular, se estudió la paralelización sobre cluster de multicores con tres modelos diferentes de comunicación: memoria compartida (utilizando OpenMP y Pthreads), memoria distribuida (usando MPI) y soluciones híbridas, y variando la distribución de datos y procesos [GAL12]. Actualmente se trabaja en la exploración de diferentes técnicas de particionado, mapping y balance de carga para tal arquitectura, y la evaluación de herramientas que incorporan dichas técnicas con el fin de obtener nuevas alternativas o mejoras en las mismas [SOL12].

➤ **Problemas de tipo N-body.** Se utilizaron las plataformas de memoria compartida GPU y cluster de multicores para la resolución de problemas con alta demanda computacional del tipo N-body. Se emplearon diferentes modelos de comunicación: memoria compartida (Pthread en CPU y CUDA en GPU), pasaje de mensajes (MPI) y soluciones híbridas (MPI-Pthread). Además, se analizó el problema desde el punto de vista energético, introduciendo los fundamentos de consumo energético. Se han mostrado los beneficios del uso de la GPU en problemas con características similares al caso planteado. Los tiempos de ejecución obtenidos son considerablemente inferiores comparados con las soluciones implementadas en CPU. Los experimentos realizados desde el punto de vista del consumo energético favorecen el uso de la GPU en tales problemas [KIR10] [MON12].

#### 4. FORMACION DE RECURSOS HUMANOS

Dentro de la temática de la línea de I/D se concluyeron 3 Trabajos Finales de Especialización y 1 Tesina de Grado de Licenciatura. Se encuentran en curso en el marco del proyecto 8 tesis doctorales, 4 de maestría, 3 trabajos de Especialización y 3 Tesinas..

Se participa en el dictado de las carreras de Doctorado en Cs. Informáticas y Magíster y Especialización en Cómputo de Altas Prestaciones de la Facultad de Informática UNLP, por lo que potencialmente pueden generarse más Tesis y Trabajos Finales.

Existe cooperación con grupos de otras Universidades del país y del exterior, y hay tesis de diferentes Universidades realizando su trabajo con el equipo del proyecto.

#### 5. BIBLIOGRAFIA

- [BEC08] Becker Alexander (Editor), "Concurrent and Parallel Computing: Theory, Implementation and Applications", Nova Science Pub Inc, 2008, ISBN-10: 1604562749, ISBN-13: 9781604562743  
 [BEN06] Ben-Ari, M. "Principles of Concurrent and Distributed Programming, 2/E". Addison-Wesley. 2006

- [BIS08] Bischof C., Bucker M., Gibbon P., Joubert G., Lippert T., Mohr B., Peters F. (eds.), *Parallel Computing: Architectures, Algorithms and Applications*, Advances in Parallel Computing, Vol. 15, IOS Press, February 2008.
- [BUR10] Burns E, Lemons S, Ruml W, Zhou R. "Best First Heuristic Search for Multicore Machines". *Journal of Artificial Intelligence Research*, Vol.39, No.1, pp. 689-743, 2010.
- [CHA07] Chapman B., *The Multicore Programming Challenge*, Advanced Parallel Processing Technologies; LNCS, Vol. 4847, p3, Springer, November 2007.
- [DUM08] Dummler J., Rauber T., Runger G., *Mapping Algorithms for Multiprocessor Tasks on Multi-Core Clusters*, Proc. 2008 International Conference on Parallel Processing IEEE CS 2008.
- [FEN05] Feng, W.C., "The importance of being low power in HPC". *Cyberinfrastructure Technology Watch Quarterly*. 2005.
- [GAL12] Gallo S., Chichizola F., De Giusti L., Naiouf M.. "Análisis de soluciones paralelas puras e híbridas en un problema de simulación". CACIC2012. ISBN: 978987-1648-34-4. Pág. 367-376. Octubre 2012.
- [KIR10] Kirk D., Hwu W. "Programming Massively Parallel Processors: A Hands-on Approach (Applications of GPU Computing Series). Morgan-Kaufmann. 2010.
- [KIR10] Kirk, D., Hwu, W. "Programming Massively Parallel Processors: A Hands-on Approach". ISBN: 978-0-12-381472-2. Elsevier, 2010.
- [KIS12] Kishimoto A., Fukunaga A., Botea A. "Evaluation of a Simple, Scalable, Parallel Best-First-Search Strategy", Arxivpreprint: arXiv 1201.3204, 2012.
- [LEI12a] Leibovich F., Chichizola F., De Giusti L., Naiouf M., Tirado, F., De Giusti, A. "Programación híbrida en clusters de multicore. Análisis del impacto de la jerarquía de memoria". CACIC2012. ISBN: 978987-1648-34-4. Pág. 306-315. 2012.
- [LEI12b] Leibovich F., Naiouf M., De Giusti L., Tinetti F., De Giusti A.. "Hybrid Algorithms for Matrix Multiplication on Multicore Clusters". Procs PDPTA'12. Vol II. ISBN 1-60132-228-3J. 2012, USA.
- [LUE08] Luebke D. "Cuda: Scalable parallel programming for high-performance scientific computing". 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2008.
- [MON12] Montes de Oca E., De Giusti L., De Giusti A., Naiouf M. "Comparación del uso de GPU y cluster de multicore en problemas con alta demanda computacional". CACIC2012. ISBN: 978987-1648-34-4. Pág. 267-275. 2012.
- [NOT09] Nottingham A. y Irwin B. "GPU packet classification using openssl: a consideration of viable classification methods". Research Conf. of the South African Inst. of Comp. Sc. and Inf. Technologists. ACM, 2009.
- [NVI08] NVIDIA. "Nvidia CUDA compute unified device architecture, programming guide v.2.0". 2008.
- [OLI08] Olivier S., Prins S., Scalable Dynamic Load Balancing Using UPC. Proc. 37th ICPP'08. CD-ROM, IEEE CS, September 2008.
- [PAR09] Parashar M., Li Xiaolin, Chandra Sumir, "Advanced Computational Infrastructures for Parallel and Distributed Applications", Wiley-Interscience, 2009 ISBN-10: 0470072946
- [PIC11] Piccoli M.F., "Computación de Alto Desempeño utilizando GPU". XV Escuela Internacional de Informática. Editorial Edulp, 2011.
- [QIU08] Qiu X., Fox G. G., Yuan H., Bae S., Chrysanthakopoulos G., Nielsen H. F. "Performance of Multicore Systems on Parallel Data Clustering with Deterministic Annealing". LNCS 4331, pags. 407-416. ISBN 978-3-540-69383-3. Springer 2008.
- [RAU10] Rauber T., Rüniger G. "Parallel programming for multicore and cluster systems". Springer. 2010.
- [RUC11] Rucci E., Chichizola F., De Giusti L., Naiouf M., De Giusti A. "DNA sequence alignment: hybrid parallel programming on a multicore cluster". Procs. ICDCC'11. España, 2011. ISBN 978-1-61804-030-5, pp 183-190.
- [RUC12a] Rucci E., Chichizola F., Naiouf M., De Giusti L., De Giusti A. "Parallel pipelines for DNA sequence alignment on a cluster of multicores. A comparison of communication models". *Journal of Communication and Computer*. David Publishing Company, CA(EEUU) 2012, Vol9, N°12, pp. 1364-1371. ISSN: 1548-7709.
- [RUC12b] Rucci E., Chichizola F., Naiouf M., De Giusti A. "Performance comparison of parallel programming paradigms on a multicore cluster". Procs PDCS 2012, Vol. 1, ACTA Press, Las Vegas, EEUU, ISBN: 978-0-88986-940-0, pp. 216-221.
- [RUC13] Rucci E., Chichizola F., Naiouf M., De Giusti A. "A Hybrid Parallel Neighbor-Joining Algorithm for Phylogenetic Tree Reconstruction on a Multicore Cluster". To appear in *Journal Parallel & Cloud Computing*. American V-King Scientific Publishing, LTD, Nueva York (EEUU). ISSN print: 2304-9464. ISSN online: 2304-9456.
- [SAN12] Sanz V., De Giusti A., Naiouf M. Performance Analysis of a Matrix Diagonalization Algorithm with Jacobi Method on a Multicore Architecture. Procs PDPTA 2012 Las Vegas, USA. ISBN: 1-60132-227-5. Vol I. pp. 883-889
- [SHA08] Shavit N., Herlihy M., "The Art of Multiprocessor Programming", Morgan Kaufmann Pub, 2008, ISBN-13: 9780123705914
- [SID07] Siddh S., Pallipadi V., Mallick A.. "Process Scheduling Challenges in the Era of Multicore Processors". *Intel Technology Journal*, Vol. 11, Issue 04, November 2007.
- [SOL12] Solar Gallardo R., "Particionamiento y balance de carga en simulaciones distribuidas de bancos de peces", Tesis Doctoral, Universitat Autònoma de Barcelona, 2012
- [STU88] J. Studier and K. Keppler, "A note on the neighbor-joining algorithm of Saitou and Nei," *Mol. BioEvol.*, vol. 5, no. 6, 1988, pp. 729-731.
- [TEL08] Teller J., Ozguner F., Ewing R., Scheduling Task Graphs on Heterogeneous Multiprocessors with Reconfigurable Hardware, Proc. 37th ICPP'08 IEEE CS, Sept. 2008

# Arquitecturas Multiprocesador en HPC: Software de Base, Métricas y Aplicaciones.

De Giusti Armando E.<sup>2</sup>, Tinetti Fernando G.<sup>1</sup>, Naiouf R. Marcelo, Chichizola Franco, De Giusti Laura C., Villagaría Horacio<sup>1</sup>, Montezanti Diego<sup>3</sup>, Frati F. Emmanuel<sup>3</sup>, Pousa Adrián, Rodríguez Ismael P., Denham Mónica M., Iglesias Luciano

Instituto de Investigación en Informática LIDI (III-LIDI) - Facultad de Informática – UNLP

{degusti, fernando, mmaouf, francoch, ldgiusti, hvw, dmontezanti, fefrati, apousa, ismael, mdenham, li}@lidi.info.unlp.edu.ar

Con la colaboración en la dirección de Tesis de Posgrado de la Universidad Autónoma de Barcelona (España) y la Universidad Complutense de Madrid (España).

## Contexto

Esta línea de Investigación está dentro del proyecto “Arquitecturas Multiprocesador Distribuidas. Modelos, Software de Base y Aplicaciones” acreditado por el Ministerio de Educación y de proyectos específicos apoyados por organismos nacionales e internacionales.

Asimismo los proyectos “Eficiencia energética en Sistemas Paralelos” y “Algoritmos Paralelos utilizando GPGPUs. Análisis de rendimiento” financiados por la Facultad de Informática de la UNLP.

En el tema hay cooperación con varias Universidades de Argentina y se está trabajando con Universidades de América Latina y Europa en proyectos financiados por CyTED, AECID y la OEI (Organización de Estados Iberoamericanos).

Se participa en iniciativas como el Programa IberoTIC de intercambio de Profesores y Alumnos de Doctorado en el área de Informática.

Por otra parte, se tiene financiamiento de Telefónica de Argentina en Becas de grado y posgrado. Asimismo el III-LIDI forma parte del Sistema Nacional de Cómputo de Alto Desempeño (SNCAD) del MINCYT.

## Resumen

Caracterizar las arquitecturas multiprocesador distribuidas enfocadas especialmente a cluster y cloud computing, con énfasis en las que utilizan procesadores de múltiples núcleos (multicores y GPUs), con el objetivo de modelizarlas, estudiar su escalabilidad, analizar y predecir performance de aplicaciones paralelas y desarrollar esquemas de tolerancia a fallas en las mismas.

Profundizar el estudio de arquitecturas basadas en GPUs y su comparación con clusters de multicores, así como el empleo combinado de GPUs y multicores en computadoras de alta performance.

Analizar la eficiencia energética en estas arquitecturas paralelas, considerando el impacto de la arquitectura, el sistema operativo, el modelo de programación y el algoritmo específico.

Analizar y desarrollar software de base para clusters de multicores y GPUs, tratando de optimizar el rendimiento.

En el año 2012 se han agregado dos líneas de interés:

- El estudio de clusters híbridos, que combinen multicores y GPUs.
- La utilización de los registros de hardware de los procesadores para la toma de diferentes decisiones en tiempo de ejecución.

Es de hacer notar que este proyecto se coordina con otros dos proyectos en curso en el III-LIDI, relacionados con Algoritmos Distribuidos/Paralelos y Sistemas de Software Distribuido.

**Palabras claves:** Sistemas Paralelos. Multicore. GPU. Cluster, Grid y Cloud Computing. Clusters homogéneos y heterogéneos. Performance y eficiencia energética. Tolerancia a fallas en Sistemas paralelos. Modelos de programación de arquitecturas paralelas. Planificación. Scheduling. Virtualización.

## Introducción

La investigación en Sistemas Distribuidos y Paralelos es una de las líneas de mayor desarrollo en la Ciencia Informática actual [1][2][3]. En particular la utilización de arquitecturas multiprocesador configuradas en clusters, multiclusters, grids y clouds, soportadas por redes de diferentes características y topologías se ha generalizado, tanto para el desarrollo de algoritmos paralelos, la ejecución de procesos que requieren cómputo intensivo y la atención de servicios WEB concurrentes [4][5][6][7].

El cambio tecnológico, fundamentalmente a partir de los procesadores multicore, ha impuesto la necesidad de investigar en paradigmas “híbridos”, en los cuales coexisten esquemas de memoria compartida con mensajes [8][9][10][11]. Asimismo la utilización de procesadores gráficos (GPGPUs) como arquitecturas paralelas presenta una alternativa para alcanzar un alto speedup en determinadas aplicaciones [12].

Es importante en este contexto re-analizar el concepto de *eficiencia* incluyendo tanto al aspecto computacional como el energético y considerar el impacto del consumo sobre arquitecturas con miles de procesadores que trabajan concurrentemente [16].

<sup>1</sup>Investigador CIC Prov. de Bs. As. <sup>2</sup>Investigador CONICET <sup>3</sup>Becario CONICET



Esto conduce a un estudio de los lenguajes, paradigmas y herramientas orientados a la optimización de sistemas paralelos. [13][14][15].

Asimismo, aparecen líneas de I/D tales como la integración de arquitecturas heterogéneas con diferente performance según el tipo de aplicación, la planificación y scheduling dinámico basado en la carga de trabajo y en el consumo del sistema paralelo, el control en tiempo real de la frecuencia de reloj de los procesadores para optimizar consumo y el uso de los registros de hardware para la toma de decisiones en tiempo de ejecución [17][50].

Por otra parte, se profundiza el estudio y desarrollo de lenguajes, compiladores, estructuras de datos y soporte de comunicaciones adecuados a estas arquitecturas, así como se enfatizan los problemas de detección y tolerancia a fallos tratando de minimizar el overhead de tiempo y aprovechando alguna redundancia en la misma arquitectura [13][19][14][15][18].

La aparición de las arquitecturas tipo Cloud obliga a poner especial atención a los problemas de planificación, virtualización y predicción de performance (para la asignación dinámica de recursos). Naturalmente a mayor potencia del Cloud, también crecen las complejidades al analizar la comunicación y el acceso a memoria en arquitecturas que están distribuidas y a su vez conformadas por placas con un número variable de procesadores multicore y/o GPU [20][21][22][51].

En el proyecto se ha abierto una línea específicamente dedicada a los problemas de configuración y administración eficiente de Cloud.

### Definiciones básicas

Un *procesador multicore* integra dos o más núcleos computacionales dentro de un mismo “chip” [23][24]. La motivación de su desarrollo se basa en incrementar el rendimiento, reduciendo el consumo de energía en cada núcleo [52][53].

Una GPU (Graphics Processing Unit) es una arquitectura multicore dedicada a procesamiento gráfico, con un gran número de cores simples. En los últimos años, estas arquitecturas, fueron utilizadas para aprovechar su potencia de cómputo en aplicaciones de propósito general logrando un alto rendimiento y dando lugar al concepto de GPGPU (General-Purpose Computing on Graphics Processing Units) [12][48][49].

Un *cluster* es un sistema de procesamiento paralelo compuesto por un conjunto de computadoras interconectadas vía algún tipo de red, las cuales cooperan configurando un recurso que se ve como “único e integrado”, más allá de la distribución física de sus componentes. Cada “procesador” puede tener diferente hardware y sistema operativo, e incluso puede ser un “multiprocesador” [26].

La combinación de multicores y GPUs en un mismo cluster es un desafío actual, por las diferencias de arquitectura y del soporte de software de ba-

se que se ha desarrollado para cada uno de ellos. Esto es lo que denomina un “cluster híbrido”. [44][45][46][47].

Un *Grid* es un tipo de sistema distribuido que permite seleccionar, compartir e integrar recursos autónomos geográficamente distribuidos [28]. Un Grid es una configuración colaborativa que se puede adaptar dinámicamente según lo requerido por el usuario, la disponibilidad y potencia de cómputo de los recursos conectados. El Grid puede verse como un “entorno de procesamiento virtual”, donde el usuario tiene la visión de un sistema de procesamiento “único” y en realidad trabaja con recursos dispersos geográficamente [29]. Actualmente las arquitecturas Grid son utilizadas mayoritariamente en entornos colaborativos, en general no orientados a HPC.

Las arquitecturas tipo “Cloud” se presentan como una evolución natural del concepto de *Clusters* y *Grids*, integrando grandes conjuntos de recursos virtuales (hardware, plataformas de desarrollo y/o servicios), fácilmente accesibles y utilizables por usuarios distribuidos, vía WEB. Estos recursos pueden ser dinámicamente reconfigurados para adaptarse a una carga variable, permitiendo optimizar su uso [21][30][31][32].

Desde el punto de vista de la investigación tecnológica el mayor problema es el acceso directo a los recursos de hardware y el software de base de un Cloud (que en general son sistemas propietarios de grandes empresas/organizaciones).

### Aspectos de interés

- A partir de la complejidad creciente del hardware, se hace más desafiante el desarrollo de capas de software eficiente, desde el middleware hasta los lenguajes de aplicación [33][34][35][40].
- El incremento en el número de procesadores disponibles en clusters, grids y clouds obliga a poner énfasis en el desarrollo de los algoritmos de virtualización de modo de explotar la arquitectura con más de una aplicación concurrente [15].
- La heterogeneidad es inevitable en estos sistemas paralelos complejos. A su vez es un factor que condiciona la predicción de performance y consumo [19]. Por otro lado la heterogeneidad puede ser un factor “buscado” en el desarrollo de procesadores de múltiples núcleos, con el objetivo de especializar algunos núcleos a funciones específicas. [42] [43]
- Los modelos de predicción de performance resultan especialmente complejos. Es de interés el estudio de esquemas sintéticos (“firmas”) propios de la aplicación para estimar tiempos y consumo, ejecutando un código mínimo frente al de la aplicación real [37].
- Los problemas clásicos de scheduling y mapeo de procesos a procesadores tienen nuevos objetivos (en particular los relacionados con el consu-



mo) y deben considerar la migración dinámica de datos y procesos en función de performance y consumo [36].

- El tema de la detección y tolerancia a fallos de hardware y software se vuelve un punto crítico al operar sobre arquitecturas con gran número de procesadores, los cuales pueden reconfigurarse dinámicamente [38][39] [41].

## Líneas de Investigación y Desarrollo

### Temas de Estudio e Investigación

- Arquitectura de procesadores multicore. Clusters de multicores. Software de base.
- Arquitecturas de GPUs y Cluster de GPUs. Software de base.
- Nuevas estructuras de Cluster que integran multicores y GPUs. Software de base y lenguajes de programación de aplicaciones.
- Cloud computing. Software de base y aplicaciones de Cloud en el área de HPC.
- Green Computing. Consumo energético a nivel instrucción / algoritmo / aplicación y su utilización como parámetro de performance.
- Análisis comparativo de performance en cluster y cloud para problemas de HPC.
- Métricas y Modelos de predicción de performance para arquitecturas basadas en procesadores de múltiples núcleos. Utilización de la "firma" de aplicaciones regulares en predicción de comportamiento.
- Utilización de los registros de hardware de los procesadores para mejorar rendimiento en tiempo de ejecución.
- Técnicas de scheduling y virtualización en clusters, grid y cloud.
- Nuevas estructuras de datos, orientadas a procesamiento paralelo sobre clusters, grids y clouds.
- Detección y Tolerancia a Fallos (de hardware y software) en clusters, grids y clouds.

### Investigación experimental

- Desarrollo y evaluación de aplicaciones sobre cluster de multicores y cluster de GPUs.
- Pruebas de consumo en cluster de multicores y GPUs, analizando eficiencia computacional, escalabilidad y eficiencia energética.
- Desarrollo de experiencias con clusters híbridos (multicores y GPUs).
- Análisis, detección y corrección de fallos de hardware en arquitecturas de cluster de múltiples procesadores.
- Experiencias en Cloud Computing, utilizando equipamiento/servicios de proveedores y desa-

rollando las capas de emulación sobre un cluster de multicores clásico.

## Formación de Recursos Humanos

En cooperación con Universidades iberoamericanas se ha implementado la Maestría en Cómputo de Altas Prestaciones y se continúa dictando la Especialización en Cómputo de altas Prestaciones y Tecnología GRID.

En esta línea de I/D existe cooperación a nivel nacional e internacional. Hay 8 Investigadores realizando su Doctorado, 2 realizando la Maestría, 4 realizando la Especialización y 4 alumnos avanzados están trabajando en su Tesina de Grado de Licenciatura. En 2012 se aprobó 4 Trabajos Finales de Especialista y 2 Tesinas de Grado en temas del proyecto.

## Bibliografía

1. Grama A, Gupta A, Karypis G, Kumar V. "Introduction to parallel computing". Second Edition. Pearson Addison Wesley, 2003.
2. Dongarra J, Foster I, Fox G, Gropp W, Kennedy K, Torczon L, White A. "The Sourcebook of Parallel Computing". Morgan Kauffman Publishers. Elsevier Science, 2003.
3. Ben-Ari, M. "Principles of Concurrent and Distributed Programming, 2/E". Addison-Wesley, 2006.
4. Juhasz Z. (Editor), Kacsuk P. (Editor), Kranzlmuller D. (Editor). "Distributed and Parallel Systems: Cluster and Grid Computing". Springer; 1 edition (September 21, 2004).
5. Miller M. "Cloud computing: web-based applications that change the way you work and collaborate online". Que Publishing. USA 2008.
6. Di Stefano M., "Distributed data management for Grid Computing". John Wiley & Sons Inc. 2005.
7. Ghosh S. "Distributed System. An Algorithmic Approach". Chapman & Hall/CRC Computer and Information Science Series. 2006.
8. Mc. Cool M. "Programming models for scalable multicore programming". 2007. <http://www.hpcwire.com/features/17902939.html>
9. Lei Chai, Qi Gao, Dhabaleswar K. Panda. "Understanding the Impact of Multi-Core Architecture in Cluster Computing: A Case Study with Intel Dual-Core System". IEEE International Symp. on Cluster Computing and the Grid 2007 (CCGRID 2007), pp. 471-478 (May 2007).
10. Leibovich F., Chichizola F., De Giusti L., Naiouf M., Tirado Fernández F., De Giusti A.. "Programación híbrida en clusters de multicore. Análisis del impacto de la jerarquía de memoria". XII Workshop de Procesamiento Distribuido y Paralelo. CACIC2012. ISBN: 978987-1648-34-4. Pág. 306-315. 2012.
11. Rucci E., De Giusti A., Chichizola F., Naiouf M., De Giusti L. "DNA Sequence Alignment: hybrid parallel programming on multicore cluster". Pro-

- ceedings of the International Conference on Computers, Digital Communications and Computing (ICDCCC '11), Vol. 1, Nikos Mastorakis, Valeri Mladenov, Badea Lepadatescu, Hamid Reza Karimi, Costas G. Helmis (Editors), WSEAS Press, September 15-17, 2011, Barcelona, ISBN: 978-1-61804-030-5, pp. 183-190.
12. General-Purpose Computation on Graphics Processing Units. <http://gpgpu.org>.
  13. De Giusti L., Chichizola F., Naiouf M., De Giusti A.E., Luque E. "Automatic Mapping Tasks to Cores - Evaluating AMTHA Algorithm in Multicore Architectures". *IJCSI International Journal of Computer Science Issues*, Vol. 7, Issue 2, No 1, March 2010. ISSN (Online): 1694-0784. ISSN (Print): 1694-0814. Págs. 1-6.
  14. Olszewski M., Ansel J., Amarasinghe S. "Kendo: Efficient Deterministic Multithreading in Software". *Architectural Support for Programming Languages and Operating Systems*, 2009.
  15. Bertogna M., Grosclaude E., Naiouf M., De Giusti A., Luque E. "Dynamic on Demand Virtual Clusters in Grids". 3rd Workshop on Virtualization in High-Performance Cluster and Grid Computing. VHPC 08 – España. Agosto 2008.
  16. Feng, W.C., "The importance of being low power in high-performance computing". *Cyberinfrastructure Technology Watch Quarterly (CTWatch Quarterly)*. 2005.
  17. Frati E., Olcos Herrero K., Piñuel Moreno L., Montezanti D., Naiouf M., De Giusti A. "Optimización de herramientas de monitoreo de errores de concurrencia a través de contadores de hardware". *Proceedings del XVII Congreso Argentino de Ciencias de la Computación (CACIC 2011)*, La Plata (Argentina), 2011, ISBN: 978-950-34-0756-1. Págs: 337-346.
  18. Shirako J. et al. "Compiler Control Power Saving Scheme for Multi Core Processors". *LNCS Mayo 2007* – pp. 362-376.
  19. Suresh Siddha, Venkatesh Pallipadi, Asit Mallick. "Process Scheduling Challenges in the Era of Multicore Processors" *Intel Technology Journal*, Vol. 11, Issue 04, November 2007.
  20. Vaquero L.M. et al. "A Break in the Clouds: Towards a Cloud Definition". *ACM SIGCOMM Computer Communication Review*, vol. 39, num. 1, páginas 50-55, ISSN 0146-4833. Enero 2009.
  21. Foster I. "There's Grid in them thar Clouds". 2 de Enero, 2008. <http://ianfoster.typepad.com/blog/2008/01/theres-grid-in.html>. Noviembre, 2010.
  22. Rodriguez I., Pettoruti J., Chichizola F., De Giusti A. "Despliegue de un Cloud Privado para entornos de cómputo científico". *Proceedings del XVII Congreso Argentino de Ciencias de la Computación (CACIC'11)*, La Plata (Argentina), 2011, ISBN: 978-950-34-0756-1. Págs: 251-260.
  23. Burger T. W. "Intel Multi-Core Processors: Quick Reference Guide". [http://cachewww.intel.com/cd/00/00/23/19/231912\\_231912.pdf](http://cachewww.intel.com/cd/00/00/23/19/231912_231912.pdf)
  24. AMD. "Evolución de la tecnología de múltiple núcleo". <http://multicore.amd.com/es-ES/AMD-Multi-Core/resources/Technology-Evolution>. 2009.
  25. Adrian Pousa, Victoria Sanz, Armando De Giusti, "Análisis de rendimiento de un algoritmo de criptografía simétrica sobre arquitecturas multicore", *Proceedings del XVII Congreso Argentino de Ciencias de la Computación (CACIC 2011)*, La Plata (Argentina), 2011, ISBN: 978-950-34-0756-1. Págs: 231-240.
  26. Zoltan J., Kacsuk P., Kranzlmuller D., "Distributed and Parallel Systems: Cluster and Grid Computing". *The International Series in Engineering and Computer Science*. Springer; 1st ed., 2004.
  27. Bertogna M. L. "Planificación dinámica sobre entornos Grid". Ph.D. thesis, Universidad Nacional de La Plata, La Plata, Argentina, 2010.
  28. Grid Computing and Distributed Systems (GRIDS) Laboratory - Department of Computer Science and Software Engineering (University of Melbourne). "Cluster and Grid Computing". 2007. <http://www.cs.mu.oz.au/678/>.
  29. Grid Computing Infocentre: <http://www.gridcomputing.com/>
  30. Dikaikos M. et al. "Distributed InterNet Computing for IT and Scientific Research". *Internet Computing IEEE*. Vol 13, Nro. 5, pp 10-13
  31. Ardissono L., Goy A., Petrone G., Segnan M. "From Service Clouds to User-centric Personal Clouds". 2009 *IEEE Second International Conference on Cloud Computing*.
  32. Hemsoth N. "Outsourcing Versus Federation: Ian Foster on Grid and Cloud". 15 de Junio, 2010. <http://www.hpcinthecloud.com/blogs/Outsourcing-Versus-Federation-Ian-Foster-on-Grid-and-Cloud-96326829.html>. Noviembre, 2010.
  33. Song Y., Kalogeropoulos S., Tirumalai P. "Design and Implementation of a Compiler Framework for Helper Threading on Multi-core Processors". *Proceedings of the 14th International Conference on Parallel Architectures and Compilation Techniques*; Sept. 2005.
  34. Vázquez Blanco C., Huedo E., Montero R. S., Llorente I. M. "Elastic Management of Cluster-based Services in the Cloud". *Proceedings pp. 19-24, ACM Digital Library 2009*. ISBN 978-1-60558-564-2.
  35. Vázquez Blanco C., Huedo E., Montero R. S., Llorente I. M. "Dynamic Provision of Computing Resources from Grid Infrastructures and Cloud Providers". *IEEE Society Press*, pp.113-120, *Workshops at the Grid and Pervasive Computing Conference, GPC 2009*. ISBN 978-0-7695-3677-4.
  36. De Giusti L., Naiouf M., Chichizola F., Luque E., De Giusti A. E. "Dynamic Scheduling in Heterogeneous Multiprocessor Architectures. Effi-

- ciency Analysis". Computer Science & Technology Series – XV Argentine Congress of Computer Science Selected Papers. Editores: Guillermo Simari, Patricia Pesado, José Paganini. Págs. 85-95. ISBN 978-950-34-0684-7. Editorial de la Universidad de La Plata (edulp). La Plata (Argentina). 2010.
37. Corredor Franco J. "Predicción de perfiles de comportamiento de aplicaciones científicas en nodos multicore". Ph.D. Thesis, Universidad Autónoma de Barcelona, Barcelona, España, Julio 2011.
  38. Lu S., Tucek J., Qin F., Zhou Y. "AVIO: detecting atomicity violations via access interleaving invariants". SIGPLAN Not, ACM, 2006, 41, 37-48.
  39. Golander A., Weiss S., Ronen R. "Synchronizing Redundant Cores in a Dynamic DMR Multicore Architecture". IEEE Transactions on Circuits and Systems II: Express Briefs Volume 56, Issue 6, 474-478. 2009.
  40. Muresano Cáceres R. "Metodología para la aplicación eficiente de aplicaciones SPMD en clústers con procesadores multicore" Ph.D. Thesis, Universidad Autónoma de Barcelona, Barcelona, España, Julio 2011.
  41. Fialho L. "Fault Tolerance configuration for uncoordinated checkpoints". Ph.D. Thesis, Universidad Autónoma de Barcelona, Barcelona, España, Julio 2011.
  42. Sáenz J. C. "Planificación de Procesos en Sistemas Multicore Asimétricos". Ph.D. Thesis, Universidad Complutense de Madrid, Madrid, España, Diciembre 2010.
  43. Annamalai A., Rodrigues R., Koren I., Kundu S., "Dynamic Thread Scheduling in Asymmetric Multicores to Maximize Performance-per-Watt," 2012 IEEE 26th International Parallel and Distributed Processing Symposium Workshops & PhD Forum, pp. 964-971, 2012 IEEE 26th International Parallel and Distributed Processing Symposium Workshops & PhD Forum, 2012.
  44. Kindratenko, V.V.; Enos, J.J.; Guochun Shi; Showerman, M.T.; Arnold, G.W.; Stone, J.E.; Phillips, J.C.; Wen-Mei Hwu, "GPU clusters for high-performance computing," Cluster Computing and Workshops, 2009. CLUSTER '09. IEEE International Conference on , vol., no., pp.1,8, Aug. 31 2009-Sept. 4 2009
  45. Kindratenko, V.V.; Enos, J.J.; Guochun Shi; Showerman, M.T.; Arnold, G.W.; Stone, J.E.; Phillips, J.C.; Wen-Mei Hwu, "GPU clusters for high-performance computing," Cluster Computing and Workshops, 2009. CLUSTER '09. IEEE International Conference on , vol., no., pp.1,8, Aug. 31 2009-Sept. 4 2009.
  46. Sinha, R.; Prakash, A.; Patel, H.D., "Parallel simulation of mixed-abstraction SystemC models on GPUs and multicore CPUs," Design Automation Conference (ASP-DAC), 2012 17th Asia and South Pacific , vol., no., pp.455,460, Jan. 30 2012-Feb. 2 2012.
  47. Lingyuan Wang, Miaoqing Huang, and Tarek El-Ghazawi. "Towards efficient GPU sharing on multicore processors". In Proceedings of the second international workshop on Performance modeling, benchmarking and simulation of high performance computing systems (PMBS '11). ACM, New York, NY, USA, 23-24.
  48. Montes de Oca E., De Giusti L., De Giusti A., Naiouf M. "Comparación del uso de GPU y cluster de multicore en problemas con alta demanda computacional". XII Workshop de Procesamiento Distribuido y Paralelo. CACIC2012. ISBN: 978987-1648-34-4. Pág. 267-275. Bahía Blanca, Buenos Aires, Argentina, Octubre 2012.
  49. Montes de Oca E., Naiouf M., De Giusti L., Chichizola F., Giacomantone J., De Giusti A. "Una implementación paralela de las Transformadas DCT y DST en GPU. Análisis de performance". XII Workshop de Procesamiento Distribuido y Paralelo. CACIC2012. ISBN: 978987-1648-34-4. Pág. 276-285. Bahía Blanca, Buenos Aires, Argentina, Octubre 2012.
  50. Frati F. E., Olcoz Herrero K., Piñuel Moreno L., Naiouf M., De Giusti A. "Detección de interleavings no serializables usando contadores de hardware". XII Workshop de Procesamiento Distribuido y Paralelo. CACIC 2012. ISBN: 978987-1648-34-4. Pág. 327-336. Bahía Blanca, Buenos Aires, Argentina, Octubre 2012.
  51. Pettoruti J. E., Rodriguez I. P., Chichizola F., De Giusti A. E. "Análisis de la degradación de las comunicaciones en algoritmos de cómputo científico en un Cloud privado". XII Workshop de Procesamiento Distribuido y Paralelo. CACIC 2012. ISBN: 978987-1648-34-4. Pág. 286-295. Bahía Blanca, Buenos Aires, Argentina, Octubre 2012.
  52. Balladini J., Rucci E., De Giusti A., Naiouf M., Suppi R., Rexachs D., Luque E. "Power Characterisation of Shared-Memory HPC Systems". XII Workshop de Procesamiento Distribuido y Paralelo. CACIC 2012. ISBN: 978987-1648-34-4. Pág. 316-326. Bahía Blanca, Buenos Aires, Argentina, Octubre 2012.
  - Casanova B., Balladini J., De Giusti A., Suppi R., Rexachs D., Luque E. "Mejora de la eficiencia energética en sistemas de computación de altas prestaciones". XII Workshop de Procesamiento Distribuido y Paralelo. CACIC 2012. ISBN: 978987-1648-34-4. Pág. 377-386. Bahía Blanca, Buenos Aires, Argentina, Octubre 2012.

## Tendencias en Arquitecturas y Algoritmos Paralelos.

Marcelo Naiouf, Franco Chichizola, Laura De Giusti, Diego Montezanti, Enzo Rucci, Emmanuel Frati, Adrián Pousa, Fabiana Leibovich, Diego Encinas, Horacio Villagarcía, Fernando Romero, Erica Montes de Oca, Javier Balladini, Armando De Giusti

**Instituto de Investigación en Informática LIDI (III-LIDI) - Facultad de Informática - UNLP**

{mnaiouf,francoch,ldgiusti,dmontezanti,erucci,fefrati,apousa,fleibovich,dencinas,hvw,fromero,emontesdeoca,degusti}@lidi.info.unlp.edu.ar; javier.balladini@gmail.com

Con la colaboración en la dirección de Tesis de Posgrado de la Universidad Autónoma de Barcelona (España) y la Universidad Complutense de Madrid (España)

### CONTEXTO

Se presenta una línea de Investigación que es parte de los Proyectos 11/F010 “Arquitecturas multiprocesador distribuidas. Modelos, Software de Base y Aplicaciones” y 11/F011 “Procesamiento paralelo y distribuido. Fundamentos y aplicaciones en Sistemas Inteligentes y Tratamiento de imágenes y video” del III-LIDI acreditados por el Ministerio de Educación.

Asimismo los proyectos “Eficiencia energética en Sistemas Paralelos” y “Algoritmos Paralelos utilizando GPGPUs. Análisis de rendimiento” financiados por la Facultad de Informática de la UNLP.

En el tema hay cooperación con varias Universidades de Argentina y se está trabajando con Universidades de América Latina y Europa en proyectos financiados por CyTED, AECID y la OEI (Organización de Estados Iberoamericanos).

Se participa en iniciativas como el Programa IberoTIC de intercambio de Profesores y Alumnos de Doctorado en el área de Informática.

Por otra parte, se tiene financiamiento de Telefónica de Argentina en Becas de grado y posgrado. Asimismo el III-LIDI forma parte del Sistema Nacional de Cómputo de Alto Desempeño (SNCAD) del MINCYT.

### RESUMEN

El eje de esta línea de I/D lo constituye el estudio de tendencias actuales en las áreas de arquitecturas y algoritmos paralelos.

Incluye como temas centrales:

- Arquitecturas híbridas (diferentes combinaciones de multicores y GPUs) y Arquitecturas heterogéneas.
- Lenguajes y Estructuras de Datos para nuevas arquitecturas de cómputo paralelo.
- Desarrollo y evaluación de algoritmos paralelos sobre nuevas arquitecturas y su evaluación de rendimiento.
- Aspectos del consumo energético, en particular en relación con clases de instrucciones y algoritmos paralelos.

- Empleo de contadores de hardware, en particular en toma de decisiones en tiempo de ejecución.

Las temáticas mencionadas se observan como aristas promisorias en el futuro del cómputo paralelo de altas prestaciones.

*Palabras clave: Sistemas Paralelos. Multicores y GPUs. Clusters híbridos. Algoritmos y lenguajes en clusters heterogéneos/híbridos. Performance y eficiencia energética. Contadores de hardware.*

### INTRODUCCION

Una de las áreas de creciente interés lo constituye el cómputo de altas prestaciones, en el cual el rendimiento está relacionado con dos aspectos: por un lado las arquitecturas de soporte y por el otro los algoritmos que hacen uso de las mismas.

A la aparición de arquitecturas con múltiples núcleos (como los procesadores multicore o los many-core como las GPU), se han sumado en el último tiempo combinaciones de estos, dando lugar a plataformas híbridas con diferentes características. [CHA2011][LIN2011].

Lógicamente, esto trae aparejado una revisión de los conceptos del diseño de algoritmos paralelos (incluyendo los lenguajes mismos de programación y el software de base), así como la evaluación de las soluciones que éstos implementan.

Las estrategias de distribución de datos y procesos necesitan ser investigada a fin de optimizar la performance.

Además de las evaluaciones clásicas de rendimiento prestacional como el Speed-up y la Eficiencia, otros aspectos comienzan a ser de interés, tales como el estudio del consumo y la Eficiencia energética de tales sistemas paralelos. [CAS2012].



Una herramienta que ha comenzado a ser utilizada para la monitorización y evaluación de soluciones paralelas en tiempo de ejecución son los contadores de hardware, que (en sus diferentes variantes) permiten detectar fallas de concurrencia o estimar problemas de consumo instantáneo o decidir una migración de datos o procesos. [BOR2005] [SPR2002].

En esta línea de I/D se trabaja sobre estos aspectos que marcan tendencias en el área.

### Algunos conceptos

Un procesador multicore integra dos o más núcleos computacionales dentro de un mismo "chip", con el objetivo de incrementar el rendimiento reduciendo el consumo de energía en cada núcleo.

Una GPU (Graphics Processing Unit) es una arquitectura dedicada a procesamiento gráfico, con un gran número de núcleos simples (suele mencionársela como many-core). En la actualidad se utiliza su potencia de cómputo en aplicaciones de propósito general logrando un alto rendimiento y dando lugar al concepto de GPGPU (General-Purpose Computing on Graphics Processing Units). [KIR2010][PIC2011].

Un cluster es un sistema de cómputo de procesamiento paralelo compuesto por un conjunto de máquinas (o nodos) interconectadas por una red, que al cooperar configuran un recurso que se ve como "único e integrado", más allá de la distribución física de sus componentes. En particular, cada nodo puede estar compuesto por diferente hardware y sistema operativo (inclusive puede ser un multiprocesador). [CHA2007][SUR2007]

### Multicores

Respecto de las arquitecturas multicore, se puede realizar la siguiente clasificación según las características de sus núcleos:

- Arquitecturas Multicores Homogéneas: Todos sus cores poseen las mismas características.
- Arquitecturas Multicores Heterogéneas: Poseen cores con diferentes características en particular con distinto ISA (instruction set architecture). La desventaja principal es que el código fuente de un programa debe ser compilado para cada tipo de core.
- Arquitecturas Multicores Asimétricas: Todos sus cores usan el mismo ISA (instruction set architecture), de manera que no es necesario compilar el código fuente para cada tipo core, pero los cores difieren en el rendimiento y tienen diferentes características de energía.

Un multicore asimétrico (AMP) típico consiste de algunos cores potentes y complejos, llamados *Fast* o *Big* cores, con características como alta frecuencia de

clock, pipeline fuera de orden, predicción de saltos y alto consumo de energía; y un gran número de cores simples, llamados *Slow* o *Small* cores, con características como baja frecuencia de clock, un pipeline simple y bajo consumo de energía. [SAE2010] [SAE2011A]

Los Fast cores son apropiados para aplicaciones secuenciales o partes secuenciales de aplicaciones paralelas de uso intensivo de CPU que realmente requieren de las características de estos cores; por otro lado, los Slow cores son adecuados en aplicaciones paralelas altamente escalables o para aplicaciones que no necesiten de las características de un Fast core como por ejemplo las intensivas en memoria o en entrada-salida.

Tener cores de diferentes tipos permite optimizar el rendimiento, además de realizar especialización distribuyendo las tareas eficientemente y logrando mayor eficiencia en la relación rendimiento/energía. Las aplicaciones paralelas tienden a ser mucho más eficientes si se ejecutan sobre muchos cores simples que sobre pocos cores más potentes y complejos que cubren un área similar y consumen la misma energía; a su vez, una aplicación paralela puede tener fases secuenciales que necesitan de la potencia de los Fast cores, en este caso la utilización de un Fast core permitiría acelerar la fase secuencial. [KOU2010].

La tarea de distribuir eficientemente los hilos de las aplicaciones es parte del planificador de tareas (scheduler) del sistema operativo, los sistemas operativos actuales no poseen la capacidad de hacer planificación de tareas en arquitecturas asimétricas. Trabajos previos agregaron nuevos planificadores de tareas en el kernel del sistema operativo Solaris, estos planificadores incluyen características que permiten determinar cuándo un hilo se debe ejecutar en un Fast core y cuando en un Slow core. [SAE2009] [SAE2011B] [FED2009A] [FED2009B]

Para determinar si un hilo debe ser migrado de un tipo de core a otro se utiliza información de los procesos dentro del sistema operativo y los contadores hardware que posee cada arquitectura. Actualmente se está trabajando en migrar los planificadores en el sistema operativo Solaris al sistema operativo Linux.

### GPU

En 2007 Nvidia lanzó la implementación de CUDA, plataforma hardware y software que extiende al lenguaje C con un reducido conjunto de abstracciones para la programación paralela de propósito general en GPU. La arquitectura de una GPU-Nvidia CUDA está organizada en Streams Multiprocessors (SMs), los cuales tienen un determinado número de Streams Processors (SPs) que comparten la lógica de control y la caché de instrucciones.

La organización de los threads puede verse por nivel o jerárquicamente de la siguiente manera: grilla o grid (formado por bloques de threads), bloque (conformado por threads) y threads. En un bloque, los hilos están organizados en warps (en general de 32 hilos). Todos los hilos de un warp son planificados juntos durante la ejecución.

El sistema de memoria de la GPU está compuesto por distintos tipos de memorias que se diferencian tanto en la forma de acceso de los hilos, las limitaciones en las operaciones permitidas y su ubicación dentro del dispositivo. Está compuesto por: Memoria Global, Memoria de Textura, Memoria Constante, Memoria Shared o compartida, Memoria Local y Registro.

CUDA permite definir una función denominada kernel que es ejecutada  $n$  veces, siendo  $n$  la cantidad de threads de la aplicación. Dicha función solo corre en la GPU o Device, y puede invocar código secuencial para ser ejecutado en la CPU. El kernel debe ser configurado con el número de hilos por bloque y por grid. Como el kernel se ejecuta en el device, la memoria debe ser alocada antes de que la función sea invocada y los datos que requiera utilizar copiados desde la memoria del host a la memoria del device. Una vez ejecutada la función kernel, los datos de la memoria del device deben ser copiados a la memoria del host. [NVI2008]

Actualmente, ha comenzado a utilizarse OpenCL como lenguaje para las GPU, ya que el mismo provee una abstracción respecto de la plataforma hardware utilizada. Sin embargo, si el objetivo final es obtener alta performance, su utilización debe ser estudiada cuidadosamente. [NOT2009]

### Multicores y GPU

Estas dos plataformas brindan un vasto conjunto de posibilidades de investigación en arquitecturas híbridas, a partir de diferentes combinaciones a saber:

- cluster de máquinas uncore cada una con placa GPU, lo que permite combinar herramientas de programación paralela como MPI-CUDA
- máquinas multicores con más de una GPU, que combinan herramientas de programación paralela como OpenMP – CUDA o Pthread – CUDA.
- cluster de máquinas multicores cada una con placa de GPU, lo que permite combinar OpenMP-MPI-CUDA o Pthread-MPI-CUDA

Los desafíos que se plantean son múltiples, sobre todo en lo referido a distribución de datos y procesos en tales arquitecturas híbridas a fin de optimizar el rendimiento de las soluciones.

### El tema del consumo

Un aspecto de interés creciente en la informática actual, principalmente a partir de las plataformas con gran cantidad de procesadores, es el del consumo energético que los mismos producen.

Muchos esfuerzos están orientados a tratar el consumo como eje de I/D, como métrica de evaluación, y también a la necesidad de metodologías para medirlo.

Entre los puntos de interés pueden mencionarse:

- ✓ Caracterización energética de las instrucciones, tanto sobre procesadores multicores como GPUs.
- ✓ Caracterización de algoritmos complejos y de sistemas paralelos, desde el punto de vista energético (potencia máxima y consumo total).
- ✓ Diseño de microbenchmarks utilizables desde el punto de vista energético, para estudiar patrones en algoritmos de HPC.
- ✓ Modelos de predicción de performance energética.
- ✓ Análisis de esquemas de distribución de procesos entre procesadores, con ajuste dinámico de la frecuencia de clock en función del consumo.

### Contadores de hardware

Todos los procesadores actuales poseen un conjunto de registros especiales denominados contadores de hardware [SPR2002]. Estos registros se pueden programar para contar el número de veces que ocurre un evento dentro del procesador durante la ejecución de una aplicación. Tales eventos pueden proveer información sobre diferentes aspectos de la ejecución de un programa (por ejemplo el número de instrucciones ejecutadas, la cantidad de fallos cache en L1, cuántas operaciones en punto flotante se ejecutaron, etc). Los procesadores actuales poseen una gran cantidad de eventos (más de 300) y la capacidad de usar hasta 11 registros simultáneamente [INT2012]. El acceso a estos recursos de monitorización se puede llevar a cabo usando diferentes herramientas en función del nivel de abstracción deseado: por ejemplo, linux provee una aplicación de usuario llamada perf [TUT2013] que puede ser utilizada para monitorizar eventos generados durante la ejecución de un programa sin necesidad de recompilar el código; PAPI [GAR2000] es una API que permite al programador obtener información sobre segmentos específicos de sus programas.

Interesa plantear la utilización de contadores de hardware en dos líneas:

- Optimización de programas. Los procesadores actuales en lugar de escalar su performance a través del incremento de su frecuencia, lo hacen duplicando recursos. Problemas como el exceso de consumo y calor debido a elevadas frecuencias de cómputo, la latencia del sistema de memoria y la ineficiencia del modelo de Von Neumann en procesadores muy rápidos que impulsaron a la industria en esta dirección siguen igual de vigentes. La dependencia que tiene un programa paralelo sobre su arquitectura para ser altamente eficiente requiere comprender los motivos que penalizan su desempeño. Los contadores hardware permiten acceder a información precisa sobre aspectos específicos de la ejecución de los programas, ayudando al programador en la tarea de encontrar esos motivos y comparar con datos concretos los beneficios de los cambios que realiza [TIN2013].

- Sintonización dinámica de aplicaciones. La información que proporciona la PMU (Unidad de Monitorización de Performance) de los procesadores puede ser utilizada para modificar el comportamiento o tomar decisiones en tiempo de ejecución. Esto permite construir algoritmos dinámicos de gran precisión, que se ajustan a los eventos que ocurren en el hardware, como por ejemplo algoritmos de planificación o balance de carga. En esta línea se está desarrollando una tesis de doctorado sobre optimización de herramientas de detección de errores de concurrencia [FRA2011] [FRA2012A] [FRA2012B], centrada en activar/desactivar herramientas de monitorización en función de los eventos que genera el protocolo de coherencia cache de los procesadores actuales.

### LINEAS DE INVESTIGACION y DESARROLLO

- Arquitecturas híbridas (diferentes combinaciones de multicores y GPUs). Diseño de algoritmos paralelos sobre las mismas.
- Arquitecturas heterogéneas. Optimización de algoritmos en función de la potencia de cómputo.
- Lenguajes y Estructuras de Datos para nuevas arquitecturas de cómputo paralelo.
- Consumo energético, en particular en relación con clases de instrucciones y algoritmos paralelos. Modelos y predicción de performance energética en sistemas paralelos.
- Contadores de hardware. Aplicaciones en la optimización de aplicaciones paralelas.

### Investigación experimental

- Desarrollo y evaluación de algoritmos paralelos sobre nuevas arquitecturas y análisis de rendimiento y consumo.

- Utilización combinada de cluster de multicores y cluster de GPUs.
- Empleo experimental de contadores de hardware, orientados a la detección de fallas de concurrencia y a la toma de decisiones sobre la frecuencia de clock de los procesadores en función del consumo.

### FORMACION DE RECURSOS HUMANOS

Dentro de la temática de la línea de I/D se concluyeron 2 trabajos de Especialización y se encuentran en curso 6 tesis de Doctorado en Ciencias Informáticas y 3 tesinas de grado de Licenciatura.

Además, se participa en el dictado de las carreras de Doctorado en Ciencias Informáticas, y Magíster y Especialización en Cómputo de Altas Prestaciones de la Facultad de Informática de la UNLP (todas acreditadas A por la Coneau), por lo que potencialmente pueden generarse nuevas Tesis de Doctorado y Maestría y Trabajos Finales de Especialización.

Existe cooperación con grupos de otras Universidades del país y del exterior, y hay tesis de diferentes Universidades realizando su Tesis con el equipo del proyecto.

### BIBLIOGRAFIA

[BOR2005] S. Y. Borkar, P. Dubey, K. C. Kahn, D. J. Kuck, H. Mulder, S. S. Pawlowski, y J. R. Rattner, «Platform 2015: Intel® Processor and Platform Evolution for the Next Decade», Intel Corporation, White Paper, 2005.

[CAS2012] Casanova B., Balladini J., De Giusti A., Suppi R., Rexachs D., Luque E.. “Mejora de la eficiencia energética en sistemas de computación de altas prestaciones”. XII Workshop de Procesamiento Distribuido y Paralelo. CACIC 2012. ISBN: 978987-1648-34-4. Pág. 377-386. Bahía Blanca, Buenos Aires, Argentina, Octubre 2012.

[CHA2007] Chapman B., The Multicore Programming Challenge, Advanced Parallel Processing Technologies; 7th International Symposium, (7th APPT'07), Lecture Notes in Computer Science (LNCS), Vol. 4847, p. 3, Springer-Verlag (New York), November 2007.

[CHA2011] Chao-Tung Yang, Chih-Lin Huang, Cheng-Fang Lin, “Hybrid CUDA, OpenMP, and MPI parallel programming on multicore GPU Clusters”, Computer Physics Communications 182 (2011) 266–269, Elsevier.

[FED2009A] Alexandra Fedorova, Juan Carlos Saez, Daniel Shelepov and Manuel Prieto. Maximizing Power Efficiency with Asymmetric Multicore Systems. Communications of the ACM, Vol. 52 (12), pp 48-57. December 2009.

- [FED2009B] Alexandra Fedorova, Juan Carlos Saez, Daniel Shelepov and Manuel Prieto. Maximizing Power Efficiency with Asymmetric Multicore Systems. ACM Queue, Vol. 7(10), pp. 30-45. November 2009.
- [FRA2011] F. E. Frati, K. Olcoz Herrero, L. P. Moreno, D. M. Montezanti, M. Naiouf, y A. De Giusti, «Optimización de herramientas de monitoreo de errores de concurrencia a través de contadores de hardware», in Proceedings del XVII Congreso Argentino de Ciencia de la Computación, La Plata, 2011, vol. XVII, p. 10.
- [FRA2012A] F. E. Frati, K. Olcoz Herrero, L. Piñuel Moreno, M. R. Naiouf, y A. De Giusti, «Unserializable Interleaving Detection using Hardware Counters», in *Proceedings of the IASTED International Conference Parallel and Distributed Computing and Systems*, Las Vegas, USA, 2012, pp. 230-236.
- [FRA2012B] F. E. Frati, K. Olcoz Herrero, L. Piñuel Moreno, M. Naiouf, y A. E. De Giusti, «Detección de interleavings no serializables usando contadores de hardware», presented at the XVIII Congreso Argentino de Ciencias de la Computación, 2012.
- [GAR2000] B. D. Garner, S. Browne, J. Dongarra, N. Garner, G. Ho, y P. Mucci, «A Portable Programming Interface for Performance Evaluation on Modern Processors», *The International Journal of High Performance Computing Applications*, vol. 14, pp. 189–204, 2000.
- [INT2012] Intel® 64 and IA-32 Architectures Software Developer's Manual, Intel Corporation, Manual 253669-043US, may 2012. Recuperado a partir de <http://www.intel.com/content/www/us/en/processors/architectures-software-developer-manuals.html>
- [KIR2010] Kirk D., Hwu W. "Programming Massively Parallel Processors: A Hands-on Approach (Applications of GPU Computing Series). Morgan-Kaufmann. 2010.
- [KOU2010] David Koufaty, Dheeraj Reddy, and Scott Hahn. Bias Scheduling in Heterogeneous Multi-core Architectures. In Proc. of Eurosys '10, 2010.
- [LIN2011] Lingyuan Wang, Miaoqing Huang, Vikram K. Narayana, Tarek El-Ghazawi, "Scaling Scientific Applications on Clusters of Hybrid"Multicore/GPU Nodes". CF '11 Proceedings of the 8th ACM International Conference on Computing Frontiers. USA 2011.
- [NOT2009] Nottingham A. y Irwin B. "Gpu packet classification using openssl: a consideration of viable classification methods". In Research Conf. of the South African Inst. of Comp. Sc. and Inf. Technologists. ACM, 2009.
- [NVI2008] NVIDIA. "Nvidia cuda compute unified device architecture, programming guide version 2.0". In NVIDIA. 2008.
- [PIC2011] M. F. Piccoli, "Computación de Alto Desempeño utilizando GPU". XV Escuela Internacional de Informática. Editorial Edulp, 2011.
- [SAE2009] Juan Carlos Saez. Planificación de procesos en sistemas Multicore asimétricos. <http://eprints.ucm.es/12668/1/T32909.pdf>
- [SAE2010] Juan Carlos Saez, Manuel Prieto, Alexandra Fedorova and Sergey Blagodurov. A Comprehensive Scheduler for Asymmetric Multicore Systems. In Proc. Of 5th ACM European Conference on Computer Systems (Eurosys '10), pp. 139-152. Paris, France. April 2010.
- [SAE2011A] Juan Carlos Sáez (Universidad Complutense de Madrid), Manuel Prieto (Universidad Complutense de Madrid), Adrian Pousa (Universidad Nacional de La Plata), Alexandra Fedorova (School of Computing Science – Simon Fraser University - Canadá). Explotación de Técnicas de Especialización de Cores para Planificación Eficiente en Procesadores Multicore Asimétricos. XXII Jornadas de paralelismo. Universidad de La Laguna, Tenerife, España.
- [SAE2011B] Juan Carlos Saez, Daniel Shelepov, Alexandra Fedorova and Manuel Prieto. Leveraging Workload Diversity Through OS Scheduling to Maximize Performance on Single-ISA Heterogeneous Multicore Systems", In *Journal of Parallel and Distributed Computing (JPDC)*, Vol. 71, pp. 114-131. January 2011.
- [SPR2002] B. Sprunt, «The basics of performance-monitoring hardware», *IEEE Micro*, vol. 22, n.o 4, pp. 64- 71, ago. 2002.
- [SUR2007] Suresh Siddha, Venkatesh Pallipadi, Asit Mallick. "Process Scheduling Challenges in the Era of Multicore Processors" *Intel Technology Journal*, Vol. 11, Issue 04, November 2007.
- [TIN2013] F. G. Tinetti, S. M. Martin, F. E. Frati, y M. Méndez, «Optimization and Parallelization experiences using hardware performance counters», in Proceedings of the 4th International Supercomputing Conference in México, Manzanillo, Colima, México, 2013, p. 5.
- [TUT2013] Tutorial - Perf Wiki. (s. f.). Recuperado 6 de marzo de 2013, a partir de <https://perf.wiki.kernel.org/index.php/Tutorial>.



# Cloud Computing en HPC

De Giusti Armando E.<sup>1</sup>, Rodríguez Ismael P., Pousa Adrian, Chichizola Franco,  
Pettoruti José, Eguren Sebastián, Naiouf Marcelo

Instituto de Investigación en Informática LIDI (III-LIDI) - Facultad de Informática - UNLP  
{degiusti, ismael, apousa, francoch, josep, seguren, mnaiouf}@lidi.info.unlp.edu.ar

Con la colaboración en la dirección de Tesis de Posgrado de la Universidad Autónoma de Barcelona (España).

## CONTEXTO

Esta línea de Investigación está dentro del proyecto “Arquitecturas Multiprocesador Distribuidas. Modelos, Software de Base y Aplicaciones” acreditado por el Ministerio de Educación y de proyectos específicos apoyados por organismos nacionales e internacionales.

Asimismo los proyectos “Eficiencia energética en Sistemas Paralelos” y “Algoritmos Paralelos utilizando GPGPUs. Análisis de rendimiento” financiados por la Facultad de Informática de la UNLP.

En el tema hay cooperación con varias Universidades de Argentina y se está trabajando con Universidades de América Latina y Europa en proyectos financiados por CyTED, AECID y la OEI (Organización de Estados Iberoamericanos).

Se participa en iniciativas como el Programa IberoTIC de intercambio de Profesores y Alumnos de Doctorado en el área de Informática.

Por otra parte, se tiene financiamiento de CONICET, CIC, SPU, UNLP y Telefónica de Argentina en Becas de grado y posgrado.

El III-LIDI forma parte del Sistema Nacional de Cómputo de Alto Desempeño (SNCAD) del MINCYT.

## RESUMEN

El objetivo principal de la línea de investigación lo constituye el estudio de las arquitecturas tipo Cloud y el desarrollo de software de base y aplicaciones eficientes en Cloud Computing, en particular en el área de cómputo paralelo de altas prestaciones (*HPC*).

Los temas de interés abarcan aspectos fundamentales tales como Modelización de Clouds, Virtualización, Plataformas de software para gestión de sistemas Cloud, Evaluación de performance y Despliegues de entornos para ejecución de algoritmos paralelos, que pueden ser homogéneos o heterogéneos.

Las áreas de experimentación se centran principalmente en el despliegue de un Cloud privado, el análisis de la performance de cómputo y de

comunicación y la migración en vivo de Clusters Virtuales en el Cloud.

Se incorporan temáticas como el análisis del consumo y la eficiencia energética en el Cloud, como así también la detección de fallas con sus respectivos mecanismos de protección ante las mismas.

Es de hacer notar que este proyecto se coordina con otros dos proyectos en curso en el III-LIDI, relacionados con Algoritmos Distribuidos/Paralelos y Sistemas de Software Distribuido

**Palabras clave:** Cloud Computing, Virtualización, Cluster Virtual, Cómputo de Altas Prestaciones, Migración en vivo, Algoritmos paralelos, Tolerancia a fallas, Eficiencia energética.

## INTRODUCCIÓN

Los avances en las tecnologías de virtualización y cómputo distribuido han dado origen al paradigma de Cloud Computing, que se presenta como una alternativa a los tradicionales sistemas de Clusters y Grids para ambientes de HPC [ROD07][DEG07][BER08][KON09].

La virtualización es un componente fundamental para implementar el paradigma Cloud Computing. Esta tecnología provee una abstracción de los recursos de hardware permitiendo ejecutar simultáneamente múltiples instancias de Sistemas Operativos (*S.O.*) en máquinas virtuales (*VMs*) sobre un único hardware físico [GOL74][NUS09].

Cloud Computing, proporciona grandes conjuntos de recursos físicos y lógicos (como ser infraestructura, plataformas de desarrollo, almacenamiento y/o aplicaciones), fácilmente accesibles y utilizables por medio de una interfaz de administración web, con un modelo de arquitectura “virtualizada” [SHA10][XIN12]. Estos recursos son proporcionados como servicios (“as a service”) y pueden ser dinámicamente reconfigurados para adaptarse a una carga de trabajo variable (escalabilidad), logrando una mejor utilización y evitando el sobre o sub dimensionamiento (elasticidad) [VEL09][VAQ09].

Existen tres modelos de despliegue de un sistema de Cloud Computing: Cloud público, Cloud privado y Cloud híbrido [CHE10][NUR09]:

<sup>1</sup>Investigador CONICET

- *Cloud público:* este modelo brinda la posibilidad de acceder inmediatamente a un conjunto de recursos con una mínima inversión, pero a su vez presenta un alto costo acumulado en el tiempo de uso del mismo y la ausencia de garantías sobre la privacidad y seguridad de la información del usuario. Del conjunto de proveedores se destacan Amazon Elastic Cloud Compute (EC2), Microsoft Windows Azure y Rackspace entre otros [EC213][MIC13][RAC13].
- *Cloud privado:* tiene la capacidad de utilizar con mayor eficiencia la infraestructura física. Esto lo logra consolidando la infraestructura por medio de la virtualización. Además, brinda un alto nivel de seguridad sobre los datos sensibles de la organización por medio de políticas y medidas de seguridad establecidas en la Intranet de la organización. Por otro lado, un Cloud privado, presenta una limitación en la escalabilidad ante la presencia de grandes demandas de servicios, dado por la capacidad de los recursos físicos subyacentes. La misma se puede evitar con un modelo de Cloud híbrido.
- *Cloud híbrido:* este modelo permite a las organizaciones obtener los beneficios combinados de un Cloud privado y uno público, logrando así el aumento de la eficiencia en el uso de la infraestructura física, un mayor nivel de seguridad de los datos y una mayor disponibilidad de recursos para atender las grandes demandas de servicios. En la actualidad, existen diversas plataformas de software que permiten desplegar y administrar un Cloud privado, como así también extender el modelo a un Cloud híbrido. Entre los más destacado se mencionan: Eucalyptus, OpenNebula y OpenStack [EUC13][OPE13a][OPE13b].

La utilización de un Cloud en entornos de Cómputo Científico presenta las siguientes características y beneficios [PET11]:

- ✓ *Recursos disponibles bajo demanda en el Cloud:* en el caso de utilizar un Cloud público, no se debe disponer ni afrontar el despliegue, instalación y mantenimiento de los recursos físicos, sólo se debe costear los servicios que se utilizan a medida que se requieren.
- ✓ *Escalabilidad y Elasticidad:* los recursos brindados por un Cloud son altamente escalables, dado que una aplicación o un usuario pueden agregar o reducir dinámicamente sus recursos en respuesta a la variación en la carga de trabajo.
- ✓ *Aprovisionamiento automático de recursos:* la distribución de los recursos se realiza de forma inmediata y automática sin requerir la intervención de personal técnico y/o administrativo. De esta forma, el personal técnico se libera de tareas repetitivas de configuración y atención a demandas, mientras que el usuario accede rápidamente a los recursos.
- ✓ *Autoservicio:* un usuario puede solicitar directamente los recursos que necesita, sirviéndose de los mismos por medio de una interfaz que le brinda control directo sobre el despliegue y configuración de los recursos, evitando así las demoras en el acceso a los mismos.

Más allá de las potenciales características y beneficios que brinda un Cloud, de por sí atractivas, es de gran interés estudiar el despliegue de entornos de ejecución para cómputo paralelo y distribuido (*Clusters Virtuales*), como así también realizar I/D en la portabilidad de las aplicaciones de HPC en el Cloud [DOE11][ROD11].

Los Clusters Virtuales (VC), están conformados por VMs configuradas e interconectadas virtualmente para trabajar en forma conjunta como un recurso de cómputo único e integrado. Cada una tiene asociado un S.O., recursos de almacenamiento, protocolos de comunicación, configuraciones de red y entornos de software para ejecución de algoritmos paralelos [VAZ09][HAC11].

Actualmente, la virtualización es un componente esencial en las actividades de administración de un sistema de Cloud Computing, dado que permite realizar balanceo de carga, ahorro de energía, recuperación tras una falla y un manejo versátil para el mantenimiento del sistema.

Por otro lado, el principal beneficio de la virtualización es la técnica de migración en vivo de una VM de un hardware físico a otro, sin interrumpir el funcionamiento de la VM. Esta técnica ha producido uno de los mayores impactos en el Cloud y el gran desafío está centrado en cómo aprovechar al máximo esta técnica para la migración de VC. En este sentido, interesa realizar I/D en la especificación, optimización y evaluación de técnicas de migración para VC [CLA05].

## LINEAS DE INVESTIGACION y DESARROLLO

- Evaluación de performance en arquitecturas de Cloud público.
- Comparación de plataformas de software para gestión y administración Cloud privados e híbridos para la ejecución de VCs.
- Análisis del overhead introducido por software de administración del Cloud para ejecución de soluciones a problemas de HPC.
- Estudio comparativo de performance en Clusters y VCs para problemas de HPC. Escalabilidad.
- Modelos y predicción de performance en arquitecturas de Cloud. Utilización de programas

sintéticos y “firmas” de programas de HPC que requirieren el uso de Cloud.

- Evaluación de rendimiento de las comunicaciones de red virtualizada.
- Hypervisors y técnicas de virtualización.
- Análisis de las técnicas de migración en vivo de VMs.
- Migración en vivo de VCs homogéneos y heterogéneos.
- Análisis de consumo y eficiencia energética en Clouds.
- Detección y tolerancia a fallos en entornos de VCs.

### RESULTADOS OBTENIDOS/ESPERADOS

- Formar RRHH en los temas del proyecto, incluyendo Tesis de postgrado y Tesinas de grado.
- Estudiar y analizar las plataformas de software que permitan el despliegue y la administración de un Cloud.
- Desplegar Clouds privados que brinden recursos para entornos de ejecución de cómputo paralelo (VCs).
- Analizar el overhead introducido por el sistema gestor del Cloud (*plataforma de software + Hypervisor*) en un entorno de ejecución de cómputo de altas prestaciones para aplicaciones científicas. Comparar Cloud y Cluster Computing, desde el punto de vista de eficiencia y rendimiento.
- Evaluar la performance de las comunicaciones de red virtual.
- Estudiar y determinar el controlador virtual de red que ofrezca las mejores prestaciones de rendimiento para la ejecución de aplicaciones paralelas, caracterizadas por tener un tiempo de comunicación significativo.
- Analizar y comparar las técnicas de migración vivo en VMs, con el fin de implementar migraciones de VCs homogéneos y heterogéneos.
- Iniciar la investigación sobre el consumo y eficiencia energética en el Cloud, planteando diversos escenarios de virtualización.
- Adicionalmente, investigar la detección de posibles fallas en el Cloud y los mecanismos de protección que se puedan implementar.

En el contexto presentado, y señalando que esta línea es nueva en el III-LIDI, pueden mencionarse las siguientes experiencias:

- ✓ Se experimentó el despliegue de un Cloud privado utilizando la plataforma de software de Gestión

Eucalyptus con el Hypervisor KVM, incluyendo dos Cluster heterogéneos, uno Dedicado y otro Virtual, integrados por 32 núcleos físicos de CPU y 116 GB de RAM [PET12a].

- ✓ Se estudió, analizó y evaluó el overhead introducido por el sistema de Cloud en la ejecución de aplicaciones de cómputo científico, como ser la suite de benchmarks NAS y una solución paralela al problema de N-Reinas [PET12a].
- ✓ Se investigó el rendimiento de las comunicaciones de red virtual en el Cloud y se determinó el controlador de red virtual que mejor prestaciones ofrece para algoritmos cuyo tiempo de comunicación es significativo con respecto al de cómputo [PET12b].
- ✓ Actualmente, se estudia la migración en vivo de VMs, con el fin de obtener migraciones en vivo de VCs.
- ✓ Se está organizando para el mes de junio la Ira Jornada de Cloud Computing (JCC 2013) en Argentina, con participación de especialistas académicos del país y del exterior y de empresas con experiencia en Cloud Computing.

### FORMACION DE RECURSOS HUMANOS

En cooperación con Universidades iberoamericanas se ha implementado la Maestría en Cómputo de Altas Prestaciones y se continúa dictando la Especialización en Cómputo de altas Prestaciones y Tecnología GRID.

En esta línea de I/D existe cooperación a nivel nacional e internacional. Hay 2 investigadores realizando la Maestría, 1 realizando la Especialización, y en el 2012 se han finalizado 2 Tesinas de Grado en temas del proyecto.

### BIBLIOGRAFÍA

- [BER08] Bertogna, M., Grosclaude, E., Naiouf, M., De Giusti, A., Luque, E.: “Dynamic on Demand Virtual Clusters in Grids”. 3rd Workshop on Virtualization in High-Performance Cluster and Grid Computing (VHPC 08). España. (2008)
- [CHE10] Chen, X., Wills, G. B., Gilbert, L., Bacigalupo, D.: TecRes Report: Using Cloud for Research: a Technical Review. In: Computing, pp. 1--52. UK. (2010).
- [CLA05] Clark, C., Fraser, K., Hand, S., Hansen, J., Jul, E., Limpach, C., Patt, I., Warfield, A.: “Live migration of virtual machines”. In: Proceedings of the 2<sup>nd</sup> conference on Symposium on Networked Systems Design & Implementation. California, USA. (2005).
- [DEG07] De Giusti, A., Naiouf, M., Pettoruti, J. E., Pousa, A., Rodríguez, I. P., De Giusti, L., Chichizola, F., Ardenghi, J., Bertogna, L., Printista, M.: Parallel Algorithms on Multi-Cluster

- Architectures using GRID Middleware. Experiences in Argentine Universities”. In: Garcia Tobio, J., Doallo Biempica, R., López Cabido, I., Martín Santamaría, M. J., Gómez Tato, A., Touriño Dominguez, J. (eds.), 1st Iberian Grid Infrastructure Conference (IBERGRID) Proceedings, pp. 322—332. Spain. (2007)
- [DOE11] Doelitzcher, F., Held, M., Sulistio, A., Reich, C. Viteraas: Virtual Cluster as a Service. In: 3<sup>rd</sup> IEEE International Conference on Cloud Computing Technology and Science. Atenas, Grecia (2011).
- [EC213] Amazon Elastic Compute Cloud (Amazon EC2). <http://aws.amazon.com/es/ec2/>. Febrero 2013.
- [EUC13] Eucalyptus: Open Source software for building AWS-compatible private and hybrid clouds. <http://www.eucalyptus.com>. Febrero 2013.
- [GOL74] Popek, G.J., Goldberg, R.P.: Formal Requirements for Virtualizable Third Generation Architectures. In: Communications in the ACM, Volume 17, Number 7, pp. 412--421. USA. (1974)
- [HAC11] Hacker, T., Mahadik, K. “Flexible Resource Allocation for Reliable Virtual Cluster Computing.” In: Supercomputing Proceedings (SC11). Seattle, USA (2011).
- [KON09] Kondo, D., Javadi, B., Malecot, P., Cappello, F., Anderson, D. P.: “Cost-benefit analysis of Cloud Computing versus desktop grids”. In: IPDPS '09 Proceedings. IEEE International Symposium on Parallel and Distributed Processing. Washington, USA (2009).
- [MIC13] Microsoft Windows Azure. <http://www.windowsazure.com>. Febrero 2013.
- [NUR09] Nurmi, D., Wolski, R., Grzegorzczak, C., Obertelli, G., Soman, S., Youseff, L., Zagorodnov, D.: The Eucalyptus Open-Source Cloud-Computing System. In: 9th IEEE/ACM International Symposium on Cluster Computing and the Grid (CCGRID '09), pp. 124-131, IEEE Computer Society, Washington (2009).
- [NUS09] Nussbaum, L., Anhalt, F., Olivier, M., Gelas, J.: Linux-based virtualization for HPC clusters. In: Montreal Linux Symposium (2009), pp. 221—234. Canada. (2009).
- [OPE13a] OpenNebula: Open Source Data Center Virtualization. <http://opennebula.org>. Febrero 2013.
- [OPE13b] OpenStack Cloud Software: Open source software for building private and public clouds. <http://www.openstack.org>. Febrero 2013.
- [PET12a] Pettoruti, J., Rodriguez, I.: Cloud Computing en aplicaciones científicas. Arquitectura, configuración y análisis experimental de costo/performance. Tesina de Grado, Facultad de Informática, UNLP. La Plata, Argentina. (2012).
- [PET12b] Pettoruti, J.E., Rodriguez, I., Chichizola, F., De Giusti, A.: Análisis de la degradación de las comunicaciones en algoritmos de cómputo científico en un Cloud privado. In: Proceedings del XII Workshop de Procesamiento Distribuido y Paralelo (WPDP) - XVIII Congreso Argentino de Ciencias de la Computación (CACIC 2012). Bahía Blanca, Argentina. (2012).
- [RAC13] RackSpace Open Cloud Company. <http://www.rackspace.com>. Febrero 2013.
- [ROD07] Rodriguez, I. P., Pousa, A., Pettoruti, J.E., Chichizola, F., Naiouf, M., De Giusti, L., De Giusti, A.: Estudio del overhead en la migración de algoritmos paralelos de cluster y multicluster a GRID. In: XIII° Congreso Argentino de Ciencias de la Computación (CACIC 2007) Proceedings. Argentina. (2007)
- [ROD11] Rodriguez, I., Pettoruti, J.E., Chichizola, F., De Giusti, A.: Despliegue de un Cloud Privado para entornos de cómputo científico. In: Proceedings del XI Workshop de Procesamiento Distribuido y Paralelo (WPDP) - XVII Congreso Argentino de Ciencias de la Computación (CACIC 2011). La Plata, Argentina. (2011).
- [SHA10] Shafer, J.: “I/O virtualization bottlenecks in cloud computing today”. In: Proceedings of the 2<sup>nd</sup> conference on I/O virtualization (VIOV10). California, USA (2010).
- [VAQ09] Vaquero, L. M., Rodero-Merino, L., Caceres, J., Lindner, M.: A Break in the Clouds: Towards a Cloud Definition. In: ACM SIGCOMM Computer Communication Review, Volume 39, Issue 1, pp. 50--55. USA. (2009).
- [VAZ09] Vázquez Blanco, B., Huedo, E., Montero, R. S., Llorente, I. M.: “Elastic Management of Cluster-based Services in the Cloud”. In: Proceedings pp.19-24, ACM Digital Library. ISBN 978-1-60558-564-2. (2009).
- [VEL09] Velte, A.T., Velte, T.J., Elsenpeter, R.: Cloud Computing: A Practical Approach, McGraw Hill Professional. (2009)
- [XIN12] Xing, Y., Zhan, Y.: “Virtualization and Cloud Computing”. In: Proceedings pp.305-312, Springer Link. ISBN 978-3-642-27323-0. (2012).



# Utilizando GPU en la Recuperación de Información Multimedia

Mercedes Barrionuevo, Mariela Lopresti, Natalia Miranda, Fabiana Piccoli, Nora Reyes

LIDIC- Univ. Nacional de San Luis

San Luis, Argentina

{mbarrio, omlopres, ncmiran, mpiccoli}@unsl.edu.ar

## 1. Contexto

Esta propuesta de trabajo se lleva a cabo dentro del proyecto de investigación “*Nuevas Tecnologías para el Tratamiento Integral de Datos Multimedia*” abarcando las líneas de investigación “Computación de Alto Desempeño”, “Procesamiento de Información Multimedia” y “Recuperación de Datos e Información Multimedia”. Dicho proyecto se desarrolla en el marco del Laboratorio de Investigación y Desarrollo en Inteligencia Computacional (LIDIC), de la Facultad de Ciencias Físico, Matemáticas y Naturales de la Universidad Nacional de San Luis.

## 2. Resumen

Este trabajo presenta las distintas líneas de trabajo abordadas considerando la GPU como computadora masivamente paralela de propósito general. Los problemas a tratar están relacionados a las Bases de Datos Métricas, en este momento nos enfocamos en la optimización de la resolución de distintos tipos de consultas.

**Palabras Claves:** Computación de alta performance, Espacios métricos, Recuperación de Información.

## 3. Introducción

Una de las principales herramientas que las computadoras ofrecen es almacenar grandes cantidades de datos organizados, siguiendo un determinado esquema o un modelo de datos que facilite su almacenamiento, recuperación y modificación de una forma rápida y ordenada.

Cualquier conjunto de datos pertenecientes a un mismo contexto y almacenados sistemáticamente para su posterior uso constituye una base

de datos tradicional, en las cuales se busca sólo a partir de una clave. Las bases de datos y repositorios de información multimedia (audio, imagen, video y texto) no pueden ser trabajadas tan eficientemente como en las bases de datos tradicionales debido a que la información multimedia debe ser recuperada por similitud; mientras que en las bases de datos tradicionales esta búsqueda es exacta.

Las consultas que pueden ser de interés en bases de datos no estructuradas se denominan búsquedas por similitud o búsquedas por proximidad. La búsqueda por proximidad es la búsqueda en Bases de Datos de elementos que sean similares o cercanos a un elemento consultado. La similitud es modelada con la función de distancia, la cual satisface entre otras propiedades la desigualdad triangular. El conjunto de elementos u objetos sobre los que se define la función de distancia es llamado *espacio métrico* [3].

El costo computacional de resolver consultas en espacios métricos se mide en la cantidad de evaluaciones de funciones de distancia necesarias para responder a la consulta. Existen diferentes alternativas de optimización, una de ellas es el desarrollo de las soluciones computacionales aplicando técnicas de computación de alto desempeño para tecnologías emergentes como son las GPUs.

El poder computacional asociado a las GPUs (Unidad de Procesamiento Gráfico), el constante avance asociado a su hardware y software, y bajo costo, han constituido una alternativa válida a las supercomputadoras paralelas [1, 17]. Para cierto tipo de aplicaciones, una tarjeta de video puede proporcionar mucho más poder de cómputo que la computadora huésped [14]. Entre las características destacadas se encuentran la arquitectura de computación de alto rendimiento, el alto bandwidth de memoria, los numerosos chips dedicados al procesamiento de datos y la administración de

las unidades de procesamiento.

Si consideramos el desarrollo de software para GPUs, no todo tipo de problemas puede ser resueltos en la GPU, los problemas más adecuados son aquellos que pueden ser implementados con procesamiento de stream y usando memoria limitada. Las aplicaciones más adecuadas son aquellas con cuantioso paralelismo de datos. El uso de la GPU para computaciones de propósito general se conoce como GPGPU (GPU de propósito general) [11, 19, 22].

La programación paralela de aplicaciones en la GPU debe considerar varias diferencias con la programación paralela en computadoras paralelas típicas, las más relevantes son: *El número de unidades de procesamiento*, *Estructura de la memoria CPU-GPU* y *Número de threads paralelos*. Respecto al número de unidades de procesamiento, a diferencia de las soluciones para computadoras masivamente paralelas donde el número de unidades de proceso coincide con el número de unidades de procesamiento (Procesadores o cores), en GPGPU esto no es considerado (Es posible la ejecución de cientos o miles de unidades de proceso en un número inferior de cores). En el caso de la estructura de la memoria del sistema CPU-GPU, existen dos espacios de memoria diferentes: la memoria del host (Compartida por todos los hilos de la CPU durante la aplicación) y la memoria de la GPU. Como los hilos o threads de la GPU ejecutan en un espacio de memoria separado a los threads del host en la aplicación, las transferencias de código y datos son necesarios entre el host. Finalmente, la programación de las GPU tiene la posibilidad de iniciar un gran número de threads con poca sobrecarga. Toda GPU tiene mecanismos transparentes y de bajo costo para la creación y administración de threads.

Existen diferentes alternativas para procesamiento de aplicaciones de propósito general en la GPU, la más ampliamente utilizada es la tarjeta Nvidia, para la cual se ha desarrollado un kit de programación en C, con un modelo de comunicación de datos y de control de hilos proporcionado por un driver, quien provee una interfaz GPU-CPU [10]. Este ambiente de desarrollo llamado Compute Unified Device Architecture (CUDA) propone un modelo de programación SIMD (Simple Instrucción, Múltiples Datos) con funcionalidades de procesamiento de vector [5].

Las líneas de investigación proponen utilizar

la GPU como computadora masivamente paralela para resolver mediante técnicas de alto desempeño problemas relacionados a la administración de base de datos métricas.

## 4. Líneas de Investigación y Desarrollo

Actualmente ha surgido la necesidad de crear bases de datos de información no estructurada, como por ejemplo imágenes, texto, sonido y video. Muchas aplicaciones computacionales necesitan buscar eficientemente información sobre estas bases de datos. Para realizar consultas en este tipo de bases de datos se han creado nuevos modelos y algoritmos de búsqueda más generales que los correspondientes a bases de datos tradicionales [3].

A estas nuevas bases de datos se las denomina base de datos métricas y las consultas se realizan según búsquedas por similitud o proximidad. La búsqueda por proximidad es la búsqueda en bases de datos de elementos similares o cercanos al elemento consultado. La similitud es modelada con una función de distancia, la cual satisface las siguientes propiedades. Si  $X$  denota el universo de objetos válidos. Un subconjunto finito de él  $U$ , de tamaño  $n$ , es el conjunto de objetos o base de datos en donde se realizan las búsquedas. La función  $d : X * X \rightarrow \mathbb{R}^+$  denota la medida de distancia entre los objetos. Esta función de distancia debe cumplir las propiedades de positividad  $\forall x, y \in X, d(x, y) \geq 0$ , simetría  $(\forall x, y \in X, d(x, y) = d(y, x))$ , reflexividad  $(\forall x \in X, d(x, x) = 0)$ , desigualdad triangular  $\forall x, y, z \in X, d(x, y) \leq d(x, z) + d(z, y)$  y en la mayoría de los casos la positividad es estricta  $(\forall x, y \in X, x \neq y \Rightarrow d(x, y) > 0)$ . El conjunto de elementos u objetos sobre los que se define la función de distancia es llamado *espacio métrico* [3].

Mientras más pequeña sea la distancia entre los objetos más similares son dichos objetos. Existen tres tipos básicos de consultas por proximidad en espacios métricos:

- Consulta por rango  $(q, r)_d$ : recupera los elementos que están a lo más a distancia  $r$  de  $q$ .
- Consulta por vecino más cercano  $NN(q)$ : recupera el vecino más cercano a  $q$  en  $U$ .
- Consulta por  $k$  vecinos más cercanos

$NN_k(q)$ : recupera los  $k$  vecinos más cercanos a  $q$  en  $U$ .

Las líneas de investigación y desarrollo que actualmente se siguen están realacionadas a la resolución de consultas en espacios métricos. Para ello nos enfocamos en distintas áreas, ellas son:

- *Algoritmos de Ordenamiento*

Cuando se trabaja con datos multimedia se utilizan Base de Datos métricas. En situaciones donde se quiere responder a consultas de estas bases de datos, no tiene sentido la búsqueda por exactitud, la idea es medir la similitud (o diferencia) entre el elemento de consulta dado y cada uno de los objetos de la base de datos. Para determinar la similitud entre los objetos se deben obtener las distancias entre un elemento particular (consulta) y cada objeto de la base de datos. Existen diferentes funciones de distancia, una de ellas es la *Distancia de Edición*, la cual computa el número mínimo de operaciones requeridas para transformar un objeto en otro.

Cuando se deben responder a cualquiera de las consultas planteadas en estas bases de datos es necesario, en la mayoría de los casos, ordenar resultados intermedios o finales. Dicha operación debe ser realizada en forma eficiente, por ello es de vital importancia lograr algoritmos con buen desempeño. Hemos considerado tres algoritmos de ordenamiento: *Quicksort*[9], *QuickSelect*[12] y *Parallel Sort Regular Sampling (PSRS)* [13], todos con diferentes características y conocidos por su buen desempeño en ámbitos secuenciales y paralelos.

Existen bibliotecas estándares incluidas dentro de los repositorios de CUDA, las cuales implementan varios de estos métodos de ordenamiento, una de ellas es *Thrust* [8]. *Thrust* ofrece una colección de algoritmos paralelos tales como *scan*, *sort*, *reduction*, entre otros. En ella, los desarrolladores realizan su computación usando una colección de algoritmos de alto nivel y delegan completamente la decisión de cómo implementar la computación a la biblioteca. Esta interfaz abstracta le permite a los programadores describir qué computar sin colocar ninguna

restricción adicional sobre cómo llevar a cabo la computación.

Una desventaja de *Thrust* es la exposición de sólo un subconjunto de capacidades del hardware debido a las abstracciones para los desarrolladores. Ante esto y el bajo desempeño obtenido en las consultas implementadas, se decidió buscar e implementar alternativas que logran un mejor rendimiento. En ambas propuestas se tiene en cuenta la naturaleza paralela de la GPU y todas sus características para lograr óptimos resultados. Actualmente se ha desarrollado y evaluado *Quicksort* y *QuickSelect*, mientras que el *PSRS* está en las etapas finales de desarrollo.

- *Consultas en Espacios Métricos*

En este punto estamos trabajando en dos direcciones, una se corresponde con la resolución de cada tipo de consulta aplicando el método de resolución de fuerza bruta y diferentes técnicas algorítmicas de computación de alto desempeño en la GPU. Esta línea se nutre de la línea anterior, donde es necesario contar con buenos algoritmos de ordenamiento para resolver las consultas.

La otra línea de investigación dentro de esta área está la optimización a través del uso de índices.

Si tenemos una base de datos de cardinalidad  $n$ , todas estas consultas pueden ser resueltas ejecutando  $n$  evaluaciones de distancia. Generalmente el número de evaluaciones de distancias ejecutadas es la medida de complejidad de los algoritmos. El desafío es diseñar un algoritmo de indexación eficiente que reduzca el número de las evaluaciones de distancia [3, 4, 6, 7, 18, 20, 21, 23].

Los algoritmos de indexación permiten construir a priori un índice, una estructura de datos capaz de ahorrar computaciones de distancias al responder consultas por proximidad.

En general todos los algoritmos de indexación particionan el conjunto  $X$  en subconjuntos  $X_i$ . Se construye un índice para permitir determinar una lista de subconjuntos  $X_i$  de candidatos potenciales a contener objetos relevantes para la consulta.

Actualmente se está trabajando con el índice de *Permutantes*, el cual permite acelerar las consultas y responderlas por aproximación. La inclusión de técnicas de computación de alto desempeño nos permite optimizar aún más, no sólo el proceso de indexación sino también la obtención de las respuestas a las consultas.

En ambas líneas de investigación, las implementaciones han mostrado muy buenos resultados, algunos de los cuales han sido publicados en congresos nacionales e internacionales [15, 16].

## 5. Resultados obtenidos / esperados

En todas las líneas de investigación se han evaluado las soluciones en la GPU y comparado con otras existentes en la literatura. Los resultados obtenidos han mostrado una mejora sustancial en las velocidades para resolver el problema. Además, las mejoras no sólo se reflejaron respecto a los parámetros de evaluación típicos de los sistemas de alto desempeño, sino también en la confiabilidad de las respuestas. Actualmente, se están desarrollando y evaluando otras soluciones y métodos de indexación.

## 6. Formación de Recursos Humanos

Los resultados esperados respecto a la formación de recursos humanos son hasta el momento una tesis de maestría y dos tesis de doctorado en desarrollo y otra en definición.

Además la finalización de beca de postgrado tipo II otorgada por el Consejo Nacional de Investigaciones Científicas (CONICET), obtenida el 01/04/11.

## Referencias

- [1] Buck, I. - GPU computing with NVIDIA CUDA - SIGGRAPH '07 - ACM SIGGRAPH 2007 courses ACM - New York, NY, USA. 2007.
- [2] Bustos Cárdenas, B.E. - Index Structures for Similarity Search in Multimedia Databases - PhD thesis, Universitat Konstanz, Fachbereich Informatik, Germany - Octubre 2006.
- [3] Chávez E., Navarro G., Baeza Yates R.A., Marroquín J.L. - Searching in metric spaces - ACM Comput. Surv. Vol 33 N3 - Pp 273:321 - 2001.
- [4] Chávez, E. and Navarro, G. - A compact space decomposition for effective metric indexing. - Pattern Recognition Letters 26(9) - Pp 1363:1376 - 2005.
- [5] NVIDIA - NVIDIA CUDA Compute Unified Device Architecture, Programming Guide Version 5 - NVIDIA. 2013.
- [6] Dohnal, V., Gennaro, C., Savino, C. and Zuzula, P. - D-index: Distance searching index for metric data sets. - Multimedia Tools and Applications (MTAP), - 21(1): 9:33 - 2003.
- [7] Figueroa Mora, K. - Indexación Efectiva de Espacios Métricos usando Permutaciones. - PhD thesis - Universidad de Chile, Santiago, Chile - 2007. - Director: Dr. G. Navarro y Dr. E. Chávez.
- [8] Hoberock, J. Bell, N. - Thrust: A Parallel Template Library - <http://www.meganeurons.com/>. 2010.
- [9] Hoare, C. - Quicksort - The Computer Journal. Vol 5, N. 1. Pp 10-16. 1962.
- [10] Joselli, M., Zamith, M., Clua, E., Montenegro, A., Conci, A., Leal-Toledo, R. Valente, L., Feijo, B., Dórnellas, M., Pozzer, C - Automatic Dynamic Task Distribution between CPU and GPU for Real-Time Systems - 11th IEEE International Conference on Computational Science and Engineering, 2008 ( CSE '08) - Pp 48:55 - July 2008.
- [11] Kirk D. and Hwu, W. - Programming Massively Parallel Processors, A Hands on Approach - Elsevier, Morgan Kaufmann. ISBN 978-0-12-381472-2. 2010.
- [12] Kuba, M. - On Quickselect, partial sorting and Multiple Quickselect - Inf. Process. Lett. Vol 99 N. 5. Pp 181-186. 2006.
- [13] Li, X., Lu, P., Schaeffer, J., Shillington, J., Wong, P., Shi, H. - On the Versatility of Parallel Sorting by Regular Sampling- Parallel Computing. Vol 19. Pp 1079-1103. 1993.



- [14] Lloyd, D., Boyd, C., Govindaraju, N. - Fast computation of general Fourier Transforms on GPUS - IEEE International Conference on Multimedia and Expo - Pp 5:8 - April 2008.
- [15] Lopresti, M., Miranda, N., Piccoli, F., Reyes, N. - Efficient Similarity Search on Multimedia Databases XVIII Congreso Argentino de Ciencias de la Computación, CACIC 2012. Pp 1079-1088. Bahía Blanca, Argentina. 2012.
- [16] Lopresti, M., Miranda, N., Piccoli, F., Reyes, N. - Solving Multiple Queries through the Permutation Index in GPU. 4th International supercomputing Conference in Mexico. Colima - México. 5-8 March 2013. In Press.
- [17] Luebke, D., Humphreys, G. - How GPUs Work Computer - vol 40, N 2 - Pp 96:100 - ISSN 0018-9162 - Feb. 2007.
- [18] Mamede, M. - Recursive lists of clusters: A dynamic data structure for range queries in metric spaces. In Proc. 20th Intl. Symp. on Computer and Information Sciences (IS-CIS'05), LNCS 3733 - pages 843:853 - 2005.
- [19] Piccoli, M.F. - Computación de alto desempeño en GPU. ISBN: 978-950-34-0759-2. Ed-lup. Octubre 2011.
- [20] Reyes, N. - Indices dinámicos para espacios métricos de alta dimensionalidad. - Master's thesis - Universidad Nacional de San Luis - San Luis, Argentina - 2002. - Director: Dr. G. Navarro.
- [21] Samet, H. - Foundations of Multidimensional and Metric Data Structures (The Morgan Kaufmann Series in Computer Graphics and Geometric Modeling). - Morgan Kaufmann Publishers Inc. - San Francisco, CA, USA - 2005.
- [22] Sanders, J. and Kandrot, E. - CUDA by Example, An Introduction to General Purpose GPU Programming - Addison Wesley. ISBN 978-0-13-138768-3. 2010.
- [23] Zezula, P., Amato, G., Dohnal, V. and Batko, M. - Similarity Search: The Metric Space Approach (Advances in Database Systems). - Springer-Verlag New York, Inc. - Secaucus, NJ, USA - 2005. - XVIII, 220 p., Hardcover - ISBN: 0-387-29146-6.

# Red de nodos sensores y actuadores para control y supervisión de consumo de energía en dispositivos electrodomésticos.

Peretti Gaston<sup>1</sup>, Gallina Sergio Hilario<sup>2</sup>, Felissia Sergio Francisco<sup>1</sup>, Villagrán Luis Daniel<sup>2</sup>, Beltramini Paola<sup>2</sup>, Herrera Martín<sup>2</sup>

- 1) Departamento de Electrónica de Facultad Regional San Francisco de la Universidad Tecnológica Nacional
- 2) Departamento de Electrónica Facultad de Tecnología y Ciencias Aplicadas de la Universidad Nacional de Catamarca

Av. De la Universidad 50- San Francisco (Pcia de Córdoba)– CP (2400)

Tel.: 03564-421147 / e-mails: [gastonperetti@gmail.com](mailto:gastonperetti@gmail.com), [sgallina@tecno.unca.edu.ar](mailto:sgallina@tecno.unca.edu.ar), [sergiofelissia@gmail.com](mailto:sergiofelissia@gmail.com), [ldvillagran@hotmail.com](mailto:ldvillagran@hotmail.com)

## Resumen

En la actualidad existen una gran variedad de sistemas domóticos que utilizan distintos protocolos para controlar los dispositivos hogareños, pero no existen desarrollos que integren la medición de energía en tiempo real de cada dispositivo y el control sobre los mismos al mismo tiempo. El presente proyecto pretende desarrollar una red de nodos sensores y actuadores configurables en tiempo real, que permitan implementar funciones de control y supervisión de consumo energético de artefactos y dispositivos hogareños.

Adicionalmente, teniendo en cuenta la disponibilidad del acceso a la red global, implementar el acceso a la información y a la configuración de los nodos a través de páginas web embebidas en cada uno de ellos, sin la necesidad de un controlador central, tal cual es habitual encontrar en los sistemas domóticos de la actualidad, brindando autonomía e inteligencia a cada uno de ellos.

Se pretende en un paso posterior al proyecto actual el estudio de algoritmos que logren reducir el consumo energético aprovechando la posibilidad y flexibilidad del desarrollo del sistema que se está proponiendo.

**Palabras clave:** Eficiencia energética - Domótica - Control - Automatización

## Contexto

La investigación está inserta dentro de la línea de comunicaciones, control y redes de información. El proyecto de investigación red de nodos sensores y actuadores para control y supervisión de consumo de energía en dispositivos electrodomésticos se lleva a cabo en el ámbito del Departamento de Electrónica de la Facultad Regional San Francisco en conjunto con el Departamento de Tecnología y Ciencias Aplicadas de la Universidad Nacional de Catamarca.

La Institución que acredita el proyecto de Investigación y desarrollo (PID) es Rectorado de la Universidad Tecnológica Nacional mediante la financiación del mismo con fecha Marzo del 2013.

## Introducción

Existe una tendencia a optimizar el consumo de energía eléctrica en las empresas, con determinación del gasto en cada sector o área y por ende también diversas empresas que prestan asesoramiento en este sentido ó que ofrecen soluciones para la supervisión y gestión de la energía [1]. Pero esta misma tendencia de optimización de la energía, no está tan fuertemente instaurada en el consumo eléctrico hogareño. Este es el principal motor de desarrollo del proyecto de investigación, debido a que no se han encontrado desarrollos de domótica que apunten a la posibilidad de conocer la información de consumo de cada elemento o dispositivo conectado a la red

eléctrica en tiempo real, por medio de nodos que formen parte de una red domótica, interconectados a una red global, que permita el control de encendido y apagado de distintos dispositivos hogareños [2].

Para la realización del proyecto se propone considerar un concepto denominado "Internet de las cosas" [3], el cual propone el acceso a través de la red global a todos los elementos y productos utilizados en la vida cotidiana. Desde este punto de vista, otro aporte que esperamos realizar es proponer los aspectos técnicos necesarios para garantizar el acceso a Internet de la red domótica y de cada nodo en particular[4].

Podemos considerar que en la arquitectura de una red de dispositivos para aplicaciones de automatización de entornos residenciales, se distinguen tres capas[5]:

1. Capa de dispositivos hardware: En esta capa se encuentran los sensores que permiten recoger información, y los actuadores que permiten interactuar con el entorno.
2. Capa correspondiente a la plataforma software de base o firmware.
3. Capa de entorno de aplicación.

Nuestro trabajo estará centrado en las capas 1 y 2 de este modelo. Los detalles a tener en cuenta en el desarrollo están basados en los siguientes puntos:

- Costo reducido.
- Sin controlador maestro.
- Hardware reducido y de mínima ocupación de espacio físico.
- Protocolo de comunicación simple y de baja ocupación de ancho de banda.
- Funcionalidades de medición de corrientes, tensiones y potencia.
- Posibilidad de supervisión y control del consumo energético.
- Control de dispositivos hogareños (on/off y control de potencia)

Se propone desarrollar e implementar en una primera fase, la conexión de tres de nodos conectados a través de una red cableada por par trenzado, utilizando una capa física en RS-485, y alternativamente,

conectividad inalámbrica utilizando tecnología zigbee [6].

Se prevee estudiar las necesidades de hardware que se requiere para cada una de las partes del sistema y luego determinar los recursos mínimos para implementar las funcionalidades previstas.

En base a esta información, se seleccionará una plataforma de microcontroladores con la capacidad y características adecuadas para los requerimientos que surgen del estudio realizado, y se seleccionarán los componentes electrónicos para implementar la red prototipo [7]- [8].

Los Tipos de Nodos de la red domótica a desarrollar serán:

- Nodo actuador
- Nodo sensor
- Nodo Gateway

#### **Nodo actuador:**

- Control de encendido/apagado de artefactos de todo tipo.
- Control de potencia (dimmer)
- Medición de valor eficaz real de corriente y tensión. Cálculo de potencia activa y reactiva. Factor de potencia.
- Memoria con capacidad para Registro temporal de valores.
- Reloj de tiempo real sincronizado con la red.
- Tabla de parámetros para configuración.
- Conectividad por par trenzado e inalámbrica.

#### **Nodo sensor:**

- Lectura de entradas digitales (on/off).
- Lectura de entradas analógicas.
- Memoria para Registro temporal de valores.
- Reloj de tiempo real sincronizado con la red.
- Tabla de parámetros para configuración.

- Conectividad por par trenzado e inalámbrica.

#### Nodo gateway:

- Interfaz de acceso a Internet vía Ethernet ó WiFi.
- Fuente de alimentación de los nodos.

Un aspecto muy importante en el desarrollo a realizar es la medición y registro de valores de corrientes, tensiones y potencias, con vistas a la supervisión del consumo energético en cada punto de la red. Se prevé utilizar sensores de corriente integrados, que basan sus mediciones en el efecto Hall, y cuyas características de funcionamiento superan ampliamente a los transformadores de corriente que se utilizan también para este tipo de mediciones.

Se analizarán las ventajas y desventajas de realizar un diseño utilizando Circuitos Integrados con capacidad para realizar mediciones del valor eficaz verdadero (RMS) ó de hacerlo a través de software mediante técnicas de Procesamiento Digital de Señales (DSP), lo cual conlleva el requerimiento de procesadores más potentes y de mayor costo, para realizar los cálculos necesarios. Se evaluará comparativamente todos los aspectos (costo, espacio físico, precisión de las medidas, inmunidad a ruidos, etc.) Se diseñará un esquema experimental de laboratorio para realizar estas comparaciones, utilizando el instrumental adecuado. La metodología utilizada y los resultados obtenidos serán susceptibles de ser publicados y presentados en eventos científicos tecnológicos.

Otro aspecto relevante, es realizar un estudio de las alternativas que se disponen para establecer la comunicación entre nodos de modo que interactúen entre sí y además que dispongan de una dirección para el acceso a la red global por medio del dispositivo que actúe como Gateway.

El aporte del proyecto en cuanto al impacto del mismo sobre la sociedad, es que el desarrollo de una red domótica de este tipo permitiría un hogar de mayor confort y manejo eficiente de la energía eléctrica. La red además de administrar y controlar los dispositivos del hogar permitiría conocer en tiempo real los consumos de dispositivos, las horas picos y el tiempo de funcionamiento. Este dato no solo le proveería al usuario del consumo de energía por artefacto, sino que podría predecir incrementos de consumo, o fallas

difíciles de detectar como por ejemplo el funcionamiento en forma continua de una heladera por algún problema en el automático, cierre de puertas, falta de ventilación, etc.

Además se podrían plantear, en proyectos posteriores al actual, algoritmos de ahorro energético a través del encendido y apagado de los dispositivos, en función de la estación del año, horarios de trabajos, período de vacaciones, etc.

En principio se podría esperar que la combinación de dispositivos eficientes, la automatización y la supervisión facilitarían ahorros importantes en el consumo energético. algo muy apreciado en nuestros días por ser la energía un bien escaso y alto impacto en la economía Argentina.

## Línea de Investigación y Desarrollo

Nuestro proyecto consta de los siguientes ejes:

1. Determinar las necesidades de hardware que se requieren para cada tipo de nodo de la red domótica.
2. Estudio de los dispositivos que pueden ser integrado en la comunicación, en base a la disponibilidad y oferta en Argentina
3. Desarrollo de los nodos e interfaces de comunicación (hardware + Software).
4. Desarrollo de software de control y datos de consumo eléctrico de los diferentes dispositivos (amigable con el usuario).

## Resultados y objetivos

**OBJETIVO GENERAL:** Proponer la utilización de una red para el hogar que aproveche la flexibilidad y la integración de sistemas que nos brindan las tecnologías aplicadas en la automatización y que no sólo considere aspectos de confort, sino que agregue elementos de interés social y comunitario como puede ser el manejo eficiente de los recursos y la energía.

**OBJETIVOS ESPECÍFICOS:** Diseñar una red con nodos sensores y actuadores configurables para el control inteligente de los elementos sensoriales y de



actuación en un hogar, combinándolo con los siguientes aspectos:

- Gestión eficiente de recursos energéticos y confort mediante el relevamiento de la información de consumo energético de cada artefacto electrodoméstico presente en el hogar, sumado a acciones de control como la regulación automática de la intensidad lumínica, la climatización, apertura o cierre de ventanas.
- Gestión de la seguridad en el hogar con algoritmos de presencia como por ejemplo para encender las luces simulando personas en el hogar y disuadir a posibles intrusos.
- Gestión para la prevención de accidentes mediante la monitorización de humo y fugas de gas, agua, etc.
- Gestión de los nodos desde internet ya sea para la activación o desactivación de algún dispositivo o para la extracción de datos en tiempo real de consumo, tiempo de utilización, horario de encendido, etc.

Se desea que la propuesta de diseño de este sistema sea un primer paso para generar lineamientos de utilización de nodos inteligentes para el control y utilización eficiente de la energía en el hogar.

**RESULTADOS:** considerando que el proyecto se encuentra en su fase inicial, los resultados que se esperan, se pueden resumir como:

- ✓ Concretar la experimentación y desarrollo de los nodos de la red domótica.
- ✓ Acceder en tiempo real desde la red global a los nodos, de modo de conocer el consumo eléctrico de los diferentes dispositivos de la red.
- ✓ Controlar el encendido, apagado y la potencia (dimmer) de cada dispositivo conectado a la red desde la red global, o a través de algún algoritmo de ahorro energético.
- ✓ Diseñar un software amigable con el usuario de modo de permitir al mismo elaborar una estrategia de reducción del consumo eléctrico hogareño.
- ✓ Planificar clases prácticas en asignaturas y cursos relacionados con la domótica y la automatización en general a fin de complementar la enseñanza en las carreras de grado de Ingeniería Electrónica e Informática.

- ✓ Realizar transferencia de tecnología a empresas del sector, desde los mismos grupos de investigación en distintas regiones (centro a través de la UTN San Francisco- Córdoba y norte del país a través de la UNCa).

## Formación de Recursos Humanos

El director del proyecto, Mg. Esp. Ing. Peretti Gastón Carlos realizó la Maestría en Ciencias de la Ingeniería, mención Telecomunicaciones en la Universidad Nacional de Córdoba. En la actualidad es docente de la carrera de Ing. Electrónica de la Universidad Tecnológica Nacional Facultad Regional San Francisco, en la cátedra de Dispositivos Electrónicos y posee publicaciones en congresos y libros referidas a las líneas de investigación sistemas de control y comunicaciones. El Codirector del proyecto, Ing. Sergio H. Gallina ha cursado los módulos correspondientes a la Maestría en Ingeniería de software y a la especialidad en Gestión Estratégica de los Servicios de Telecomunicaciones, actualmente trabaja en la tesis de la especialización manteniendo una estrecha relación con otras instituciones tales como el Dpto. Electrónica de la Facultad de Ciencias Exactas de la Universidad Nacional de Córdoba y el Dpto. Electrónica de la UTN Facultad Regional San Francisco. El Ing. Felissia Sergio, se encuentra actualmente cursando la Maestría en Sistemas de Control aplicado en la UTN Facultad Regional Paraná y es docente de la misma universidad en las cátedras de de Informática I, Sistemas de Control Aplicado y Control de Procesos. El grupo de investigación también está conformado por alumnos avanzados de la carrera Ingeniería Electrónica de la Facultad Regional San Francisco y la Universidad Nacional de Catamarca.

El impacto esperado del proyecto se basa en la fundación de un grupo de trabajo interdisciplinario orientado al control y los sistemas de comunicaciones con un objetivo único y general que es el de generar conciencia en el manejo eficiente de los recursos energéticos en el entorno residencial. Respecto al potencial humano que conforma el grupo de trabajo, está formado por docentes y alumnos de la Universidad Nacional de Catamarca y Universidad Tecnológica Nacional Facultad

Regional San Francisco (Pcia. de Córdoba). Además se propone interactuar con grupos de trabajo en el área de otras universidades, principalmente con el Laboratorio de Investigación Matemática Aplicada a Control de la Facultad de Ciencias Exactas Físicas y Naturales (LIMAC-FCEFN de la UNCba).

## Referencias

- [1] Romero Cristobal, Vazquez Francisco, *Domótica e inmótica*, Edición Alfaomega, Madrid, 2011.
- [2] Silveira Rubén, *Automatización de viviendas y edificios*, Edición Ceac, Barcelona, España, 2009.
- [3] Russel Jesse, Cohn Ronald, *Internet 0*, Edición Boskntlen publishin, Estados Unidos, 2012.
- [4] Junstrand Stefan, Passaret Xavier, Vazquez Daniel, *Domótica y Hogar Digital*, Edición Paraninfo, Madrid, 2005.
- [5] Hoya Huidobro, Tejedor Ramón, *Edificios inteligentes*, Creaciones Copyright, Madrid, 2004.
- [6] Zigbee Aliance, [www.zigbee.org](http://www.zigbee.org)
- [7] OSGI Aliance <http://www.osgi.org/>
- [8] Foro UPnP <http://www.upnp.org>.

## Desarrollo e implementación de un sistema informático para el análisis del movimiento humano orientado al estudio de la marcha con procesamiento digital de imágenes.

Sampallo G.M., Cleva M.S., González Thomas A.O., Acosta C.A.  
 Grupo de Investigación Sobre Inteligencia Artificial  
 Facultad Regional Resistencia  
 Universidad Tecnológica Nacional  
 French 414 -H3500CHJ- Resistencia, Chaco, Argentina  
 gsampallo@gmail.com

### Resumen

El objetivo del proyecto es desarrollar e implementar un sistema informático para analizar el movimiento humano con énfasis en la marcha sin requerimiento de marcadores para la posición de los centros articulares en un ambiente no necesariamente de laboratorio con procesamiento digital de imágenes.

El método propuesto se basa en el procesamiento digital de la secuencia de imágenes de la marcha, con el objeto de determinar las sucesivas posiciones y orientaciones de los segmentos que representan las extremidades inferiores del cuerpo.

Se emplearán cámaras web conectadas a una PC, para registrar el movimiento del sujeto desde distintos puntos de vista.

A partir de las imágenes secuenciales de video, se realizará un preprocesamiento de imágenes para extraer la figura humana en cada imagen. Con esta información se implementarán y/o desarrollarán algoritmos para determinar la posición y orientación espacial de los segmentos representativos de las extremidades, con los cuales se obtendrán los parámetros cinemáticos del movimiento.

Los resultados logrados se validarán con datos referenciales obtenidos por los métodos convencionales.

Como aplicación del sistema se prevé emplearlo para la comparación de la marcha (movimiento de extremidades) en diferentes condiciones físicas, y correlacionar los resultados con diagnósticos conocidos.

**Palabras clave:** Procesamiento de imagen. Cinemática. Biomecánica

### Contexto

La línea de I+D presentada es una continuación y profundización de proyectos desarrollados previamente por el grupo. Todos dentro de la temática de procesamiento digital de imágenes y la utilización de Inteligencia Artificial para la

interpretación de los datos obtenidos por ese medio.

### Introducción

El análisis del movimiento es importante para los estudios en biomecánica y se usa habitualmente para el diagnóstico de trastornos musculoesqueléticos y mecánicos del movimiento (1, 2). También se usa para la evaluación de tratamientos de rehabilitación (3).

A pesar de ser ampliamente conocido el uso clínico del análisis de movimiento, su uso rutinario no ha sido adoptado exitosamente, debido principalmente a varios factores como la accesibilidad de la tecnología requerida, y el tiempo que demanda la obtención de los datos experimentales, su procesamiento y posterior interpretación (4).

Un avance importante en el análisis del movimiento sería el desarrollo de un sistema que no esté basado en el uso de marcadores. Un sistema con estas características ayudaría a que el análisis de la marcha se haga más rutinario al eliminar el tiempo de preparación del paciente y no induciría cambios en la marcha natural. Elimina además el problema de la oclusión parcial de los marcadores durante el movimiento cuando se trabaja con una sola cámara.

Si bien el campo de la visión por computadora provee las bases conceptuales y herramientas para llevar a cabo esta tarea, este tipo de sistemas no está ampliamente difundido. Tienen la ventaja de ser no invasivos y permiten la captura del movimiento de los sujetos en su entorno natural.

El campo del análisis del movimiento sin el uso de marcadores tiene otras áreas de aplicación como vigilancia inteligente (smart surveillance), control por identificación (identification control), animación de personajes (character animation), realidad virtual (virtual reality) e interpolación visual (visual interpolation)(5,6).

En estos últimos 20 años, las técnicas de visión por computadora aplicadas a la detección y seguimiento del movimiento humano (human

tracking movement) se han desarrollado continuamente y han sido propuestos numerosos sistemas. Varían en la cantidad de cámaras usadas, la representación de los datos obtenidos, los tipos de algoritmos, los modelos empleados, y el estudio de una parte específica o de todo el cuerpo.

Mundermann y col (7) citan en un review diferentes tipos de algoritmos propuestos para estudiar el movimiento humano: constraint propagation (8), optical flow (9,10), medial axial transformation (11), stochastic propagation (12), search space decomposition based on cues (13), statistical model of background and foreground (14) silhouette contours (15), annealed particle filtering (16), silhouette based techniques (17,18), shape-encoded particle propagation (19) y fuzzy clustering process (20). Moeslund y col (5) hacen una revisión de más de 130 papers relacionados con la captura de movimiento publicados entre 1980 y 2000. Realizan una taxonomía basada en las fases del proceso de determinación de la postura. Wang L y col (6), hacen un estudio de papers publicados entre 1997 y 2001 clasificándolos según el sistema de visión empleado.

Los contenidos teóricos de la técnica de visión por computadora son suficientes para su aplicación al análisis del movimiento sin marcadores. Sin embargo, no hay un desarrollo apropiado en el área que se propone en el proyecto. En general los trabajos y/o aplicaciones desarrolladas están orientados al campo de la vigilancia (surveillance) y con una sola cámara. Estos sistemas sacrifican precisión por rapidez.

El sistema que se pretende desarrollar en este proyecto proveerá de una herramienta de bajo costo, que podrá ser empleada en estudios del movimiento humano en áreas vinculadas con la medicina, y con potenciales aplicaciones en otras como ser: vigilancia, animación de personajes, deportología, etc. También se prevee su empleo como herramienta didáctica.

Permitirá además, la formación de recursos humanos en un área de vacancia a nivel regional, con un enfoque multidisciplinario del problema.

### **Líneas de investigación y desarrollo**

A partir de una secuencia de imágenes obtenida de un video se realizará el seguimiento de las sucesivas posiciones y orientaciones espaciales de los segmentos representativos de las extremidades del sujeto en el movimiento de marcha.

Los ejes temáticos sobre los que se trabajará son básicamente:

1. La segmentación de la imagen. Es un tema sensible porque presenta varios niveles de dificultad:
  - a. Diferenciación entre el fondo y el sujeto.
  - b. Individualización de cada segmento.
  - c. En cada segmento se debe detectar las características propias.
2. Implementar diferentes algoritmos para el seguimiento de los segmentos que constituyen las extremidades sin el uso de marcadores que entorpezcan la marcha
3. Los descriptores de la trayectoria de los segmentos.
4. Estandarizar la forma en que se analiza el movimiento de distintos segmentos articulares
5. Construcción de una base de conocimiento a partir de los descriptores, de modo que cada vez que una nueva secuencia de imágenes se procese, incorpore y actualice los datos en forma eficiente.
6. Construcción de un sistema experto basado en reglas que use la base para realizar una evaluación de la marcha.
7. Generación de informes.

### **Objetivos**

- Desarrollar un sistema experto de visión que no requiera el uso de marcadores para el análisis de la cinemática del movimiento orientado a la marcha humana, que podrá ser utilizado en áreas de la Medicina: Traumatología, Ortopedia, Kinesiología, Medicina Deportiva, etc.

- Que el mismo sea de bajo costo, de implementación sencilla y sin el empleo de una infraestructura compleja.

- Establecer protocolos y criterios de normalización para el proceso de adquisición de los videos y su procesamiento.

- Desarrollo de un software amigable para la adquisición de la secuencia de marcha, que permita presentar los resultados en un formato que facilite su interpretación.

### **Formación de Recursos Humanos**

En la implementación del proyecto se prevee la incorporación de becarios que sean alumnos avanzados de la carrera de Ingeniería en Sistemas de Información, y si fuera posible becarios graduados.

Actualmente integran el equipo de trabajo cuatro investigadores categorizados estables, tres becarios de grado y uno posgrado. La relación con otros grupos del país se realiza a través de presentaciones a Congresos y vinculación a través



de convenios institucionales. Integrantes del grupo han dictado cursos de posgrado acreditados por la UTN y de capacitación en procesamiento digital de imágenes.

### Referencias

- 1 - Andriacchi TP, Alexander EJ: Studies of human locomotion: Past, present and future. *Journal of Biomechanics* 2000, 33(10):1217-1224.
- 2 - Harris GF, Smith PA: *Human Motion Analysis: Current Applications and Future Directions*. New York: IEEE Press; 1996.
- 3 - Mündermann A, Dyrby CO, Hurwitz DE, Sharma L, Andriacchi TP: Potential strategies to reduce medial compartment loading in patients with knee OA of varying severity: Reduced walking speed. *Arthritis and Rheumatism* 2004, 50(4):1172-1178.
- 4 - Simon RS: Quantification of human motion: gait analysis benefits and limitations to its application to clinical problems. *Journal of Biomechanics* 2004, 37:1869-1880.
- 5 - Moeslund G, Granum E: A survey of computer vision-based human motion capture. *Computer Vision and Image Understanding* 2001, 81(3):231-268.
- 6 - Wang L, Hu W, Tan T: Recent Developments in Human Motion Analysis. *Pattern Recognition* 2003, 36(3):585-601.
- 7 - Mündermann L, Corazza S, Andriacchi T: The evolution of methods for the capture of human movement leading to markerless motion capture for biomechanical applications. *Journal of NeuroEngineering and Rehabilitation*. 2006, 3, 6.
- 8 - O'Rourke J, Badler NI: Model-based image analysis of human motion using constraint propagation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1980, 2:522-536.
- 9 - Bregler C, Malik J: Tracking people with twists and exponential maps. *Computer Vision and Pattern Recognition* 1997:568-574.
- 10 - Yamamoto M, Koshikawa K: Human motion analysis based on a robot arm model. *Computer Vision and Pattern Recognition* 1991.
- 11 - Bharatkumar AG, Daigle KE, Pandy MG, Cai Q, Aggarwal JK: Lower limb kinematics of human walking with the medial axis transformation. *Workshop on Motion of Non-Rigid and Articulated Objects: Austin, TX 1994*.
- 12 - Isard M, Blake A: Visual tracking by stochastic propagation of conditional density. *4th European Conference on Computer Vision: Cambridge, UK 1996:343-356*.
- 13 - Gavrilu D, Davis L: 3-D model-based tracking of humans in action: a multi-view approach. *Conference on Computer Vision and Pattern Recognition: San Francisco, CA 1996*.
- 14 - Wren CR, Azarbayejani A, Darrel T, Pentland AP: Pfunder: Realtime tracking of the human body. *Trans on Pattern Analysis and Machine Intelligence* 1997, 19(7):780-785
- 15 - Legrand L, Marzani F, Dusserre L: A marker-free system for the analysis of movement disabilities. *Medinfo* 1998, 9(2):1066-1070.
- 16 - Deutscher J, Blake A, Reid I: Articulated body motion capture by annealed particle filtering. *Computer Vision and Pattern Recognition: Hilton Head, SC 2000*.
- 17 - Bottino A, Laurentini A: A silhouette based technique for the reconstruction of human movement. *Computer Vision and Image Understanding* 2001, 83:79-95.
- 18 - Cheung G, Baker S, Kanade T: Shape-from-silhouette of articulated objects and its use for human body kinematics estimation and motion capture. *IEEE Conference on Computer Vision and Pattern Recognition: Madison, WI: IEEE* 2003:77-84.
- 19 - Moon H, Chellappa R, Rosenfeld A: 3D object tracking using shape-encoded particle propagation. *Intl Conf on Computer Vision: Vancouver, BC 2001*.
- 20 - Marzani F, Calais E, Legrand L: A 3-D marker-free system for the analysis of movement disabilities – an application to the legs. *IEEE Trans Inf Technol Biomed* 2001, 5(1):18-26.

## Metodologías de Diseño para Sistemas Embebidos

**Marcelo A. Tosini, Elias Todorovich, Martín Vázquez, Lucas Leiva, Claudio Aciti, José Marone, Oscar Goñi, Luis Pantaleone, Nelson Acosta, Hugo Curti, Juan Manuel Toloza, Carlos Kornuta**  
**Grupo Inca/Intia**  
**Departamento de Computación y Sistemas**  
**Facultad de Ciencias Exactas**  
**Universidad Nacional del Centro de la Provincia de Buenos Aires**  
**Tandil – Buenos Aires - Argentina**  
*mtosini@exa.unicen.edu.ar*

### RESUMEN

Se presentan los aspectos relevantes de un proyecto de investigación y desarrollo abordado en el grupo Inca del Instituto Intia de la UNCPBA. Este proyecto fue oportunamente aprobado en el marco de los incentivos a la investigación y busca como objetivos generales el trabajo conjunto tendiente a desarrollar tecnologías para el manejo de dispositivos en sistemas embebidos.

Conjuntamente se realiza la formación de varios de los integrantes quienes se encuentran realizando sus estudios de posgrado en temas afines al proyecto.

El proyecto busca desarrollar tareas de investigación y desarrollo en temas relacionados a los sistemas embebidos, tanto en aspectos de software como de hardware. De este modo se busca desarrollar soluciones a problemas en áreas como visualización, optimización de sistemas operativos de tiempo real, sensores, etc. no solo desde el punto de vista algorítmico sino a partir de su implementación en sistemas de hardware programable como FPGAs.

**Palabras clave:** *sistemas embebidos, tiempo real, procesamiento de imágenes, aritmética decimal.*

### CONTEXTO

El proyecto presentado aborda distintos aspectos de metodologías de diseño de sistemas embebidos tales como tratamiento de imágenes y video, tiempo real y manejo de prioridades en sistemas operativos en entornos embebidos, aritmética digital, y otros.

Formalmente se enmarca dentro del sistema de incentivos a la investigación como proyecto tri-anual llevado adelante en el grupo de trabajo INCA (Investigación en Computación Aplicada) del Instituto INTIA (Instituto de Investigación en Tecnología Informática Avanzada) de la Facultad de Ciencias Exactas de la Universidad Nacional del Centro de la Provincia de Buenos Aires.

### INTRODUCCION

El contexto tecnológico de esta propuesta es la generación de soluciones informáticas que resuelvan problemas en el ámbito de los llamados sistemas embebidos, en general, y relacionados con la lógica programable, en particular.

En un sentido amplio, “Sistema embebido” es la denominación aplicable a los equipos electrónicos que incluyen procesamiento de datos, pero que, a diferencia de una

computadora de propósito general, están diseñados para satisfacer una función específica, como en el caso de un reloj digital, un reproductor de MP3, un teléfono celular, un router, el sistema de control de un automóvil (ECU), etc.

Componentes fundamentales de un sistema embebido son la arquitectura de hardware subyacente, los dispositivos de acceso (fundamentalmente sensores) y el software de control, en general un sistema operativo de tiempo real de altas prestaciones y optimizado para funcionar en sistemas de tiempos óptimos de respuesta, bajo consumo y alta fiabilidad.

De este modo, un objetivo general de la propuesta es el estudio de soluciones de hardware (sensores, aritmética, procesadores dedicados y coprocesadores) y software (sistemas operativos para sistemas embebidos y aplicaciones de control) orientados al desarrollo de sistemas dedicados de alta performance. La aplicación de los desarrollos logrados se realizará, preferentemente, en tareas de análisis, detección y control a partir del procesamiento de imágenes.

## LÍNEAS DE TRABAJO

Los objetivos específicos del proyecto definen las líneas de trabajo principales que se pueden detallar según los siguientes enfoques:

Diseño de sistemas digitales: Estudio, desarrollo e implementación de técnicas de síntesis de circuitos digitales para sistemas embebidos orientados a la implementación en hardware sobre dispositivos programables como FPGA's (Field Programmable Gate Array), (programación del hardware, o microcontroladores, programación del software.

Aritmética. Realizar unidades aritméticas en FPGA, con circuitos dedicados para las operaciones básicas de multiplicación y división; y especiales como exponenciales,

logarítmicas y trigonométricas, de altas prestaciones en tiempo y precisión. Un enfoque particular se está dando hacia la aritmética decimal por su reciente interés en aplicaciones comerciales, financieras, en ingeniería y científicas.

Verificación: Estudiar, aplicar y proponer técnicas modernas de verificación funcional de sistemas digitales para los diseños en el proyecto. En particular van a considerarse técnicas de verificación basada en aserciones y verificación dirigida por la cobertura por un lado y lenguajes de verificación como SystemVerilog o SystemC y frameworks de verificación como OVM (Open Verification Methodology) y UVM (Universal Verification Methodology) por otro.

Sistemas operativos embebidos. Estudio de sistemas operativos embebidos/de tiempo real de tipo FOSS (Free & Open Source Software). Desarrollo de aplicaciones sobre estos. Con particular énfasis, en el estudio, desarrollo e implementación de técnicas de planificación de tareas con prioridades dinámicas en dispositivos con sistemas operativos de tiempo real embebidos.

Imágenes. Estudio de técnicas para la implementación de sistemas de machine vision (sistemas de visión industriales), reconocimiento de escenarios 3D orientados al desarrollo de algoritmos en sistemas embebidos para la asistencia en el control de UAV's (Unmanned Aerial Vehicles) y UGV's (Unmanned Ground Vehicles).

## RESULTADOS ESPERADOS

Los sistemas embebidos como tecnología son un motor clave de la industria y del desarrollo tecnológico, y es un área que en los últimos años ha crecido notablemente en la Argentina.

En el país existen varios grupos de trabajo en el tema en instituciones de prestigio como la UBA, el INTI, la UNS, la UNMDP, el ITBA, la UNLP, la UNSJ y

UNLaM entre otros. La UNCPBA no está ausente de estas iniciativas en el tema ya que desde hace algunos años se ha iniciado desde el grupo INCA un contacto intenso con investigadores trabajando en temáticas afines al uso de dispositivos programables, en un principio, y al desarrollo de sistemas embebidos, en los últimos tiempos. El Dr. Todorovich tuvo participación activa en el Comité Organizador de la edición 2011 del Simposio Argentino de Sistemas Embebidos que se llevó a cabo en Buenos Aires a principios de año.

En lo académico, se busca acrecentar la formación de algunos de los integrantes que se encuentran inscriptos o realizando sus respectivos doctorados. En este sentido, se espera que durante el periodo de vigencia del proyecto terminen sus doctorados Martín Vázquez, Lucas Leiva, Claudio Aciti y Marcelo Tosini; y su maestría José Marone. Además, la experiencia adquirida en el tema permitirá aumentar la oferta académica de grado y/o posgrado del Departamento de Computación y Sistemas de la Facultad de Ciencias Exactas a partir del dictado de cursos de especialización y optativas con temas acordes como manejo de sistemas operativos de tiempo real, programación de microcontroladores, diseño de sistemas de aplicación específica, procesamiento de imágenes, verificación de diseños de hardware, etc..

Se hará hincapié también en la captación de nuevos recursos humanos a partir de una oferta amplia de temas de tesis de grado de la carrera de Ingeniería de Sistemas de la facultad, por un lado, y la incorporación (en la medida de lo posible) de becarios alumnos trabajando en temas específicos del proyecto.

Respecto de la transferencia, se espera una creciente vinculación con instituciones académicas y empresas interesadas en el desarrollo de microemprendimientos productivos que involucran la generación de soluciones con soporte de sistemas embebidos.

## RECURSOS HUMANOS

El proyecto Metodologías de Diseño para Sistemas Embebidos consta de ocho investigadores y cinco becarios de postgrado que trabajan activamente en tareas de investigación y desarrollo en la temáticas relativas al proyecto: robótica, sistemas autónomos, inteligencia artificial, aritméticas, arquitecturas de hardware, sistemas de visión, tiempo real, etc.

Asimismo, el grupo tiene una estrecha vinculación con investigadores de la Universidad Autónoma de Madrid que trabajan en el área de aritméticas en FPGA y con los cuales se han realizado varias publicaciones conjuntas.

En lo académico, se busca acrecentar la formación de algunos de los integrantes que se encuentran inscriptos o realizando sus respectivos doctorados. En este sentido, se espera que durante el periodo de vigencia del proyecto terminen sus doctorados cuatro de los integrantes; y su maestría otro de ellos. Además, la experiencia adquirida en el tema permitirá aumentar la oferta académica de grado y/o posgrado del Departamento de Computación y Sistemas de la Facultad a partir del dictado de cursos de especialización y optativas con temas afines como manejo de sistemas operativos de tiempo real, programación de microcontroladores, diseño de sistemas de aplicación específica, procesamiento de imágenes, verificación de diseños de hardware y otros.

## BIBLIOGRAFÍA

- ❖ **JP. Deschamps, G. Bioul, G. Sutter**, “*Synthesis of Arithmetic Circuits*”, - ASIC, FPGA, Embedded Systems”, John Wiley & Sons, New-York, February 2006.



- ❖ **Jean-Pierre Deschamps, José Luis Imaña and Gustavo D. Sutter** “Hardware Implementation of Finite-Field Arithmetic”, McGraw-Hill. ISBN: 0-071-54581-6 / 978-007154581-5, 348 pages. March 2009.M.F. Cowlishaw: “Decimal floating-point: algorithm for computers”, Proc. 16th IEEE Symposium on Computer Arithmetic, June 2003, pp. 104–111.
- ❖ **Ahmed Nabil Belbachir (Ed.)**, “Smart Cameras”, Springer, ISBN: 978-1-4419-0952-7, 2010.
- ❖ **M. Graves, B. Batchelor**, “Machine Vision For The Inspection Of Natural Products”, Springer-Verlag, Springer, 2004.
- ❖ **Kehtarnavaz, N. and Gamadia, M.** "Real-Time Image and Video Processing: From Research to Reality", Morgan & Claypool Publishers, 2005
- ❖ **B. Kisacanin, S. S. Bhattacharyya, and S. Chai**, “Embedded Computer Vision”. Springer Publishing Company, Incorporated, 2008.
- ❖ **Michael F. Cowlishaw**. “Decimal Floating\_Point\_ Algorithm for Computers”. IEEE Proceedings of the 16th Symposium on Computer Arithmetic, pp. 104-111, IEEE, June 2003.
- ❖ **A. Vázquez, E. Antelo, and P. Montuschi**. “A New Family of High-Performance Parallel Decimal Multipliers”. Proceedings of the 18th IEEE Symposium on Computer Arithmetic, pp. 195-204, Jun. 2007.
- ❖ **A. Vazquez and E. Antelo**. “A High-Performance Significant BCD Adder with IEEE 754-2008 Decimal Rounding”. 19th IEEE Symposium on Computer Arithmetic, ARITH 2009, pp. 135-144, June 2009.
- ❖ **Duy-Nguyen Ta Wei-Chao Chen Natasha Gelfand Kari Pulli** , “SURFTrac: Efficient Tracking and Continuous Object Recognition using Local Feature Descriptors” - 2009
- ❖ **Chuntao Wang; Jin-Hyung Kim; Keun-Yung Byun; Sung-Jea Ko** , “Robust digital image stabilization using feature tracking” - 2009
- ❖ **Delrieux C., Manera J., Rodriguez L., Coppo R.**, “Adquisición y Procesamiento de Imágenes Aéreas para Sensado Remoto” - 2009
- ❖ **Alan C. Bovik**, “Handbook of Image and Video Processing”– ISBN-10: 0121197905 -2000
- ❖ **Michael Keating and Pierre Bricaud**, "Reuse Methodology Manual for System-on-a-Chip Designs", 3rd Ed, Kluwer Academic Publishers, 2002. (ISBN: 1-4020-7141-8)
- ❖ **Harry D. Foster, Adam C. Krolnik, David J. Lacey**, “Assertion-Based Design”, 2nd edition, Springer, 2004. (ISBN: 1402080271).
- ❖ **Janick Bergeron**, “Writing Testbenches: Functional Verification of HDL Models”, Second Edition, Springer, 2003. (ISBN: 1402074018).
- ❖ **Ashenden, Peter J. and Jim Lewis**, “VHDL-2008 just the new stuff”, ISBN: 9780123742490, Amsterdam, Elsevier Morgan Kaufmann, 2008.
- ❖ **Stuart Sutherland, Simon Davidmann, Peter Flake**, “SystemVerilog for design a guide to using systemVerilog for hardware design and modelling”, 2nd ed., NY Springer, 2006 (ISBN: 9780387333991).

- ❖ **Chen, M. and Lin, K.**, “Dynamic priority ceilings: A concurrency control protocol for real-time systems”. Real-Time Systems, Kluwer Academic Publishers, (no. 2):325–346, (1990).
- ❖ **Liu, C. and Layland, J.**, “Scheduling algorithms for multi-programming in a hard realtime environment”. Journal of the Association for Computing Machinery, vol. 20(no. 1):46–61. (1973).

## Desarrollo de Sistemas de Apoyo para la Toma de Decisiones en Procesos Productivos

Enrique E. Tarifa, Sergio L. Martínez, Susana A. Chalabe  
 Facultad de Ingeniería / Universidad Nacional de Jujuy / CONICET  
 Av. Italia y Av. Martiarena / S. S. de Jujuy / Provincia de Jujuy/Tel. 0388-4221591  
 etarifa@fi.unju.edu.ar, smartinez@fi.unju.edu.ar, susana.chalabe@gmail.com

### Resumen

La complejidad creciente de los escenarios planteados por la globalización, la competencia, los avances tecnológicos, los problemas ambientales y sociales, entre otros, requieren de herramientas especializadas que ayuden a tomar decisiones.

Los procesos tecnológicos y productivos requieren tomar decisiones inmediatas y efectivas. Independientemente del área particular de aplicación, las decisiones se toman ejecutando las siguientes etapas: 1) Identificación y análisis del problema; 2) Identificación de los criterios; 3) Asignación de peso a los criterios; 4) Generación de alternativas; 5) Evaluación de las alternativas; 6) Selección de la mejor alternativa; 7) Implementación de la decisión; 8) Evaluación de los resultados.

Debido a esta independencia, es posible y conveniente el esfuerzo coordinado entre grupos multidisciplinarios, dejando luego a cada grupo el desarrollo de los aspectos particulares del sistema.

Este proyecto está orientado al desarrollo de sistemas de apoyo para la toma de decisiones en procesos productivos y cuenta con el trabajo coordinado de tres grupos de investigadores; uno especialista en Ingeniería de Procesos y en Tecnología de Alimentos, otro especialista en Gestión de Cuencas Hidrográficas, y otro especialista en Informática.

Los destinatarios de los sistemas a desarrollar son las personas a cargo de procesos productivos vinculados a la industria química, a la industria de alimentos y a cuencas hidrográficas.

**Palabras clave:** Toma de decisiones, simulación, optimización, inteligencia artificial, teoría de juegos, ingeniería de procesos, tecnología de alimentos, gestión de cuencas hidrográficas.

### Contexto

El presente proyecto se enmarca en el plan de Proyectos de Innovación Científica y Tecnológica Orientados de la Agencia Nacional de Promoción Científica y Tecnológica. Este proyecto se identifica como PICTO-2008-00154 categoría 1-A, y es financiado por la Agencia de Promoción Científica y Tecnológica (ANPCyT) y la Universidad Nacional de Jujuy (UNJu).

El proyecto se inscribe en el Área Prioritaria N° 1: “Desarrollo Económico–Social Regional Sustentable”, y comprende cuatro áreas disciplinarias, a saber: 1) Ingeniería de Procesos, 2) Tecnología de Alimentos, 3) Gestión de Cuencas Hidrográficas y 4) Informática.

En tal contexto, el proyecto se organiza en base a tres subproyectos para sustentar las áreas disciplinares mencionadas que comprende:

- Áreas disciplinares 1 y 2 – subproyecto N° 1 “Optimización del diseño, la operación y la gestión de procesos productivos”. Dirigido por el Dr. Enrique E. Tarifa (Categoría II – UNJu).
- Área disciplinar 3 – subproyecto N° 2 “Optimización de la gestión de cuencas”. Dirigido por la Geól. Susana Chalabe (categoría III – UNJu).
- Área disciplinar 4 – subproyecto N° 3 “Desarrollo de sistemas de supervisión, control y gestión para procesos productivos”. Dirigido por el Ing. Sergio L. Martínez (categoría III – UNJu).

### Introducción

Para tomar una buena decisión, es fundamental contar con información adecuada en cantidad y calidad para realizar un buen diagnóstico. También es necesario contar

con un profundo conocimiento del comportamiento del sistema bajo supervisión para poder plantear alternativas de acción realistas y decidir los mejores cursos de acción. Con estos criterios se enfoca el desarrollo de este proyecto que tiene como objetivo principal el desarrollo de sistemas de soporte a la toma de decisiones (DSS – *Decision Support Systems*) (Turban & Aronson, 2001; Hepting, 2007; Li *et al.*, 2007; Phillips-Wren *et al.*, 2009).

El desarrollo de un DSS requiere del aporte de diversas áreas, como ser: Sistemas Transaccionales (OLTP – *On Line Transaction Processing*), Procesamiento Analítico en Línea (OLAP – *On Line Analytical Processing*), datawarehouse, descubrimiento de conocimiento a partir de bases de datos (KDD – *Knowledge Discover from Data bases*) (Chapman *et al.*, 2000; Hernández Orallo, 2004), simulación (Tarifa *et al.*, 1998; Tarifa *et al.*, 1999), optimización (Reklaitis, 1983; Edgar & Himmelblau, 2001) y teoría de juegos (Tarifa *et al.*, 2006 a-b; Gibbons, 1992; Vega Redondo, 2000) entre otras.

Los DSS son sistemas informáticos interactivos que ayudan a tomar decisiones utilizando datos y modelos para resolver problemas no estructurados. Estos sistemas se componen de una base de datos, una base de modelos, un administrador para cada base y una interfaz de usuario (Turban & Aronson, 2001; Frombo *et al.*, 2009). Las investigaciones y desarrollos planteados para este proyecto contemplan la obtención de datos de sistemas reales, series históricas o simulación, según disponibilidades; por otro lado, se prevé el desarrollo de modelos de simulación y optimización adecuados para el sistema productivo en estudio.

Las actividades planificadas para el presente proyecto están diseñadas para realizar aportes científicos en todas las áreas involucradas mientras se desarrollan las tecnologías necesarias para resolver los problemas que dieron origen al proyecto. Luego, los resultados obtenidos serán transferidos a las industrias, productores y comunidades involucradas, como así también al gobierno de la provincia de Jujuy.

## Líneas de Investigación y Desarrollo

Entre las líneas prioritarias establecidas por la provincia de Jujuy desde el 2005 para los Proyectos Federales de Innovación Productiva (PFIP) se destacan: 1) Manejos de sistemas de Cuencas en la Quebrada de Humahuaca, 2) Innovaciones en el manejo ganadero subtropical y de zonas áridas, 3) Tecnologías de riego para el uso eficiente del agua, 4) Cultivos tradicionales (papa, oca, maíz, quinua, kimicha, yacón), 5) Bio-combustibles.

Por otra parte, el Plan Nacional de Ciencia, Tecnología e Innovación 2006-2010 “Bicentenario” establece la necesidad de impulsar a las siguientes temáticas:

- Competitividad de la Industria y Modernización de sus Métodos de Producción.
- Competitividad y Diversificación Sustentable de la Producción Agropecuaria.
- Agroindustrias y Agroalimentos.
- Medio Ambiente, Tecnologías Limpias y Remediación de la Contaminación Ambiental.

Las citadas líneas prioritarias tienen un elemento en común: la necesidad de tomar decisiones adecuadas tanto a nivel de gestión como a nivel de producción, tanto a nivel gubernamental como a nivel privado.

En este marco, el presente proyecto tiene como objetivo general desarrollar sistemas de apoyo para la toma de decisiones (DSS – *Decision Support Systems*) en procesos productivos, particularmente en plantas químicas, industria de alimentos y actividades productivas vinculadas a cuencas hidrográficas. Por ello, todos los estudios y sistemas que se desarrollan están relacionados directa o indirectamente con las líneas de investigación y temáticas antes mencionadas.

## Objetivos y Resultados

El presente proyecto tiene como objetivo general desarrollar sistemas de apoyo para la toma de decisiones (*decision support system*, DSS) en procesos productivos, es especial en plantas químicas, en la industria de alimentos y actividades productivas



(agropecuaria o ganadera) vinculadas a cuencas hidrográficas.

A continuación se presentan los objetivos asociados a cada subproyecto:

Subproyecto 1. "Optimización del diseño, la operación y la gestión de procesos productivos":

- 1) Optimizar el diseño de procesos productivos.
- 2) Optimizar la operación de procesos productivos.
- 3) Optimizar la gestión de procesos productivos.

Subproyecto 2. "Optimización de la gestión de cuencas":

- 1) Optimizar la gestión de la producción agrícola y ganadera.
- 2) Optimizar la gestión del agua potable y de riego.
- 3) Optimizar el manejo de caudales torrenciales.
- 4) Identificar vulnerabilidades y políticas óptimas de gobierno con la participación de la comunidad.

Subproyecto 3. "Desarrollo de sistemas de supervisión, control y gestión para procesos productivos":

- 1) Desarrollar sistemas de diagnóstico de fallas.
- 2) Desarrollar sistemas predictivos.
- 3) Desarrollar sistemas de control.
- 4) Desarrollar sistemas de supervisión.

Para el grado de desarrollo actual del proyecto, se realizaron diversos trabajos que dieron lugar a transferencias, cursos, presentaciones y publicaciones en congresos nacionales e internacionales y revistas científicas.

#### **Publicaciones internacionales**

"Teoría de Juegos aplicada a Políticas Públicas. Caso de Estudio: Bienes Públicos", Strategy & Management Business Review, ISSN: 0718-8714, 2 (1), 25-37, 2011.

"Flexibility Study for a MSF by Monte Carlo Simulation", Expanding Issues in Desalination, Robert Y. Ning (Ed.), ISBN: 978-953-307-624-9, InTech, 2011.

"Planteamiento de elaboración de una Infraestructura de datos espaciales a nivel del

Gobierno Provincial en Jujuy. (Argentina)", II Jornadas Ibéricas de Infraestructuras de Datos Espaciales. Barcelona. España 2011.

"Derivación de Reglas de Control MIMO para un Controlador Fuzzy sin Conocimiento Experto", Cuarto Simposio Internacional de Investigación, UCSE-DASS. Jujuy, Argentina, 2011.

"Simulación de un Intercambiador de Calor Utilizando Volúmenes Finitos Móviles", Cuarto Simposio Internacional de Investigación, UCSE-DASS. Jujuy, Argentina, 2011.

"Análisis de sensibilidad por simulación del proceso de deshidratación de una planta de acondicionamiento de gas natural", Avances en Ciencias e Ingeniería, ISSN: 0718-8706, aceptado para vol. 3(3), 2012. La Serena, Chile.

"Comparación de Desempeño de Controladores Fuzzy y Controladores Proporcionales", Avances en Ciencias e Ingeniería, ISSN: 0718-8706, vol. 3 (3), 2012.

"Simulación estacionaria de un evaporador doble efecto para jugo de tomate", Avances en Ciencias e Ingeniería, ISSN: 0718-8706, vol. 3 (4), 2012.

"Endulzamiento de Gas Natural con Aminas. Simulación del Proceso y Análisis de Sensibilidad Paramétrico", Avances en Ciencias e Ingeniería, ISSN: 0718-8706, vol. 3 (4), 2012. La Serena, Chile.

#### **Publicaciones nacionales**

"Desarrollo de un Sistema de Diagnóstico de Fallas para Equipos de Procesos basado en Estados Anormales Caracterizados", WICC 2011, Universidad Nacional de Rosario, Rosario, Santa Fe, 2011.

"Desarrollo de sistemas de apoyo para la toma de decisiones en procesos industriales", WICC 2011, Universidad Nacional de Rosario, Rosario, Santa Fe, 2011.

"Control neuronal tipo MIMO aplicado a un mezclador de corrientes líquidas", VII Jornadas de Ciencia y Tecnología de Facultades de Ingeniería del NOA, Catamarca, 2011.

"Estudio de flexibilidad de un evaporador doble efecto para jugo de tomate", VII Jornadas de Ciencia y Tecnología de Facultades de Ingeniería del NOA, Catamarca, 2011.

“Optimización de un evaporador doble efecto para jugo de tomate”, VII Jornadas de Ciencia y Tecnología de Facultades de Ingeniería del NOA, Catamarca, 2011.

### **Transferencias**

Solicitante y destinatario: YPF. Estudio para la optimización de la red de hidrógeno de la refinería de Luján de Cuyo de YPF (Mendoza, Argentina). Participó el Dr. Enrique E. Tarifa (UNJu, CONICET) y el Dr. Carlos Vera (UNL, CONICET).

Solicitante y destinatario: Universidad Politécnica de Madrid, Dictado del seminario “Técnicas estadísticas de modelización: optimización y sistemas inteligentes”, Madrid, España, 2011.

Solicitante: Secretaria de Planificación, Ministerio de Infraestructura y Producción. Destinatario: Grupo de trabajo del Programa de Riesgos y Desastres. Se integra y coordina el grupo de trabajo provincial para la elaboración de Cuadros de Situación de riesgos y desastres que sirvan de base para la elaboración del Plan estratégico Territorial.

Solicitante: Dirección Provincial de Recursos Hídricos. Destinatario: Sector Hidrología. Se presta asistencia técnica para la recopilación de información histórica de 130 estaciones que registraron datos en determinadas épocas.

### **Formación de recursos humanos**

El equipo de trabajo está formado por profesores, graduados y pasantes de la Facultad de Ingeniería y la Facultad de Ciencias Agrarias de la UNJu, quienes desarrollaron las siguientes actividades en la formación de recursos humanos:

#### **Becas doctorales**

- Ing. Julieta Martínez, con beca postdoctoral de CONICET, desde 2012. En 2012, se doctoró con la tesis doctoral “Desarrollo de modelos de simulación y optimización para la toma de decisión en la gestión de cuencas en la Quebrada de Humahuaca”. Director: Dr. Enrique Tarifa. Se vincula con el proyecto PICTO a

través del subproyecto “Optimización de la gestión de cuencas”.

- Ing. Álvaro F. Núñez, con beca doctoral de la ANPCyT, desde 2009. Cursa el Doctorado Regional en Ciencia y Tecnología de Alimentos en la Facultad de Ingeniería de la UNJu. Tesis doctoral “Simulación, optimización y control de procesos de la industria alimentaria de la región”. Director: Dr. Enrique Tarifa. Se vincula con el proyecto PICTO a través del subproyecto “Optimización del diseño, la operación y la gestión de procesos productivos”.
- Ing. Leonel Benítez, con beca doctoral de CONICET, desde 2011. Cursa el Doctorado en Ingeniería en la Facultad de Ingeniería de la UNSa. Tesis doctoral “Desarrollo de procedimientos óptimos de operación para el sector de ajuste de punto del rocío en plantas de acondicionamiento de gas natural”. Director: Dr. Enrique Tarifa. Se vincula con el proyecto PICTO a través del subproyecto “Optimización del diseño, la operación y la gestión de procesos”.
- Lic. Jorgelina F. Argañaraz. Obtuvo una beca de iniciación de la ANPCyT. Cursa el Doctorado en Ciencias Sociales en la Facultad de Humanidades y Ciencias Sociales de la UNJu. Tesis: “Identificación de vulnerabilidades y políticas óptimas de gobierno con participación de la comunidad en la gestión de Cuencas”. Director: Dr. Omar Jerez. Se vincula con el proyecto PICTO a través del subproyecto “Optimización de la gestión de cuencas”.

#### **Becas en investigación**

- Sr. Marco Antonio Acuña, con beca de Estímulo a las Vocaciones Científicas, CIN, desde 2011. Título del plan de trabajo “Simulación Estacionaria Avanzada en HYSYS”. Se vincula con el proyecto PICTO a través del subproyecto “Desarrollo de sistemas de supervisión, control y gestión para procesos productivos”.
- Srta. María Gabriela Franco, con beca de Estímulo a las Vocaciones Científicas,

CIN, desde 2011. Título del plan de trabajo "Estudios y análisis de las Vulnerabilidades y elementos expuestos al riesgo utilizando sistemas de información geográficos". Directora: Susana A. Chalabe. Se vincula con el proyecto PICTO a través del subproyecto "Optimización de la gestión de cuencas".

- Sr. Farid Astorga, con beca de Estímulo a las Vocaciones Científicas, CIN, desde 2011. Título del plan de trabajo "Catalogador de Metadatos como servicio CSW del Geoportal IDE-UGICH basado en normas OGC e ISO". Directora: Susana A. Chalabe. Se vincula con el proyecto PICTO a través del subproyecto "Optimización de la gestión de cuencas".

### Conferencias

"A 25 años del desastre nuclear en Chernobyl", IX Semana Nacional de la Ciencia y la Tecnología, 16 de junio de 2011, Facultad de Ingeniería (UNJu), auspiciada por la ANPCyT, UNJu, 2011.

"Accidentes en Plantas Químicas", Asociación Jujeña de Estudiantes de Ingeniería Química (A.J.E.I.Q.), 20 de octubre de 2011, Facultad de Ingeniería, UNJu, 2011.

### Referencias

Chapman P., Clinton J., Kerber R., Khabaza T., Reinartz T., Shearer C., Wirth R., "CRISP-DM 1.0: Step-by-step data mining guide", *SPSS White paper* - technical report CRISPWP-0800, SPSS Inc., 2000.

Edgar, T.F. and Himmelblau D.M., "Optimization of Chemical Processes", 2° edition, *McGraw-Hill*, New York, 2001.

Frombo F., Minciardi R., Robba M., Sacile, R., "Decision support system for planning biomass-based energy production", *Energy*, (34) 362-369, 2009.

Gibbons R., "A Primer in Game Theory", *Prentice Hall*, London, 1992.

Hepting D.H., "Decision support for local environmental impact assessment", *Envi-*

*ronmental Modelling & Software*, Volume 22, Issue 4, 436-441, 2007.

Hernández Orallo J., Ramírez Quintana J., Ferri Ramírez C., "Introducción a la Minería de Datos", *Ed. Prentice Hall*, España, 2004.

Li J.Q., Borenstein D., Mirchandani P.B., "A decision support system for the single-depot vehicle rescheduling problem", *Computers & Operations Research*, Volume 34, Issue 4, 1008-1032, 2007.

Phillips-Wren G., Mora M., Forgionne G.A., Gupta J.N.D., "An integrative evaluation framework for intelligent decision support systems", *European Journal of Operational Research*, (195) 642-652, 2009.

Reklaitis G.V., Ravindran A., Ragsdell K.M., "Engineering Optimization - Methods and Applications", *John Wiley & Sons*, New York, 1983.

Tarifa E., Humana D., Madregal S., Cayo N., "Módulos para Capacitación en Tecnología", *III Congreso Internacional de (Tele) Informática Educativa*, Santa Fe. 1999.

Tarifa E.E., Battezzati M.V., Verazay A.R.N., "Juego Comercio - Ampliación del duopolio de Cournot mediante simulación", *II Jornadas de Ciencia y Tecnología del NOA*, Catamarca, Argentina, ISBN 987-9170-60 1, vol II, 542-547, 2006 (a).

Tarifa E.E., Battezzati M.V., Verazay A.R.N., "Juego Comercio - Determinación de puntos de equilibrios por simulación", *II Jornadas de Ciencia y Tecnología del NOA*, Catamarca, Argentina, ISBN 987-9170-60 1, vol II, 536-541, 2006 (b).

Tarifa E.E., Humana D., Vázquez G., Franco S., Scenna N.J., "Simulación Dinámica en Tiempo Real. Pasteurizador HTST", UNJu, Jujuy. 1998.

Turban E., Aronson J.E., "Decision Support Systems And Intelligent Systems", sixth edition, *Prentice Hall*, New Jersey, 2001.

Vega Redondo F., "Economía y Juegos", *Antoni Bosh*, Barcelona, 2000.

## Aplicación de interfases lectoras de bioseñales en el contexto de la domótica

Jorge Ierache<sup>1,2</sup>, Gustavo Pereira<sup>1</sup>, Iris Satollo, Juan Iribarren<sup>1</sup>

Instituto de Sistemas Inteligentes y Enseñanza Experimental de la Robótica (ISIER)<sup>1</sup>  
Laboratorio de sistemas de información avanzados Facultad de Ingeniería Universidad de Buenos Aires<sup>2</sup>

ISIER, Facultad de Informática Ciencias de la Comunicación y Técnicas Especiales  
Universidad de Morón, Cabildo 134, (B1708JPD) Morón, Buenos Aires, Argentina  
54 11 5627 200 int 189  
jierache@unimoron.edu.ar

*Resumen: En este trabajo presenta los las líneas de investigación aplicada al control de artefactos, a través de bioseñales cerebrales. Se basa en los resultados iniciales de la investigación de tecnologías e interfases disponibles que facilitan la lectura de las señales del cerebro del usuario y su asociación a comandos explícitos. Nuestro trabajo presenta una solución, aplicando el desarrollo de un framework de comunicación y control de artefactos, sobre la base de la experiencia inicial del control de robots.*

**Key words:** Demotic, Brain Machine Interface, Bio-Electrical Signal, Human Machine Interfaces.

### 1. Contexto.

La línea de investigación y desarrollo se orienta al estudio y desarrollo de soluciones ingenieriles para aplicar Brain-Machine Interfase (BMI) al control de artefactos y robots. Este trabajo es financiado por la

Universidad de Morón en el contexto del PID 01-001/12 radicado en el Instituto de Sistemas Inteligentes y Enseñanza Experimental de la Robótica (ISIER) con la cooperación del Laboratorio de sistemas de información avanzados del departamento de computación de la Facultad de Ingeniería de la Universidad de Buenos Aires.

### 2. Introducción.

La aplicación de bioseñales para el control de sistemas, robots, aplicaciones, juegos y otros dispositivos, presenta un enfoque novedoso al abrir las puertas para la interacción entre humanos y computadoras en una nueva dimensión, donde se explotan específicamente biopotenciales eléctricos registrados en el usuario, a través de: el electro-miograma, el electro-encefalograma y el electro-oculograma, que son bioseñales eléctricas generadas por los patrones de actividad de los músculos, el cerebro y los ojos del usuario

En este orden, se presentaron diversos trabajos: los primeros, recurrieron a implantar electrodos intracraneales en la corteza motora de primates [1], [2]. Los trabajos no invasivos para humanos recurrieron a señales de Electroencefalogramas (EEG), aplicados a ejercicios de comandos mentales, como mover el cursor de una computadora [3], [4] basados en el empleo de Brain-Machine Interface (BMI). Millan et. al. [5] demuestra como dos personas pueden mover un robot usando un simple electroencefalograma, sobre



la base de reconocer tres estados mentales, los que se asocian a comandos del robot. Los trabajos de Saulnier et. al. [6] se enfocaron en controlar la velocidad de un robot y extender su aplicación para inferir el nivel de stress del usuario, y a partir de éste influir en el comportamiento social de robots domésticos, en este caso una aspiradora robot.

El trabajo seminal de Millan et. al. [5], emplea como única bioseñal el electroencefalograma, sobre la base del trabajo de dos personas para apoyar la navegación de un robot, a diferencia de éste, nuestro trabajo presenta el resultado preliminar empleando un BMI de bajo costo, utilizado en trabajos secundarios como el de Saulnier et. al. [6], que incluye las bioseñales correspondientes al electroencefalograma, electro-oculograma y electromiograma.

A diferencia del trabajo de Saulnier et. al. [6], que implementa un control de velocidad sobre la base del electromiograma e infiere el estado de stress del usuario a través del electroencefalograma, nuestro trabajo se enfoca inicialmente en la ejecución de comandos para la navegación por parte de un robot [7] y actualmente en el control de artefactos en un contexto de domótica.

Controlar artefactos, mover robots o facilitar la aplicación de dispositivos para discapacitados sin aplicar controles manuales y alcanzar el control sólo a través de la actividad mental, fascinó a los investigadores, si bien alcanzar una plasticidad con un BMI requiere de tiempo por parte del usuario, en nuestras experiencias para facilitar el empleo a un usuario con mínimo entrenamiento se desarrolló el control por auto foco con la finalidad que el robot NXT sea guiado con el empleo de un BMI-NIA [8], para cumplir con un patrón de navegación, logrando una mejora en los tiempos de control mental, superando ligeramente al manual, en las pruebas de ejecución del mismo patrón de navegación [9], [10], [11]. En trabajos realizados en el año 2011 se experimentó el control remoto vía internet de un robot NXT [11] vía Internet con la aplicación de

bioseñales con el BMI de NIA [8]. En la figura N° 1, se muestra el BMI (izq) y la Banda Lectora (der).



Figura N° 1 BMI-Banda NIA

## 2. Resultados Obtenidos

Se efectuaron demostraciones de los resultados durante el TEYET 2012 [12], empleando el BMI de EMOTIV figura N° 2.

El mencionado BMI “EMOTIV” [13] se aplicó para el control de un robot NXT (Adelante, Atrás, Izquierda, Derecha).



Figura N° 2 “BMI Emotiv”

Para mantener la premisa de empleo del BMI por parte de un usuario sin experiencia, se plantearon dos comandos: uno que permitiese controlar la selección de comportamientos y otro, la ejecución de los mismos sobre la base de sus propios controladores (por ejemplo moverse hacia adelante o girar a la derecha).

Se estableció una arquitectura experimental que contempla dos vías de comunicación, la primera vía la denominamos vía de comunicación de alto nivel ó “usuario-computadora”. Esta vía, al igual que en las investigaciones previas se instrumentó en sus inicios con un BMI de bajo costo OCZ NIA [8] de empleo experimental en video juegos

que permitió la asociación de patrones de bioseñales con el teclado de la computadora.

Sobre estas facilidades determinamos un perfil simple para el manejo de un robot o un artefacto, que asocia y caracteriza en primer lugar, el control para la ejecución del comando mental sobre la base de la detección de señales musculares, en nuestro caso a través de un leve movimiento de los párpados. En segundo lugar la selección de los comandos de alto nivel aplicados a un robot o a un artefacto, en este caso se trabajó sobre la base de las ondas cerebrales Alfa. Este tipo de bioseñales no aseguraron un adecuado control al usuario en los desplazamientos, a través de un panel ú de selección de comandos del framework de control del robot. Frente a esta situación se implementó en el framework la opción de aplicación de auto foco [9], [10], [11], para el modo control a través de BMI, mejorando el gobierno del usuario en el proceso de selección de comandos, técnica aplicable para el control de artefactos.

La vía de comunicación “computadora–robot/artectos” (segunda vía de comunicación denominada de bajo nivel). Se efectuó a través de comunicación en Bluetooth, nativa en el kit del robot Lego NXT y para la cual hay una gran cantidad de librerías disponibles para su comunicación con una computadora mediante diversos lenguajes de programación. La comunicación de bajo nivel para artefactos se efectuó a través de control IR

El BMI (Brain Machine Interface) EMOTIV EPOC, articulado con su SDK (figura N° 3) el que consta de un panel de control para generar el usuario, y registrar su perfil, además de facilitar la visualización del estado de conexión de los sensores, se cuenta con distintos patrones de registro expresivo, afectivo y cognitivo. En el expresivo se puede visualizar un avatar en el cual se podrán entrenar distintos signos de expresión de nuestra cara (pestañar, guiño izquierdo, guiño derecho, mirar a izquierda, mirar hacia la derecha, mover las cejas hacia arriba, mover las cejas hacia abajo, sonreír, apretar los

dientes). El afectivo permite verificar distintos estados de ánimo que van sucediendo en un tiempo determinado (compromiso, excitación instantánea, excitación a largo tiempo entre otras). El cognitivo permite realizar el entrenamiento de una acción a través sobre la base de un pensamiento, sobre el cual se pueden entrenar

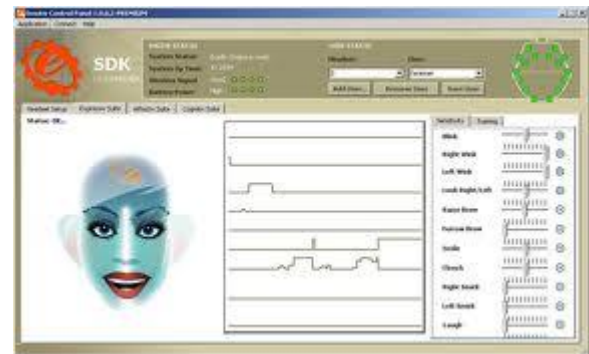


Figura N° 3 SDK-Emotiv

hasta trece acciones, de las cuales seis son movimientos direccionales (empujar, tirar hacia adelante, izquierda, derecha, arriba y abajo), seis rotacionales (giro en el sentido de las agujas del reloj, giro contrario en el sentido de las agujas del reloj, rotación a la izquierda, rotación a la derecha, adelante, atrás) y uno imaginario que es desaparecer. Otras de las herramientas con las que cuenta el SDK es Emokey (figura N° 4), que permite vincular una acción del emotiv con cualquier tecla y así poder funcionar como interfase con cualquier aplicación.

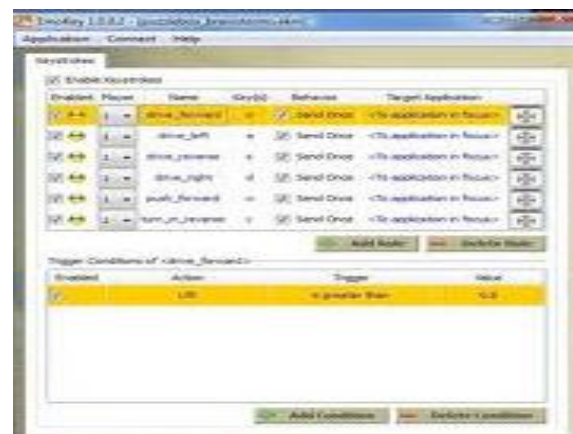


Figura N° 4 “Emokey”

Para la presentación de TEYET 2012 [13] se entrenaron dos acciones cognitivas y dos acciones expresivas para el control de un robot LEGO versión NXT mediante bluetooth a través del framework desarrollado por el Instituto de Sistemas Inteligentes y Enseñanza Experimental de la Robótica en lenguaje C#.

El framework fue diseñado para ser adaptado a cualquier comunicación y cualquier robot, con sólo agregar las librerías necesarias para realizar las primitivas de movimiento y conexión permite mover el robot con nuestros pensamientos y gestos.

Utilizando las facilidades del emokey con el empleo del dispositivo USB-UIRT [15] (figura N° 5) para obtener el código infrarrojo de un artefacto y luego poder transmitirlo, se inicio el desarrolló de un framework de domótica que facilite la comunicación del Emotiv con artefactos controlados mediante señales de infrarrojo como televisores, equipos de música, equipos de aire acondicionado, entre otros (figura N° 6).



Figura N° 6 Dispositivo USB-UIRT

El framework con una interfaz amigable facilitara la incorporación de artefactos, donde claramente se puede definir cada uno de los comandos del o los dispositivos facilitando a personas discapacitadas el control de artefactos, al igual que la interacción y control de estos usuarios con dispositivos ubicados en sitio remotos.

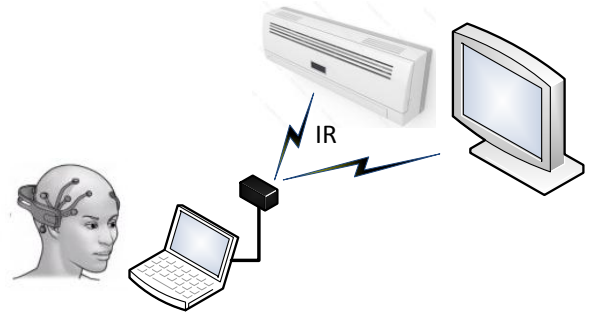


Figura N° 6 – Aplicación domótica con EMOTIV/USB-UIRT

### 3 Líneas de Investigación y Desarrollo

En el marco de las líneas de investigación actuales se implementarán mayores prestaciones de control extendidas tanto a robots como artefactos remotos de control infrarrojo (Aire Acondicionado, TV, entre otros dispositivos), apuntando tanto a dispositivos locales como a los localizados en sitios remotos mediante un único framework multiplataforma que integre librerías para una gran cantidad de dispositivos conocidos, importación y exportación de perfiles de comandos y un asistente para el entrenamiento y familiarización del usuario con el Emotiv.

### 4. Formación de Recursos Humanos

Este proyecto se encuentra financiado por la Facultad de Informática, Ciencias de la Comunicación y Técnicas Especiales de la Universidad de Morón. A su vez propicia la formación de recursos, con la participación de estudiantes de grado y posgrado para la continuación de las líneas de investigación relacionadas. Este PID lo integran cuatro investigadores y dos estudiantes de tesis de grado, orientadas al control remoto de robots y artefactos en un contexto de domótica con la aplicación de Brain-Machine Interface (BMI).

## 5. Referencias

- 1- J. Wessberg, C. R. Stambaugh, J. D. Kralik, P. D. Beck, M. Laubach, J. K., "Real-time prediction of hand trajectory by ensembles of cortical neurons in primates," *Nature*, vol. 408, pp. 361–365, 2000.
- 2- M. A. L. Nicolelis, "Brain-machine interfaces to restore motor function and probe neural circuits," *Nature Rev.Neurosci.*, vol. 4, pp. 417–422, 2003.
- 3- R. Wolpaw, D. J. McFarland, and T. M. Vaughan, "Brain-computer interface research at the Wadsworth center," *IEEE Trans. Rehab. Eng.*, vol. 8, pp. 222–226, 2000.
- 4- J. del R Millán, "Brain-computer interfaces," in *Handbook of Brain Theory and Neural Networks*, 2nd ed, M.A. Arbib, Ed. Cambridge, MA: MIT Press, 2002.
- 5- José Millán, Frédéric Renkensb, Josep Mouriñoc, and Wulfram Gerstnerb. Non-Invasive Brain-Actuated Control of a Mobile Robot by Human EEG. *IEEE Trans. on Biomedical Engineering*, Vol 51, June 2004.
- 6- Paul Saulnier, Ehud Sharlin, and Saul Greenberg. Using Bio-electrical Signals to Influence the Social Behaviours of Domesticated Robots. *HRI'09*, 2009, USA.ACM 978-1-60558-404-1/09/03.
- 7- Ierache Jorge, Pereira Gustavo, Iribarren Juan, Sattolo Iris, "Robot Control on the Basis of Bio-electrical Signals" : "International Conference on Robot Intelligence Technology and Applications" (RiTA 2012) Gwangju, Korea on December 16-18, 2012. Series *Advances in Intelligent and Soft Computing of Springer*.
- 8- [http://www.ocztechnology.com/products/ocz\\_peripherals/nia-neural\\_impulse\\_actuator](http://www.ocztechnology.com/products/ocz_peripherals/nia-neural_impulse_actuator).
- 9- Ierache, J., Dittler M., Pereira G., García Martínez R.,(2009) "Robot Control on the basis of Bio-electrical signals" XV Congreso Argentino de Ciencias de la Computación 5 al 9 octubre CACIC 2009, Universidad Nacional de Jujuy, Facultad de Ingeniería, ISBN 978-897-24068-3-9 ,pag 30.
- 10- Ierache, J., Dittler, M. García-Martínez, R., "Control de Robots con Basado en Bioseñales". XII Workshop de Investigadores en Ciencias de la Computación WICC 2010: Universidad Nacional de la Patagonia San Juan Bosco, Calafate, Santa Cruz, Argentina. 5 y 6 de mayo de 2010,ISBN 978-950-34-0652-6, pag 641.
- 11- Ierache., J, Pereira.,G, Sattolo.,I , Guerrero., A, D'Alitto J, Iribarren., J. Control vía Internet de un Robot ubicado en un sitio remoto aplicando una Interfase Cerebro-Máquina". XVII Congreso Argentino de Ciencias de la Computación 5 al 9 octubre CACIC 2011, Universidad Nacional de La Plata, Facultad de Informática, ISBN 978-950-34-0756-1, paginas 1373-1382
- 12- Ierache Jorge, Pereira Gustavo, Iribarren Juan del articulo "Demostración de los resultados en la integración de Interfases Lectoras de Bioseñales aplicadas al Control de un Robot" VII Congreso Educación en Tecnología y Tecnología en Educación 2012 Universidad Nacional del Noroeste de la Provincia de Buenos Aires. UNNOBA, 2012, demos educativas. ISBN 978-987-28186-3-0.
- 13- [www.facebook.com/isierum](http://www.facebook.com/isierum)
- 14- <http://www.emotiv.com/>
- 15- USB-UIRT: <http://www.usbuirt.com/>



## Sistemas de Tiempo Real con Requerimientos Heterogéneos: Integración Hardware-Software

Javier D. Orozco<sup>2</sup>, José M. Urriza<sup>1</sup>, Ricardo Cayssials<sup>2</sup>, Elsa Fernández<sup>3</sup>, Mariano Ferrari<sup>3</sup>, Javier Echaiz<sup>4</sup>, Carlos E. Buckle<sup>1</sup>, Damián P. Barry<sup>1</sup>, Francisco E. Páez<sup>1</sup>, Gabriela Olguín<sup>1</sup>, Luis Diaz<sup>1</sup>, Eduardo Schorb<sup>1</sup>, Lucas Schorb<sup>1</sup>, Sebastián Lucas<sup>1</sup>, Edgardo Constabel<sup>1</sup>

<sup>1</sup> Depto. de Informática, Fac.de Ingeniería, Universidad Nacional de la Patagonia San Juan Bosco.

Puerto Madryn, Argentina

+54 280-4472885 – Int. 117.

josemurriza@unp.edu.ar, cbuckle@unpata.edu.ar

<sup>2</sup> Depto. de Ingeniería Eléctrica y Computadoras - Universidad Nacional del Sur

Bahia Blanca, Argentina

+54 291-4595000 - Int. 3371.

jadorozco@gmail.com

<sup>3</sup> Depto. de Matemática, Fac.de Ingeniería, Universidad Nacional de la Patagonia San Juan Bosco.

Puerto Madryn, Argentina

+54 280-4472885 – Int. 117.

<sup>4</sup> Depto. de Ciencias e Ingeniería de la Computación - Universidad Nacional del Sur

Bahia Blanca, Argentina

+54 291-4595000 - Int. 3371.

### 1 Resumen

Este proyecto plantea el desarrollo de técnicas de modelado, diseño, análisis, optimización y testeo de sistemas embebidos con restricciones de tiempo y requerimientos heterogéneos, para diferentes plataformas de software y hardware, en la búsqueda de un apropiado balance entre prestaciones, rendimiento y eficiencia.

**Palabras clave:** Sistemas de Tiempo Real, Sistemas Operativos de Tiempo Real, Sistemas Embebidos, Planificación de CPU.

### 2 Contexto

El proyecto de investigación *Sistemas de Tiempo Real con Requerimientos Heterogéneos: Integración Hardware - Software* es dirigido por el Dr. Orozco de la Universidad

Nacional del Sur (UNS) y es codirigido por el Dr. Urriza de la Universidad Nacional de la Patagonia San Juan Bosco (UNPSJB). Las líneas de investigación de este proyecto, coinciden con las desarrolladas por los integrantes del *Real Time Systems Group*, perteneciente a la Facultad de Ingeniería de la UNPSJB Sede Puerto Madryn y con líneas de investigación desarrolladas en el Laboratorio de Sistemas Digitales del Departamento de Ing. Eléctrica y de Computadoras de la UNS del cual los doctores Orozco y Cayssials pertenecen. Además, participan dos doctores del departamento de matemática de la UNPSJB y un doctor del Departamento de Ciencias e Ingeniería de la Computación de la UNS. El proyecto es financiado por la Secretaría de Ciencia y Técnica de la UNPSJB.

### 3 Introducción

Una nación industrializada debe dominar imprescindiblemente las tecnologías de los Sistemas de

Tiempo Real (*STR*). En la actualidad los *STR* han dejado de pertenecer de forma exclusiva a grandes aplicaciones de las industrias automatizadas, la exploración espacial, dispositivos e instrumentación militar, aviónica, redes de comunicaciones, etc., para pertenecer a la gran mayoría de dispositivos electrónicos que se utilizan hoy en día. Se los puede encontrar en computadoras portátiles, tablets, teléfonos celulares, GPS, MP3s, en las computadoras de los automóviles, etc.

Los dispositivos de tiempo real por lo general cuentan con un Sistema Operativo de Tiempo Real (*SOTR*). En este, se implementan diversas técnicas y métodos como: el desarrollo de políticas de tolerancia a las fallas, atención de tareas esporádicas y aperiódicas, atención de tareas mandatorias / opcionales, calidad de servicio (*QoS*), por mencionar solamente algunas. Estas técnicas, que son parte fundamental del dispositivo, son implementadas en el núcleo del Sistema Operativo, siendo las que gobiernan parte de las decisiones del planificador de tareas.

Los *STR*, embebidos o no, por lo general están constituidos por tareas periódicas que incluyen entre sus parámetros, los instantes máximos en que las mismas deben finalizar su ejecución. Este parámetro extra se denomina vencimiento. Si finaliza después de este tiempo, se dice que la tarea ha perdido su restricción temporal o vencimiento.

La clasificación de los *STR* se realiza dependiendo de si se permite que la ejecución de una tarea finalice luego de su vencimiento o no. Si la ejecución termina en un instante posterior se los denomina *STR* blandos, si la ejecución debe terminar antes del vencimiento se los denomina *STR* duros o críticos. Si solo se permite una determinada cantidad de pérdidas de vencimientos por parte de cada tarea, se los denomina firmes.

Consecuentemente, una de las principales funciones del núcleo de un *SOTR*, es la planificación de las tareas a ejecutar garantizando, en principio, el cumplimiento de su vencimiento. Para esto, el planificador debe elegir a que tarea otorgar el derecho de ejecución en el microprocesador. Cada vez que una tarea se instancia o termina, el planificador debe examinar la cola de tareas listas y dependiendo de la política de ejecución implementada, elegir una para ejecutarla.

Por otro lado, para lograr una heterogeneidad de aplicaciones a ejecutar, es necesario utilizar el tiempo ocioso que deja libre el *STR* en su ejecución. Con este tiempo es posible atender aplicaciones no críticas y así ofrecer soporte para una diversidad de aplicaciones sobre el mismo dispositivo sin comprometer al *STR*.

En la actualidad, se puede encontrar ejemplos como: que un teléfono celular moderno puede mantener una agenda, tener juegos, sacar fotos, filmar videos, además de la principal aplicación de tiempo real que es digitalizar la voz y enviarla de manera consistente sin que las aplicaciones secundarias produzcan cortes o micro cortes en la transmisión.

En el último tiempo, ha surgido un nuevo e importante requerimiento que es el ahorro de energía en dispositivos móviles que funcionan a baterías. El incremento en la heterogeneidad de las aplicaciones, ha traído aparejado un importante incremento en la potencia de cálculo de los microprocesadores, para poder atender las necesidades de cómputo de estas nuevas aplicaciones.

Consecuentemente, la gran mayoría de los sistemas móviles necesitan técnicas y métodos de ahorro de energía que le permitan extender la carga de la batería de los mismos. De no ser así, el consumo de energía por parte del microprocesador a máximo desempeño, podría agotar la batería rápidamente, quedando el dispositivo sin utilidad (algunas técnicas puede consultarse en [1, 2, 3, 4]).

Para lograr esta heterogeneidad de aplicaciones, se necesitan herramientas de cálculo que permitan conocer, en tiempo de ejecución, cuánto tiempo ocioso existe y dónde se encuentra, para poder disponerlo por sobre el requerido por el *STR*. Existen diversas técnicas (algunas pueden verse en [2, 3, 4]) que permiten adelantar este tiempo ocioso, de manera de agruparlo y distribuirlo según sea necesario, logrando así flexibilizar al *STR*.

En particular, en los *STR* críticos que realizan planificación heterogénea, las tareas de tiempo real deben ser ejecutadas de manera obligatoria, respetando y garantizando que sus restricciones temporales se cumplan, dada la importancia de las mismas para la integridad del sistema o de la función del mismo.

Para poder realizar esto, es necesario garantizar a priori que el sistema pueda funcionar con esta heterogeneidad de aplicaciones, de manera que el planificador no ponga en riesgo las restricciones temporales que imponen las tareas de tiempo real por sobre las otras tareas ([5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17]).

Para satisfacer los requerimientos previamente planteados, el planificador debe contar con métodos y técnicas que calculen en tiempo de ejecución el tiempo ocioso disponible para la atención de tareas heterogéneas. De esta manera, el planificador puede anticipar el comportamiento del sistema y mantener la garantía de que las restricciones temporales de las tareas se cumplan, permitiendo brindar, por ejemplo, una *QoS* aceptable a las tareas heterogéneas.

Por otro lado, el tiempo de ejecución de las tareas no es el mismo en todas las ocasiones y aunque para el cálculo de los métodos antes mencionados se toma siempre el peor tiempo de ejecución, cuando este no ocurre, es necesario recuperarlo como tiempo ocioso, para que el sistema pueda utilizarlo.

Lamentablemente, garantizar tanto la ejecución de las tareas de tiempo real, como la ejecución de tareas heterogéneas (muchas de ellas complejas temporal y espacialmente) no es una labor sencilla. Trae aparejada la necesidad de incrementar la potencia de cálculo, no sólo para soportar la sobrecarga introducida por las nuevas aplicaciones, sino también la requerida para garantizar la factibilidad del *STR* heterogéneo en tiempo de ejecución.

Además, la eficiencia que estos sistemas tienen está aún muy lejos en varios aspectos de ser óptima y eficiente, y existen diversos problemas aún sin resolver de manera satisfactoria. Muchos de estos problemas son NP-Complejos, por lo cual su resolución es solo sub-óptima y es en general un compromiso entre la complejidad espacial (memoria utilizada por el método), la complejidad temporal (tiempo utilizado por el método) y los recursos con que cuenta el *SO* y el dispositivo.

En diversos centros de investigación en el mundo, este tema de trabajo es investigado de manera intensiva, ya que es fundamental proveer herramientas que permitan implementaciones tecnológicas que maximicen las prestaciones de los recursos computacionales, alcanzando así el menor costo posible. Además, un diseño adecuado puede permitir cumplir con los requerimientos establecidos por la aplicación, con otros adicionales como un correcto manejo de la calidad de servicio o el ahorro de energía y consecuentemente resulta de sumo interés para la industria.

Encontrar un apropiado balance entre una sencilla implementación, robustez y eficiencia, posibilita extender el ciclo de vida del software de tiempo real y mejorar las herramientas para el desarrollo de nuevas tecnologías. Por otro lado, es necesario contar con herramientas de validación de la especificación, diseño, optimización y testeo.

Finalizando, encontrar una integración del software y del hardware que permita satisfacer de manera óptima o subóptima lo planteado en el párrafo anterior será el tema principal de este proyecto de investigación.

#### 4 Líneas de Investigación y Desarrollo

Se trata de un proyecto de investigación científica y aplicada, con desarrollo experimental en el área de Sistemas de Tiempo Real, particularmente enfocado a la planificación de tareas con requerimientos heterogéneos y como se integran el software con el hardware. El principal campo de aplicación es la producción y tecnología de dispositivos de propósito dedicado y de propósito general con requerimientos de tiempo real.

#### 5 Resultados y Objetivos

El proyecto comienza a mediados de 2013, consecuentemente, a la actualidad no existen resultados del mismo. Sin embargo, se espera obtener buenos resultados del mismo.

Los objetivos más importantes se pueden resumir en:

- Investigar los *SOTR* de propósito general y dedicado y cómo implementan la administración de sus recursos, cumpliendo con las constricciones temporales que le imponen las tareas que conforman el sistema. También investigar las plataformas en las cuales estos *SOTR* se utilizan.

- Determinar la factibilidad y aplicabilidad de los métodos teóricos en los entornos prácticos estudiados. Proponer mejoras o nuevas técnicas y/o reformulaciones a las técnicas existentes para el manejo de recursos, temporales y espaciales.
- Implementar y validar las técnicas y métodos propuestos sobre plataformas de desarrollo concretas.
- Publicar las investigaciones realizadas. Además, fomentar, incentivar y difundir las tareas de investigación en la Universidad y consecuentemente mejorar la formación de recursos humanos, brindando gran participación a los alumnos del proyecto.

Como resultados del proyecto, además de los trabajos académicos y de divulgación científica, se espera mejorar la eficiencia de dispositivos de propósito dedicado desarrollando, mejorando e implementando nuevos métodos y técnicas de planificación dentro del núcleo de un Sistema Operativo de Tiempo Real.

Al momento de la redacción de este trabajo, en el proyecto que antecede a este, y que es base para la ejecución del mismo, se ha construido el simulador de *STR*, un generador de *STR* para alimentar al simulador y se han publicado varios trabajos en congresos nacionales e internacionales, así como trabajos en revistas ([18, 19, 20, 21, 22, 23, 24, 25, 26, 27]). Actualmente se está trabajando en la implementación de métodos de *Slack Stealing* en MarteOS (en PC), FreeRTOS (Microchip) y con el kit de Lego Mindstorm NXT 2.0 que seguramente su culminación serán parte de este proyecto.

#### 6 Formación de Recursos Humanos

En este proyecto de investigación participan 10 (Diez) docentes, 6 son docentes con el grado de doctor, 3 están en proceso de obtener un posgrado. Además, 5 (cinco) alumnos de la carrera Licenciatura en Informática de la Sede Puerto Madryn de la UNPSJB de los cuales 4 (cuatro) de ellos han comenzado sus tesis de grado en el marco de este proyecto que antecede y terminarán o publicarán su trabajo bajo este proyecto. Uno de los docentes ha obtenido una Beca Doctoral Tipo I de CONICET de comienzo en 2012, en temas afines a este proyecto y es dirigido por otros dos docentes del proyecto, uno de la UNS y otro de la UNPSJB.

#### 7 Referencias

- [1] B. Novelli, J. C. B. Leite, J. M. Urriza, and J. D. Orozco, "Regulagem Dinâmica de Voltagem em Sistemas de Tempo Real," in *XXXII Seminário Integrado de Software e Hardware (SBC 2005 SEMISH)*, Unisinos -Sao Leopoldo, Brazil, 2005.
- [2] J. M. Urriza, R. Cayssials, and J. D. Orozco, "Optimización del Cálculo del Tiempo Ocioso en Planificadores DVS con Tiempos de Ejecución Variables," in *XXXII Conferencia*

- Latinoamericana de Informática, Clei 2006*, Santiago, Chile, 2006.
- [3] J. M. Urriza, R. Cayssials, J. D. Orozco, and J. C. B. Leite, "Modelo de Tareas para recuperacion de Slack para Aplicaciones en Sistemas Enbebidos con DVS," Dep. de Ing. Eléctrica y Computadoras, Universidad Nacional del Sur, Argentina., Bahía Blanca, Reporte Interno22 de Mayo 2005.
- [4] J. M. Urriza, B. Novelli, J. C. B. Leite, and O. Javier Dario, "Economia de energia em dispositivos móveis," in *VI Workshop de Comunicação sem Fio e Computação Móvel*, Fortaleza, CE, Brasil, 2004, pp. 48–56.
- [5] R. I. Davis, "Approximate Slack Stealing Algorithms for Fixed Priority Pre-Emptive Systems," Real-Time Systems Research Group, University of York, York, England, Internal Report 1994.
- [6] R. I. Davis, "Dual Priority Scheduling: A Means of Providing Flexibility in Hard Real-Time Systems," Department of Computer Science, University of York, York, England, Internal Report1995.
- [7] R. I. Davis, K. W. Tindell, and A. Burns, "Scheduling Slack Time in Fixed-Priority Preemptive Systems," *Proceedings of the Real Time System Symposium*, pp. 222-231, 1993.
- [8] J. P. Lehoczky and S. Ramos-Thuel, "An Optimal Algorithm for Scheduling Soft-Aperiodic Tasks in Fixed-Priority Preemptive Systems," in *IEEE Real-Time Systems Symposium*, Phoenix, Arizona, EUA, 1992, pp. 110-123.
- [9] S. Ramos-Thuel and J. P. Lehoczky, "Algorithms for Scheduling Hard Aperiodic Tasks in Fixed-Prioriys Systems using Slack Stealing," in *Real-Time Systems Symposium*, 1994, pp. 22-33.
- [10] S. Ramos-Thuel and J. P. Lehoczky, "On-Line Scheduling of Hard Deadline Aperiodic Tasks in Fixed-Priority Systems," in *Real-Time Systems Symposium*, 1993, pp. 160-171.
- [11] B. Sprunt, L. Sha, and J. P. Lehoczky, "Aperiodic Task Scheduling for Hard Real-Time Systems," *The Journal of Real-Time Systems*, vol. 1, pp. 27-60, 1989.
- [12] J. M. Urriza, R. Cayssials, and J. D. Orozco, "A Fast Slack Stealing Method for embedded Real-Time Systems," Dep. de Ing. Eléctrica y Computadoras, Universidad Nacional del Sur, Argentina., Bahía Blanca, Internal ReportMay 31 2005.
- [13] J. M. Urriza, J. D. Orozco, and R. Cayssials, "Fast Slack Stealing methods for Embedded Real Time Systems," in *26th IEEE International Real-Time Systems Symposium (RTSS 2005) - Work In Progress Session*, Miami, EEUU, 2005, pp. 12-16.
- [14] R. M. Santos, J. M. Urriza, J. Santos, and J. D. Orozco, "New methods for redistributing slack time: applications and comparative evaluations," *The Journal of Systems & Software*, vol. 70-2, pp. 115-128, 2004.
- [15] J. M. Urriza, R. Cayssials, and J. D. Orozco, "Optimización on-line de Sistemas de Tiempo Real con Computación Imprecisa Basados en Recompensas.," in *32 JAIIO AST2003*, Buenos Aires, Argentina, 2003.
- [16] J. M. Urriza, J. Santos, and J. D. Orozco, "Un Algoritmo para la Diagramación de Tareas No-Duras mediante el Cálculo del Slack Time Disponible en cada Instante," in *XXIX Conferencia Latinoamericana de Informática. CLEI*, Bolivia, La Paz, 2003.
- [17] T.-S. Tia, J. W. Liu, and M. Shankar, "Aperiodic Request Scheduling in Fixed-Priority Preemptive Systems," Department of Computer Science, University of Illinois, Internal Report UIUCDCS-R-94-1859, 1994.
- [18] J. M. Urriza, J. D. Orozco, R. Cayssials, and L. Schorb, "Reduced Computational Cost in the Calculation of Worst Case Response Time for Real Time Systems," *Journal of Computer Science & Technology*, vol. 9, pp. 72-81, 2009.
- [19] J. M. Urriza, J. D. Orozco, C. Buckle, and R. Cayssials, "Ahorro de Energía en Dispositivos con un SO de Tiempo Real que planifican en RM o DM," in *Encuentro Chileno de Computacion*, Santiago, Chile, 2009.
- [20] J. M. Urriza, R. Cayssials, and E. Ferro, "Hardware Co-Processing Unit For Real-Time Scheduling Analsys," presented at the VI Southern Conference on Programmable Logic 2010 - Designer Forum 2010, Ipojuca Porto Galinhas Beach, Brazil, 2010.
- [21] J. M. Urriza, F. E. Paez, R. Cayssials, J. D. Orozco, and L. Schorb, "Low Cost Slack Stealing Method dor RM/DM," *International Review in Computers and Software (IRECOS)*, vol. 5, pp. 660-667, 2010.
- [22] G. Olguín, L. Biscayart, and J. M. Urriza, "Generador de Conjuntos de Tareas para Simulación en Sistemas de Tiempo Real," presented at the JAIIO 39 - Jornadas de Informática Industrial (JII) 2010, Buenos Aires, 2010.
- [23] C. E. E. Buckle, J. M. Urriza, and F. E. Paez, "Transitando Hacia las Bases de Datos de Tiempo Real," presented at the JAIIO 39 - Jornadas de Informática Industrial (JII) 2010, Buenos Aires, 2010.
- [24] L. Moreno, C. Geymonat, and J. M. Urriza, "Conceptos de Tiempo Real Aplicados a la Informática Industrial," presented at the CACIC 2011 - XVII Congreso Argentino de Ciencias de la Computación, La Plata, Arg., 2011.
- [25] F. E. Paez, J. M. Urriza, J. D. Orozco, and C. R. Buckle, "Un Modelo de Eventos Discretos para la Simulación de Sistemas de Tiempo Real," presented at the CACIC 2011 - XVII Congreso Argentino de Ciencias de la Computación, La Plata, Arg., 2011.
- [26] C. E. Buckle, J. M. Urriza, D. P. Barry, and F. E. Paez, "Tipo de Dato Abstracto para Sistemas de Bases de Datos de Tiempo Real," presented at the CACIC 2011 - XVII Congreso Argentino de Ciencias de la Computación, La Plata, Arg., 2011.
- [27] G. Olguin, L. Biscayart, and J. M. Urriza, "Generación de Tareas Periódicas y Aperiódicas para Simulación de Sistemas de Tiempo Real," presented at the JAIIO 40 - Jornadas de Informática Industrial 2011, Cordoba, Arg., 2011.



## Análisis de Perturbaciones de Sistemas de Tiempo Real en Aplicaciones de Control

Ricardo Cayssials<sup>2</sup>, José M. Urriza<sup>1</sup>, Francisco E. Páez<sup>1</sup>, Eduardo Schorb<sup>1</sup>, Lucas Schorb<sup>1</sup>, Sebastián Lucas<sup>1</sup>

<sup>1</sup> Depto. de Informática, Fac.de Ingeniería, Universidad Nacional de la Patagonia San Juan Bosco.

Puerto Madryn, Argentina  
+54 280-4472885 – Int. 117  
josemurriza@unp.edu.ar

<sup>2</sup> Depto. de Ingeniería Eléctrica y Computadoras - Universidad Nacional del Sur

Bahia Blanca, Argentina  
+54 291-4595000 - Int. 3372  
ricardo.cayssials@uns.edu.ar

### 1 Resumen

Este proyecto plantea el desarrollo de técnicas y mecanismos de planificación de tiempo real flexibles, que permitan, de acuerdo a las condiciones de trabajo de una aplicación en determinado instante, modificar las características de tiempo real del sistema. Principalmente se abarcará la utilización de sistemas de tiempo real en aplicaciones de control y procesamiento de señales.

**Palabras clave:** Sistemas de Tiempo Real, Aplicaciones de Control, Procesamiento de Señales, Sistemas Embebidos, Planificación de CPU.

### 2 Contexto

El proyecto de investigación *Análisis de perturbaciones de Sistemas de Tiempo Real en Aplicaciones de Control* es dirigido por el Dr. Cayssials de la Universidad Nacional del Sur (UNS) y codirigido por el Dr. Urriza de la Universidad Nacional de la Patagonia San Juan Bosco (UNPSJB). Las líneas de investigación del presente proyecto concuerdan tanto con las desarrolladas por los integrantes del *Real Time Systems Group*, perteneciente a la Facultad de Ingeniería

de la UNPSJB, Sede Puerto Madryn, como con las emprendidas por el *Laboratorio de Sistemas Digitales* del Departamento de Ingeniería Eléctrica y de Computadoras de la UNS, al cual pertenece el Dr. Cayssials. El proyecto es financiado por la Secretaría de Ciencia y Técnica de la UNPSJB.

### 3 Introducción

Tradicionalmente los Sistemas de Tiempo Real (*STR*) han sido concebidos para ser utilizados de forma casi exclusiva en grandes aplicaciones críticas en industrias automatizadas, en la exploración espacial, en dispositivos e instrumentación militar, en aviónica, sistemas de control, redes de comunicaciones, etc. Actualmente, sin embargo, los *STR* forman parte de creciente número de dispositivos electrónicos utilizados en el día a día. Se los puede encontrar tanto en dispositivos móviles (teléfonos celulares, *tablets*, GPS) como en las computadoras de los automóviles, etc. Se hace evidente que el dominio de recursos humanos y tecnológicos en el área de los *STR* es una necesidad imprescindible para cualquier nación industrializada en la actualidad.

Por lo general, los *STR* están constituidos por tareas periódicas. Las mismas deben finalizar su ejecución antes de un instante máximo denominado *vencimiento*. Este parámetro extra es parte de sus parámetros funcionales. Si

la tarea finalizara en un tiempo posterior a dicho instante, se dice que la misma ha perdido su vencimiento.

Los *STR* son clasificados dependiendo de si es permitido o no que la ejecución de una tarea finalice luego de perder su vencimiento. Así, si la ejecución puede finalizar en un instante posterior, el *STR* es denominado *blando*. Si bajo ninguna circunstancia se puede perder el vencimiento, el *STR* se clasifica como *duro* o *crítico*. Si únicamente una determinada cantidad de pérdidas de vencimiento es aceptable, el *STR* se denomina *firme*.

En la mayoría de las aplicaciones clásicas de *STR* previamente mencionadas, las decisiones del sistema provienen de algoritmos de control. Debido a que estas aplicaciones son consideradas *críticas*, estos *STR* son de tipo *duro*. Actualmente, sin embargo, los *STR* son utilizados en diversas áreas, cada una con requerimientos particulares [1]. Más específicamente, una aplicación puede requerir diferentes comportamientos del *STR* de acuerdo a las condiciones de funcionamiento que encuentre. Más aún, las restricciones temporales pueden ser sólo una entre varias condiciones de diseño a ser satisfechas. Por lo tanto, es necesario proponer mecanismos de planificación flexibles que permitan la adaptación del *STR* a diversas condiciones de trabajo, y al mismo tiempo minimicen el impacto en el resto de los requerimientos de diseño [2-4].

Generalmente, para garantizar el cumplimiento de los vencimientos, en los *STR duros*, las condiciones son conservadoras y diseñadas para el instante crítico de funcionamiento. En consecuencia, los sistemas resultan sobredimensionados debido al pesimismo introducido por estas condiciones. Este sobredimensionamiento genera una capacidad ociosa de procesamiento, que puede ser aprovechado por tareas denominadas *blandas* [5-7].

Muchas de las aplicaciones donde el uso de un *STR duro* es necesario, incluyen a los sistemas de control y el procesamiento de señales. El teorema del muestreo permite la utilización de sistemas de procesamiento digital en el mundo real, y en estos sistemas, la periodicidad de las tareas debe ser rigurosamente satisfecha a fin de garantizar una estricta relación entre el mundo analógico (real) y el sistema discreto (digital). De lo contrario, el *STR* introduce perturbaciones difíciles de modelar o predecir, que pueden imposibilitar recomponer digitalmente la naturaleza analógica de la aplicación. Para minimizar la influencia de estas perturbaciones, deben desarrollarse mecanismos de planificación que tengan en cuenta dichas perturbaciones. Esto es, adecuar los modelos de planificación utilizados en los *STR*, al modelo temporal planteado en la modelización discreta de aplicaciones de tiempo continuo.

Los efectos que produce la planificación de tareas en tiempo real, sobre diferentes aplicaciones, ha sido abordado en varios trabajos. En [8] se propone un algoritmo de asignación de prioridades que maximiza el desempeño teniendo en cuenta las características de control. El "*control server*", presentado en [9], facilita la implementación de tareas de control en *STR* flexibles. En

ambos trabajos se analizan las perturbaciones que un *STR* produce sobre una aplicación de control. Un modelo que separa las tareas de control en tres sub-tareas, con distintas características de tiempo real, es expuesto en [3]. Una estrategia de planificación, realimentada para múltiples tareas de control, se presenta en [10]. Esta estrategia optimiza el desempeño conforme a cual es la carga del sistema. Un estudio por medio de simulaciones donde se analizan los efectos sobre una aplicación de control es llevado a cabo en [11]. Todos los trabajos anteriores proponen modelos de tareas de control para reducir el *jitter*, pero ninguno refleja la perturbación que dicho *jitter* provoca sobre las condiciones de trabajo de una aplicación.

Por otro lado, algoritmos adaptativos *on-line* son propuestos en [4] y [10], que modifican el periodo de una tarea según la carga del sistema. Sin embargo, no se considera la degradación que producen sobre el desempeño de la aplicación.

Un mecanismo que modifica la estrategia de control en tiempo de ejecución, para compensar los efectos del *jitter*, se propone en [12]. En dicho trabajo los efectos variables del muestreo de entrada no son tenidos en cuenta. Cuando el sistema es implementado, al no ser continua la entrada, la modificación del período introduce un error de muestreo que no es considerado.

Este proyecto de investigación, busca resolver parte los problemas planteados en los párrafos anteriores, dentro de las diversas sub-disciplinas que poseen los *STR* y los sistemas de propósito dedicado (*embedded systems*). Para ello se analizarán los diferentes aspectos que provocan perturbaciones indeseadas en una aplicación, abordando métricas que las permitan cuantificar. Esto permitirá una evaluación eficiente de los *STR* empleados en aplicaciones de control, y proponer mejoras en las arquitecturas de procesamiento empleadas.

## 4 Líneas de Investigación y Desarrollo

Se trata de un proyecto de investigación científica y aplicada, con desarrollo experimental en el área de *Sistemas de Tiempo Real*, particularmente enfocado a la planificación flexible de tareas con requerimientos heterogéneos, en diferentes condiciones de funcionamiento y requerimientos externos. El principal campo de aplicación es en la industria de dispositivos de propósito dedicado, especialmente en aplicaciones de control industrial y procesamiento de señales.

## 5 Resultados y Objetivos

El proyecto se orienta principalmente a analizar y proponer tanto mecanismos como metodologías de planificación de tiempo real flexibles aptas para adaptarse a diferentes condiciones de funcionamiento y requerimientos externos, abarcando tanto aplicaciones de

control como de procesamiento de señales, en las cuales el comportamiento temporal debe satisfacerse de manera rigurosa, a fin de recomponer la naturaleza analógica de las aplicaciones.

La fecha de inicio del proyecto está estipulada para mediados de 2013, por lo cual no existen, a la actualidad, resultados directos del mismo.

Sin embargo, los objetivos específicos más importantes que se persiguen se pueden resumir en:

- Analizar y proponer métricas para la evaluación de las perturbaciones que los STR producen en el procesamiento de señales de control, a fin de realizar evaluaciones de eficiencia y proponer nuevos mecanismos de adaptación.
- Determinar la factibilidad de implementación de las metodologías propuestas en estructuras de diagramación tradicionales de tiempo real.
- Proponer arquitecturas de procesamiento adecuadas para el desarrollo de aplicaciones embebidas, que sumen valor a procesos productivos de la región.
- Publicar y difundir los resultados de las investigaciones realizadas en el marco del proyecto.
- Fomentar, incentivar y difundir las tareas de investigación en la Universidad.
- Interactuar con otros grupos de investigación en tareas conjuntas de investigación y desarrollo.
- Mejorar la formación de recursos humanos altamente calificados, con capacidades de investigación y desarrollo, brindando gran participación a los alumnos del proyecto.

Además de los trabajos académicos y de divulgación científica, se espera mejorar e implementar nuevos métodos y técnicas de planificación dentro del núcleo de un Sistema Operativo de Tiempo Real, a fin de mejorar la eficiencia de aquellos dispositivos de propósito dedicado controlados mediante este tipo de sistemas. Actualmente el grupo de investigación en STR cuenta con experiencia en MarteOS y FreeRTOS, entre otros.

## 6 Formación de Recursos Humanos

En este proyecto de investigación participan 3 (tres) docentes, dos de la UNPSJB y uno de la UNS, junto con 3 (tres) alumnos de la carrera Licenciatura en Informática de la Sede Puerto Madryn de la UNPSJB. Uno de los docentes ha obtenido una Beca Doctoral Tipo I de CONICET, de comienzo en 2012, en temas afines a este proyecto, siendo otro de los docentes del proyecto su co-director. Además, en 2010, este mismo docente obtuvo una beca anual de iniciación a la investigación de la Secretaría de Ciencia y Técnica e Innovación Productiva de la Provincia del Chubut, dirigida también por un docente del presente proyecto. Todos los alumnos participantes de este proyecto, tienen planeado enmarcar sus tesinas de grado dentro de la temática de los STR.

## 7 Referencias

- [1] Lui Sha, Tarek Abdelzaher, Karl-Erik Arzen, Anton Cervin, Theodore Baker, Alan Burns, Giorgio C. Buttazzo, Marco Caccano, John Lehoczky, and Aloysius Ka-Lau Mok, "Real Time Scheduling Theory: A Historical Perspective," *The Journal of Real-Time Systems*, vol. 28, pp. 101-155, 2004.
- [2] Giorgio Buttazzo and Luca Abeni, "Adaptive Workload Management through Elastic Scheduling," *Real-Time Syst.*, vol. 23, pp. 7-24, 2002.
- [3] Patricia Balbastre, Ismael Ripoll, Josep Vidal, and Alfons Crespo, "A Task Model to Reduce Control Delays," *Real-Time Syst.*, vol. 27, pp. 215-236, 2004.
- [4] M. Caccamo, G. Buttazzo, and Sha Lui, "Elastic feedback control," in *Real-Time Systems, 2000. Euromicro RTS 2000. 12th Euromicro Conference on*, 2000, pp. 121-128.
- [5] G. Bernat, A. Burns, and A. Liamsi, "Weakly hard real-time systems," *Computers, IEEE Transactions on*, vol. 50, pp. 308-321, 2001.
- [6] G. Bernat and Ricardo Cayssials, "Guaranteed on-line weakly-hard real-time systems," in *Real-Time Systems Symposium, 2001. (RTSS 2001). Proceedings. 22nd IEEE*, 2001, pp. 25-35.
- [7] S. K. Baruah, "A general model for recurring real-time tasks," in *Real-Time Systems Symposium, 1998. Proceedings., The 19th IEEE*, 1998, pp. 114-122.
- [8] P. Albertos, A. Crespo, I. Ripoll, M. Valles, and P. Balbastre, "RT control scheduling to reduce control performance degrading," in *Decision and Control, 2000. Proceedings of the 39th IEEE Conference on*, 2000, pp. 4889-4894 vol.4885.
- [9] A. Cervin and J. Eker, "The control server: a computational model for real-time control tasks," in *Real-Time Systems, 2003. Proceedings. 15th Euromicro Conference on*, 2003, pp. 113-120.
- [10] D. Henriksson and A. Cervin, "Optimal On-line Sampling Period Assignment for Real-Time Control Tasks Based on Plant State Information," in *Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC '05. 44th IEEE Conference on*, 2005, pp. 4469-4474.
- [11] Bjorn Wittenmark and A. Urquhart, "Adaptive extremal control," in *Decision and Control, 1995., Proceedings of the 34th IEEE Conference on*, 1995, pp. 1639-1644 vol.1632.
- [12] P. Marti, J. M. Fuertes, G. Fohler, and K. Ramamritham, "Jitter compensation for real-time control systems," in *Real-Time Systems Symposium, 2001. (RTSS 2001). Proceedings. 22nd IEEE*, 2001, pp. 39-48.

## Reconocimiento Automático de Patrones, Análisis de Imágenes y Generación de Características

Javier Giacomantone, Lucia Violini, Luciano Lorenti, Marcelo Naiouf,  
Oscar Bria, María José Abásolo, Cristina Manresa-Yee<sup>1</sup>

**Instituto de Investigación en Informática LIDI (III-LIDI)**  
**Facultad de Informática – UNLP**

{jog, lviolini, llorenti, mnaiouf, obria, mjabasolo, frapiva}@lidi.info.unlp.edu.ar

### CONTEXTO

Esta línea de investigación y desarrollo (I/D) forma parte del proyecto “Procesamiento Paralelo y Distribuido y Aplicaciones en Sistemas Inteligentes y Tratamiento de Imágenes y Video”, del Instituto de Investigación en Informática LIDI acreditado por la UNLP en el marco del Programa de Incentivos. Asimismo se integra al proyecto “A1/037910/11 Formación de Recursos Humanos e Investigación en el Área de Visión por Computador e Informática Gráfica (FRIVIG)”, financiado por la Agencia Española de Cooperación Internacional (AECID) dentro del Programa de Cooperación Interuniversitaria e Investigación Científica.

### RESUMEN

Este trabajo describe una línea de I/D y los resultados esperados de la misma. El objetivo principal es estudiar, desarrollar y evaluar métodos computacionales de reconocimiento estadístico de patrones en particular clasificación supervisada, no supervisada, reducción de dimensión y análisis de rendimiento. El segundo objetivo general es investigar métodos de generación de características a partir de señales en general e

imágenes digitales en particular, desde el punto de vista del método de clasificación subyacente, su poder de discriminación y su posible aplicación en tiempo real.

**Palabras Clave:** Reconocimiento de Patrones. Análisis de Imágenes. Generación de Características. Tiempo Real. Métodos de Agrupamiento. Núcleos Dispersos. Series.

### 1. INTRODUCCION

Reconocimiento de patrones y en particular reconocimiento estadístico de patrones es un área de investigación interdisciplinaria tanto en la investigación básica de métodos fundamentales [1][2][3], como en sus aplicaciones [4][5]. El objetivo principal de un sistema de reconocimiento automático de patrones es descubrir la naturaleza subyacente de un fenómeno u objeto, describiendo y seleccionando las características fundamentales que permitan clasificarlos en una categoría determinada. Sistemas automáticos de reconocimiento de patrones permiten abordar problemas en informática, en ingeniería y en otras disciplinas científicas, por lo tanto el diseño de cada etapa requiere de criterios de análisis conjuntos para validar los resultados [6][7]. Las principales áreas de aplicación son, reconocimiento remoto, reconocimiento óptico de caracteres y escritura manuscrita, identificación de patrones en imágenes médicas,

<sup>1</sup> Universitat de les Illes Balears. Unidad de Gráficos, Visión por Computador e Inteligencia Artificial.



sistemas de clasificación en bioinformática, sistemas de identificación biométrica y clasificación de series temporales. Un modelo general clásico de un sistema automático esta constituido por tres etapas, sensor, selector de características y clasificador. La primera etapa puede ser considerada a su vez como la que trata de obtener la representación más fiel del fenómeno estudiado, y un módulo que permite generar características del mismo. La etapa de generación de características aplica desde métodos simples directos hasta características obtenidas en forma indirecta de distinto tipos de señales como imágenes digitales 2D,  $2^{1/2}$  D y 3D.

### 1.1 Reconocimiento de Patrones

En esta línea de investigación el trabajo actual lo podemos clasificar en tres tópicos bien diferenciados correspondientes a tres sub-disciplinas dentro del área. La primera es clasificación supervisada donde el énfasis de nuestro trabajo es en métodos de clasificación basados en núcleos dispersos, en particular máquinas de soporte vectorial. La segunda es en clasificación no supervisada donde la principal línea de trabajo son las técnicas de agrupamiento basadas en teoría espectral de grafos y su aplicación a imágenes de rango. El tercer tópico de fundamental importancia es el de reducción de dimensión en particular selección de características. Las máquinas de soporte vectorial (SVM) [8] son herramientas fundamentales en sistemas de aprendizaje automático, permitiendo el tratamiento de problemas actuales en reconocimiento de patrones y minería de datos tales como, reconocimiento y caracterización de texto manuscrito, detección ultrasónica de fallas en materiales, clasificación de imágenes médicas [9], sistemas biométricos [10], clasificación en bioinformática [11] y en física de altas energías [12]. Las SVM implementan reglas de decisión complejas, por medio de una función no lineal que permite mapear los puntos de entrenamiento a un espacio de mayor dimensión. En el nuevo espacio de caracte-

rísticas las clases son separadas por un hiperplano, siendo este el que maximiza la distancia entre el mismo y los puntos de entrenamiento. La distancia se denomina margen y esos vectores son los vectores de soporte. Las SVM cumplen un rol muy importante en teoría de aprendizaje estadístico y cuando es necesario entrenar un clasificador no lineal en un espacio de características de considerable dimensión con un número limitado de muestras. Podemos diferenciar dos aspectos importantes que en general reciben la denominación de máquinas de soporte vectorial, el uso de SVM en clasificación SVC y el uso de las mismas en regresión SVR. La línea de investigación propuesta estudia ambos aspectos y en particular en el caso de clasificación mediante SVM tiene como objetivo diseñar sistemas con alta capacidad de generalización. Entre las tendencias actuales podemos mencionar las investigaciones sobre SVM paralelas y secuenciales (PSVM, SSVM) [13]. Una serie temporal es una secuencia de puntos medidos a intervalos sucesivos, normalmente de tiempo, y en general a intervalos regulares. Las series temporales son el resultado de medidas de distintos fenómenos físicos en la naturaleza pero también son comunes en econometría, marketing, control industrial y como resultado de métodos de monitoreo y diagnóstico en medicina. La clasificación de las series temporales obtenidas a partir de estudios funcionales del cerebro, son un ejemplo de abordaje multidisciplinario, donde uno de sus aspectos fundamentales es el de reconocimiento de patrones. La aplicación, adaptación y adecuada selección de kernels de SVM a series temporales es un tema de investigación actual [14][15]. Las series pueden ser unidimensionales o multidimensionales con correlación tanto temporal como espacial. En el último caso se plantea un problema de reconocimiento de patrones. Entre los temas actuales de investigación que involucran los conceptos anteriores podemos citar estudios en neurociencias por medio de resonancia magnética funcional (fMRI), electroencefa-

logramas (EEGs) y magnetoencefalogramas (MEGs) [16][17].

Los métodos espectrales de agrupamiento explotan la estructura de autovalores de la matriz de semejanza u otras matrices derivadas, generando una partición en agrupamientos disjuntos. El primer paso en la construcción de un algoritmo de clustering basado en teoría espectral de grafos es definir la matriz de similitud basada en los pesos de las aristas entre nodos [18]. En segundo término se construye el laplaciano del grafo y luego se realiza el correspondiente análisis de autovalores y autovectores [19]. En esta línea de I/D se investigan tres aspectos fundamentales, la construcción de la matriz de semejanza, el análisis de métodos de segmentación [20] y el estudio de alternativas para reducir la complejidad computacional para casos particulares [21]. El tercer tema central de la línea de I/D descrita en este artículo es el de reducción de dimensión, asociado directamente con problemas que presentan alta dimensión en sus patrones con respecto al número de muestras en el conjunto de diseño y los problemas en el que el número de parámetros es relativamente alto con respecto al número de patrones disponibles para el diseño del sistema de clasificación. Se consideran dos abordajes para resolver el problema de reducción de dimensión, extracción de características y selección de características. Se estudian dos tipos de problemas supervisados y no supervisados, y transformaciones lineales y no lineales del espacio de características [22].

## 1.2 Generación de Características

Un patrón es una entidad que podemos describir mediante un conjunto de características. Esta línea de investigación estudia no solo la adquisición directa del conjunto de características denominado descriptor sino los métodos de generación de características indirectos obtenidos a partir del análisis de distinto tipo de señales. El segundo objetivo es caracterizar la calidad de los mismos

mediante cinco parámetros principales: unicidad, congruencia, invariancia, abstracción y sensibilidad. La naturaleza de las imágenes digitales analizadas es amplia desde imágenes naturales obtenidas por reflexión o transmisión, imágenes multi-espectrales, imágenes de resonancia magnética o resonancia magnética funcional [23] como también imágenes de rango obtenidas mediante cámaras de tiempo de vuelo [24][25]. A pesar de trabajar en esta línea de investigación con imágenes generadas por distintos fenómenos y que representan problemas diferentes podemos identificar una metodología común. Es necesario segmentar la imagen, dividirla en sus partes constitutivas u objetos que la componen donde el nivel de segmentación depende del problema analizado. Una vez determinados los elementos de la imagen que componen los objetos de interés y el fondo es posible investigar que método, o proponer un nuevo método que permita obtener un descriptor [26][27]. En el caso de imágenes digitales podemos clasificarlos en externos, internos y mixtos según sean generados a partir de la frontera, la región del objeto de interés o ambas. El tipo de información que se utiliza es muy dependiente del problema abordado. Los principales métodos utilizados tienen en cuenta características como el color, la textura, el espectro de frecuencias, propiedades geométricas y análisis en distintas escalas. Se investigan métodos de generación de características que además de mejorar el rendimiento con respecto a los parámetros anteriores puedan también satisfacer restricciones temporales [28] para problemas específicos, como interfaces hombre-máquina no convencionales [29].

## 2. LINEAS DE INVESTIGACION Y DESARROLLO

- Clasificación supervisada. Discriminadores lineales y no lineales.
- Métodos de estimación de parámetros para clasificadores Bayesianos.
- Clasificación no supervisada. Técnicas de agrupamiento (clustering).

- Selección y extracción de características.
- Métricas, pseudométricas y distancias ultramétricas en clasificación supervisada, no supervisada y selección de características.
- Criterios de evaluación de desempeño en sistemas de clasificación automática.
- Criterios y algoritmos para combinación de clasificadores.
- Estudio de métodos para el tratamiento de datos en cámaras basadas en tiempo de vuelo.
- Máquinas de soporte vectorial. Kernels y algoritmos de optimización.
- Clasificación de series temporales y clasificación contextual.
- Generación de características para Interfaces Basadas en Visión
- Análisis de Señales Digitales

### 3. RESULTADOS OBTENIDOS /ESPERADOS

- Desarrollar modelos y optimizar algoritmos particulares de clasificación supervisada y no supervisada.
- Evaluación de los métodos de análisis de desempeño y su aplicación sobre los clasificadores y conjuntos de datos propuestos.
- Construcción de una mesa multi-táctil basada en visión por computador para su uso en educación especial [29].
- Se estudiaron y propusieron métodos para detección en series temporales de fMRI [23]
- Se realizó la evaluación de rendimiento en sistemas de reconocimiento de patrones supervisados y de clasificación binaria [16].
- Se propuso un método de segmentación de imágenes de rango [20]
- Obtener mejoras y adecuar las técnicas de selección y extracción para el tratamiento de datos en espacios multidimensionales.

- Promover la integración entre las distintas líneas de investigación en particular para aplicaciones en tiempo real.

### 4. FORMACION DE RECURSOS HUMANOS

En el marco de esta línea de investigación alumnos de grado y postgrado tienen la posibilidad de realizar tesinas y tesis en temas específicos. Actualmente hay dos investigadores realizando su doctorado y tres alumnos realizando tesinas relacionadas a aspectos particulares en sistemas automáticos de reconocimiento de patrones y análisis de imágenes.

### 5. AGRADECIMIENTOS

Esta línea de investigación es parcialmente financiada por el proyecto "A1/037910/11 Formación de Recursos Humanos e Investigación en el Área de Visión por Computador e Informática Gráfica (FRIVIG)" de la Agencia Española de Cooperación Interuniversitaria e Investigación Científica.

### 6. BIBLIOGRAFIA

1. Fukunaga K. "Introduction to Statistical Pattern Recognition". Second Edition. Academic Press, (1990).
2. Batagelj V, Bock H, Ferligoj A. "Data Science and Classification". Springer, (2006).
3. Devijer P, Kittler, J. "Pattern Recognition: theory and applications". Springer, (1986).
4. Anke Meyer-Base. "Pattern Recognition for Medical Imaging". Academic Press, (2004).
5. Kim H.Y., Giacomantone J. O., Cho, Z. H. Robust Anisotropic Diffusion to Produce Enhanced Statistical Parametric Map, Computer Vision and Image Understanding, v.99, pp.435-452 (2005).
6. Kim H.Y., Giacomantone J. O., A New Technique to Obtain Clear Statistical Parametric Map by Applying Anisotropic Diffusion to fMRI, IEEE, International Conference on Image Processing. Proceedings, Genova, Italy, v.3, pp.724-727 (2005).

7. Corte C, Vapnik V, Support vector networks. *Machine Learning* v.20, pp.273-297 (1995).
8. Vapnik, V. *The Nature of Statistical Learning Theory*. N. Y. Springer (1995)
9. Li S., Fevens L., Krzyzak A., Li S. Automatic Clinical Image Segmentation Using Pathological Modelling, PCA and SVM, MLDN, LNAI 3587 pp.314-324, (2005).
10. Lei Z., Yang Y., Wu Z.. Ensembles of Support Vector Machine for Text-Independent Speaker Recognition, *IJCSNS* v.6 n.5A pp. 163-167, (2006).
11. Boekhorst R., Abnizova I., Wernich L. Discrimination of regulatory DNA by SVM on the basis of over and under-represented motifs, *ESANN* pp. 481-486 (2008).
12. Vossen Anselm. *Support Vector Machines in High Energy Physics*, CERN, Geneva, Switzerland, pp.23-33 (2005).
13. Wang L., Chang M., Feng J.. Parallel and Sequential Support Vector Machines for Multi-label Classification, *International Journal of Information Technology*, v.11 n.9 pp. 11-18, (2005).
14. Rüping S., SVM kernels for time series analysis, *G1-Workshop-Woche Lernen-Lehren-Wissen-Adaptivitet*, pp.43-50 (2001).
15. Yang K., Shahabi C. A pca-based kernel for kernel pca on multivariate time series, *IEEE Intern. Conf. on Data Mining* (2005).
16. Giacomantone J., De Giusti A., ROC performance evaluation of RADSPM technique, *XIV Congreso Argentino de Ciencias de la Computación (CACIC)*, Chilecito (2008).
17. Chaovaitwongse W., Pardalos P. On the Time Series Support Vector Machine using Dynamic Time Warping Kernel for Brain Activity Classification, *Cybernetics and Systems Analysis* v.44 pp.125-138 (2008).
18. Von Luxburg U. A Tutorial on Spectral Clustering. *Statistics and Computing*, 17(4), (2007).
19. Shi J., Malik J. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(8), 888-905, (2000).
20. Lorenti L., Manresa-Yee C., Giacomantone J. Segmentación de objetos parcialmente ocluidos utilizando cámaras TOF. *X Workshop Computación Gráfica, Imágenes y Visualización*. pp. 387-395. Bahía Blanca, Buenos Aires, Argentina, Octubre, (2012).
21. Zare H., Shooshtari P., Grupta A., Brinkman R. Data reduction for spectral clustering to analyse high throughput flow cytometry data. *Bioinformatics*, 11-403, (2010)
22. Han Y., Feng X., Baciú G. Variational and PCA based natural image segmentation. *Pattern Recognition* 46, pp. 1971-1984 (2013).
23. Giacomantone J., Tarutina T. Diffuse Outlier Detection Technique for Functional Magnetic Resonance Imaging. *Computer Science and Technology Series. XVI Argentine Congress of Computer Science Selected Papers*. pp. 255-265 (2011).
24. Penne et al. Robust real time 3-D respiratory motion detection using time of flight cameras. *International Journal of Computer Assisted Radiology and Surgery* v.3 pp.427-431. (2008).
25. Kollors, A., Penne, J., Hornegger, J., Barker A. Gesture Recognition with a Time-of-Flight camera. *International Journal of Intelligent Systems Technologies and Applications*, v. pp.334-343, (2008).
26. Li J., Wang Y. Visual tracing and learning using speeded up robust features. *Pattern Recognition Letters* 33, pp. 2094-2101 (2012)
27. Prato M., Zanni L. A practical use of regularization for supervised learning with kernel methods. *Pattern Recognition Letters* 34, pp. 610-618 (2013).
28. Antunez E., Marfil R., Bandera A. Combining boundary and región features inside the combinatorial pyramid for topology preserving perceptual image segmentation. *Pattern Recognition Letters* 33, pp. 2245-2253 (2012).
29. Cristina Manresa-Yee, Ramón Mas, Gabriel Moyá, María J. Abásolo, Javier Giacomantone. Interactive multi-sensory environment to control stereotypy behaviours. *Computer Science & Technology*, pp. 121-128, (2012).



# Tópicos de Diseño de Aplicaciones de Tiempo Real

Fernando Romero, Diego Encinas, Luciano Iglesias, Armando De Giusti<sup>1</sup>,  
Fernando G. Tinetti<sup>2</sup>

Instituto de Investigación en Informática LIDI (III-LIDI)  
Facultad de Informática – UNLP

{fromero, dencinas, li, degiusti, fernando}@lidi.info.unlp.edu.ar

## RESUMEN

El objetivo de esta línea de investigación es el estudio y desarrollo de sistemas de software que poseen restricciones temporales, como son los Sistemas de Tiempo Real (STR) [4] [6] [15] haciendo hincapié en los aspectos de simulación, planificadores y comunicaciones.

## CONTEXTO

Esta línea de Investigación forma parte del proyecto "Arquitecturas multiprocesador distribuidas. Modelos, Software de Base y Aplicaciones" del Instituto de Investigación en Informática LIDI acreditado por la UNLP.

**Palabras Claves:** Tiempo Real, Sistemas Operativos de Tiempo Real, Sistemas Embebidos, Posix, comunicaciones, simulaciones, GNU/Linux.

## 1. INTRODUCCION

El diseño de aplicaciones de tiempo real [2] [3] implica aspectos que, de no ser cuidados, pueden producir una falla del sistema y no solo de la performance. En

particular, la planificación de tareas a fin de cumplir con los plazos de tiempo es una premisa esencial. Para investigar estos temas se dispone de herramientas de simulación como también sistemas reales. En esta línea, los desarrollos teóricos no siempre van acompañados de implementaciones en sistemas reales. A su vez, los cambios en el hardware implican nuevas posibilidades y variantes, tales como el hecho de disponer de varios núcleos de procesador en donde ejecutar tareas. Se estudia la implementación de cambios en un sistema operativo GNU/Linux con parche RT, apto para tiempo real.

La planificación impacta en todo el sistema con lo que surgen otros temas, tales como inversión de prioridades, lo cual también es tema de este trabajo. Los sistemas de tiempo real (STR) involucran tareas con restricciones de tiempo de diferente nivel de criticidad. Esto conlleva a un cuidadoso diseño de los sistemas operativos empleados, en particular al planificador del mismo, que será en gran parte responsable del cumplimiento de los plazos.

---

<sup>1</sup> Investigador CONICET

<sup>2</sup> Investigador CIC Prov. de Buenos Aires.

Comparando con un sistema tradicional, se debe tener especial control sobre la utilización de recursos, tiempos de respuesta conocidos, manejo de prioridades, fallos y comunicaciones.

Existen algunas diferencias importantes entre un sistema de software tradicional y uno de tiempo real. Las principales diferencias son:

- Planificación de tareas dirigida al cumplimiento de las restricciones de tiempo en vez de al rendimiento.
- Control de dispositivos externos, como por ejemplo, brazo robot, una planta industrial, etc., con un mayor uso de E/S.
- Procesamiento de mensajes que pueden arribar en intervalos regulares e irregulares.
- Seguridad y confiabilidad son conjuntas en sistemas de este tipo, debe haber detección y control de condiciones de falla a fin de evitar daños y un funcionamiento sin interrupciones. Debido a que muchas veces estos sistemas no cuentan con supervisión humana (autónomos) el manejo de excepciones es crítico. Se espera que un STR se ejecute en forma continua, automática y segura, teniendo un impacto en los costos de desarrollo y la seguridad.
- Modelización de condiciones concurrentes, ya que al mezclarse gran número de eventos regulares e irregulares es frecuente su concurrencia como la de los procesos cuya ejecución desencadenen dichos eventos, con lo que se debe tratar la alocaación y control de procesos concurrentes.
- Manejo de las comunicaciones en Sistemas Distribuidos permitiendo

reaccionar a los cambios de la red, como throughput o delay.

- Protección de datos compartidos.
- Por lo general se trata de Sistemas Embebidos que forman parte de un entorno mayor dentro del cual están inmersos y que lo controlan.

## 2. LINEAS DE INVESTIGACION Y DESARROLLO

Se plantean como temas de estudio:

- Durante la etapa de verificación y validación del hardware, pueden encontrarse medidas que impliquen la realización de cambios o modificaciones en los mismos. Una manera de disminuir la complejidad y la probabilidad de errores en la producción de hardware es desarrollar una simulación específica de éstos.
- Se está experimentando la implementación de sistemas de software sobre microcontroladores, utilizando Free RTOS y GNU/Linux con kernel RT en procesadores tipo ARM y x86.
- Políticas de planificación de CPU [1] [13]: en STR se trabaja en simulaciones de diferentes algoritmos de planificación de uso de la CPU, utilizando la herramienta Cheddar [7], con la que es posible simular procesos con diferentes restricciones temporales y prioridades, empleando diferentes planificadores. En base a estas simulaciones, se innovará planificaciones experimentales en GNU/Linux RT [14] [17]. En particular, se está trabajando en implementar la política EDF (Earliest Deadline First scheduling).
- Recursos compartidos: Inversión de prioridades: se llevan a cabo experimentos sobre SOTR, a fin de

evaluar ventajas y desventajas de los mecanismos implementados en los mismos para evitar este efecto.

- Se está simulando tráfico de red con restricciones temporales (streaming de audio y/o video), a fin de evaluar la conveniencia de los emisores [18] [19] [20].

### 3. RESULTADOS OBTENIDOS/ESPERADOS

De acuerdo con las tareas desarrolladas y a desarrollar, los resultados se enfocarán en varias direcciones relacionadas con sistemas operativos y con sistemas embebidos de tiempo real:

- Se estudian los sistemas operativos FreeRTOS y GNU/Linux RT [16], explorando sus posibilidades para el desarrollo de aplicaciones de tiempo real [9], realizando ejemplos concretos.
- Identificación y definición de pautas para estrategias en Planificadores y su impacto en los sistemas en STR, ensayando las mismas en Cheddar. En particular, el planificador linux con parche preemptive previsto para adaptarse al estandar Posix 1003 [5] [11].
- Uso de robots y simuladores de robots Kephra. Relación de la problemática de aplicaciones de robots y los sistemas embebidos de tiempo real.
- Los Sistemas Distribuidos de Tiempo Real Críticos deben ejecutar algoritmos en plazos de tiempo adecuados a los requerimientos de la implementación. Durante la etapa de verificación y validación del hardware, pueden encontrarse medidas que impliquen la realización de cambios o modificaciones en los mismos. Una manera de disminuir la complejidad y la probabilidad de errores en la

generación de hardware es desarrollar una simulación específica de éstos. Se propone el diseño e implementación de un modelo de simulación utilizando el framework SystemC [12].

- El modelo de simulación se realiza considerando un análisis de planificación de transmisión de mensajes [8] para redes CAN pero imponiendo restricciones propias del desarrollo de un simulador. La programación del modelo se efectúa utilizando la metodología Transaction Level Modeling (TLM) para describir al dispositivo a nivel de sistemas. La validación se efectúa utilizando especificaciones y normas estandarizadas del protocolo junto a herramientas de visualización de señales producidas por el simulador. De esta manera se cuenta con una herramienta para predecir el comportamiento del sistema de comunicaciones [10] ante, por ejemplo, posibles fallas.

### 4. FORMACION DE RECURSOS HUMANOS

En base a estos temas se están desarrollando un Doctorado y dos tesis de Magister en Redes de Datos y posibles tesinas de grado. También aportan trabajos de alumnos de la materia Diseño de Sistemas de Tiempo Real.

### 5. BIBLIOGRAFIA

[1] N.C. Audsley , A. Burns , M. F. Richardson , A. J. Wellings. "Hard Real-Time Scheduling: The Deadline Monotonic Approach". 1991. Proc. IEEE Workshop on Real-Time Operating Systems and Software.

- [2] L. Buhr. "An Introduction to Real Time Systems". Prentice Hall, 1999.
- [3] A. Burns & A. Wellings. "Real-Time Systems and Programming Languages". Addison Wesley, ISBN 90-201-40365-x/
- [4] A. Burns, A. J. Wellings. "Designing hard real-time systems". 1992. Proceedings of the 11th Ada-Europe international conference on Ada. ISBN:0-387-55585-4
- [5] Giorgio C. Buttazzo University of Pravia Italy. Rate Monotonic vs EDF, Real-Time Systems, 29, 5–26, 2005 2005 Springer Science + Business
- [6] Coulouris G., Dollimore J., Kindberg T. "Distributed Systems Concepts and Design", Addison Wesley 1994
- [7] The Cheddar project: A free real time scheduling analyzer, <http://beru.univbrest.fr/~singhoff/cheddar/>
- [8] R. Davis, «Controller Area Network (CAN) schedulability analysis: Refuted, revisited and revised.» Real-Time Systems. Springer, vol. 35, n° 3, pp. 239-272, 2007.
- [9] S. Ellison, "Developing Real-Time Embedded Software". Wiley, 1994
- [10] J. Eickhoff, Simulating Spacecraft Systems, Springer, 2009.
- [11] Martin Hervé, Bothorel Tests de faisabilité basés sur le taux d'occupation CPU des principaux algorithmes d'ordonnancement temps réel. Sebastien. Maitrise Informatique-année 2003-2004.
- [12] IEEE Standards Association, IEEE Standard for standard SystemC language reference manual, IEEE Computer Society, 2012.
- [13] C. L. Liu and J. W. Layland. Scheduling algorithms for multiprogramming in a hard-real-time environment. J. ACM, 20(1):46-61, 1973.
- [14] Robert Love Linux Kernel Development Third Edition. Addison Wesley. 2010. Media, Inc. Manufactured in The Netherlands.
- [15] A. Silberschatz, J. Peterson, P. Galvin, "Sistemas Operativos. Conceptos Fundamentales", 3rd Ed. Addison-Wesley Iberoamericana, 1994.
- [16] Siro Arthur. Carsten Emde. Nicholas Mc Guire. "Assessment of the Realtime Preemption Patches (RT-Preempt) and their impact on the general purpose performance of the system". 20 nov 2007. 9th RT Workshop.
- [17] Wolfgang Mauerer. Professional Linux@Kernel Architecture. Wiley Publishing, Inc. 2008.
- [18] D Miras. Network QoS needs of advanced internet applications a survey. Technical report, Internet2, November 2002.
- [19] William Stallings. High-speed networks and internets: performance and quality of service. Prentice Hall, 2002.
- [20] Cisco visual networking index: Forecast and methodology, 2011-2016 [Visual networking index (VNI)]. Technical report.



## Gestión Automática de incidentes e inventarios G.A.I.I.

Gibellini, Fabián<sup>1</sup>; Ruhl, Analía Lorena<sup>2</sup>; Di Gionantonio, M. Alejandra<sup>3</sup>; Rapallini, Marco<sup>4</sup>;  
Flores, Nora Viviana<sup>5</sup>; Serna, Mónica Mariel<sup>6</sup>

*Laboratorio de Ingeniería en Sistemas de Información (LabSis)  
Departamento de Ingeniería en Sistemas de Información  
Universidad Tecnológica Nacional, Facultad Regional Córdoba (UTN-FRC)*

(<sup>1</sup>)fgibellini@bbs.frc.utn.edu.ar; (<sup>2</sup>)lruhl@bbs.frc.utn.edu.ar; (<sup>3</sup>)ing.alejandradg@gmail.com;  
(<sup>4</sup>)marco@bbs.frc.utn.edu.ar; (<sup>5</sup>)ingnoraflares@gmail.com; (<sup>6</sup>)sernamonicam@gmail.com

### Resumen

*Uno de los problemas que enfrentan los grandes ambientes informáticos es conocer en tiempo real el parque informático del cual dispone, el estado de cada componente, las condiciones en las cuales se encuentran, y si están en reparación o desuso. [1] [2]*

*Actualmente, se han implementado algunas aplicaciones para inventariar los equipos y el software que forman parte del activo informático de diversos organismos. Estas aplicaciones permiten registrar en forma manual las distintas actualizaciones del inventario, ya sea por compras, bajas, reparaciones, etc.*

*Frente a esta problemática se diseñará una aplicación que permitirá implementar en forma automatizada el registro y control del inventario del hardware en cualquier establecimiento tanto público como privado, pero fundamentalmente, alertar ante el reemplazo y/o sustracción no autorizada de piezas en los equipos. Dicho Sistema de Alertas, será un subsistema del Sistema creado por el proyecto ya desarrollado durante período 2008-2010, Seguridad en Ambientes Informáticos (SAI) [3].*

### Palabras Claves

*Sistema de Inventario, seguridad, notificaciones automáticas, tiempo real, sistema de alerta, almacenamiento de hardware, ambientes informáticos, seguridad en tiempo real.*

### Contexto

Esta aplicación se desarrolla en los laboratorios de la UTN Facultad Regional Córdoba de la Universidad Tecnológica Nacional, donde se realizarán las primeras pruebas sobre su funcionamiento. Inicialmente se implementará en las aulas del Laboratorio de Sistemas y luego se hará extensivo a los laboratorios de otras especialidades. Este Sistema de Inventario automatizado, será un subsistema del Sistema creado por el proyecto ya desarrollado en el mismo Laboratorio durante período 2008-2010, Seguridad en Ambientes Informáticos (SAI) [3]. El mismo, fue acreditado en la Facultad Regional Córdoba de la Universidad Tecnológica Nacional (Argentina), siendo ésta, la institución financiadora del proyecto. Código: EIPRCO757.

## 1. Introducción

Es prioritario en todo ambiente informático implementar normas de seguridad para resguardar los datos que en él se manejan, sin embargo, no podemos dejar de pensar en todo el equipo instalado que da soporte físico, como las estaciones de trabajo, servidores, computadoras personales y sus componentes. Por lo que en el Laboratorio de Ingeniería en Sistemas de Información de la Universidad Tecnológica Nacional Facultad Regional Córdoba, nos encontramos trabajando en un Proyecto de Seguridad integral que cubra todos los aspectos posibles de considerar sobre este tema.

Uno de los factores que impulsó a llevar a cabo este proyecto fue el creciente número de estudiantes que hacen uso de las instalaciones del Laboratorio de Ingeniería en Sistemas de Información, de la Universidad Tecnológica Nacional –Facultad Regional Córdoba, lo que demanda la instalación de nuevas computadoras, enlaces y diversos equipos. Frente a esta realidad, muy satisfactoria, se dificultan las tareas de los administradores y la cantidad de recursos que se deben controlar. [4][5]

El crecimiento constante de la plataforma tecnológica para cubrir esta realidad, requiere ser acompañado por la implementación de un sistema integral de seguridad que permita conocer de manera sencilla y detectar el accionar de intrusos interesados en los componentes del hardware de los equipos del LabSis. El sistema GAI a desarrollar debe alertar ante cambios fuera de la configuración normal del hardware del equipo, reportando en tiempo real los cambios sufridos para una posterior revisión y toma de decisiones.

### 1.1 Integración con el sistema de Seguridad en Ambientes Informáticos (SAI)

El Sistema de Alertas, no es un sistema independiente, sino que el mismo es un subsistema que se integra al SAI. Este último, ya está desarrollado, y cuenta con tres subsistemas que conforman una suite de herramientas posibles de colaborar en la prevención y detección, sustracción y ataques a los equipos de un ambiente informático. Recordando que uno de los objetivos es lograr un sistema integral que sea de libre uso y de fácil acceso para cualquier organismo que lo requiera, haciendo hincapié en un sistema seguro y de muy bajo costo.

SAI está constituido por los siguientes subsistemas:

- Detección de Apertura de Gabinetes DAG [6].
- Sistemas alternativos de video vigilancia [7].
- Sistema Distribuido de Seguimiento Local y Remoto [8].

Este sistema integral SAI, a través de DAG permite detectar si se realiza la apertura de un dispositivo; el sistema de video vigilancia, permite el monitoreo a muy bajo costo; y el sistema remoto es la interfaz necesaria para el control de DAG.

Para optimizar y completar esta suite de herramientas, desarrollaremos una aplicación que permitirá obtener información de todos los componentes con los que se cuenta y mantenerlos actualizados de modo automatizado, apuntando a la posibilidad de alertar ante el reemplazo de alguno de ellos. Esta herramienta será escalable ya que no hay limitaciones en la cantidad de componentes a monitorear.

## 2. Líneas de Investigación y Desarrollo

El método utilizado para el presente trabajo de investigación está basado en el método empírico, de orientación cuantitativa observacional en la toma, análisis y

asociación de los datos cuantitativos arrojados por las aplicaciones bajo estudio. Pues, la informática utiliza los métodos empíricos, que toman conocimiento del objeto el uso de la experiencia [9]. Por tal motivo decidimos utilizar **OCS Inventory (Open Computer and Software Inventory)**[10], el cual es un software libre que se publica bajo la GNU GPLv2<sup>1</sup>, OCS-NG<sup>2</sup> recopila información sobre el hardware y software de equipos que hay en la red que ejecutan el programa de cliente OCS ("agente OCS de inventario"). Para analizar la forma de trabajo de la aplicación y determinar de qué manera nos servirán los datos que ésta nos entrega, instalamos la máquina virtual que se encontraba en la página oficial de OCS INVENTORY REPORT, "OCS-NG-Server-2.0.5-Debian-6.0.4-32bits.tar.gz" [11], la cual se ejecuta desde la aplicación VirtualBox [12].

### 3. Resultados y Objetivos

Los resultados obtenidos hasta el momento, luego de realizar todas las pruebas pertinentes dieron satisfactorias; en cuanto a la aplicación que se utilizará para desarrollar el módulo de alertas, el Sistema OCS ha sido la herramienta seleccionada para nuestro sistema.

El sistema GAII tomará los datos referente al hardware de la base de datos provista por el OCS-Server para generar un historial de inventarios, el cual nos proporcionará la información necesaria para generar reportes de manera automática de los cambios detectados en el hardware de la red.

<sup>1</sup>GNU GPLv2 (GNU: Is Not Unix)(GPL: General Public License)

<sup>2</sup>OCS-NG (Open Computer and Software Inventory Next Generation)

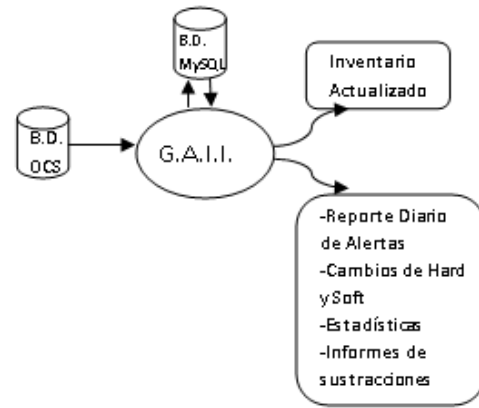


Fig. 1 Esquema básico del Sistema de Gestión Automática de incidentes e inventarios G.A.I.I.

#### Objetivo General:

Detectar y generar alertas sobre cambios y/o reemplazos no autorizados del hardware en equipos de computación, ubicados en aulas informáticas y ambientes públicos y/o privados, que requieran de un sistema de seguridad de bajo costo y fundamentalmente, de poca inversión inicial.

#### Objetivos Particulares:

- Brindar información actualizada de cada dispositivo tecnológico presente en el sistema del Labsis.
- Informar sobre los cambios en los históricos.
- Generar alertas vía mails al personal designado, para recibirlo y actuar en consecuencia.
- Generar alertas en tiempo real, enviadas por mail a otros dispositivos móviles.
- Detectar y documentar el estado real de las piezas de las computadoras después de cortes de energía eléctrica, o situaciones semejantes.
- Generar reportes con información que sea relevante para la toma de decisiones de la institución.
- Generar intercomunicación con el Sistema DAG[6]
- Aportar los datos para facilitar la gestión eficiente de incidentes e

inventarios, logrando la generación y reglas de asociación [13]

El mismo, nos brindará diariamente un reporte actualizado del contenido de cada PC, con sus movimientos, enviará un mail al administrador o a la persona encargada del lugar con el detalle de la alerta del cambio de algún elemento o modificación del equipo, inclusive si éste ha sido cambiado de lugar sin autorización o notificación del cambio.

#### 4. Formación de Recursos Humanos

Este proyecto contribuirá a la formación y crecimiento de la carrera de investigador de los integrantes del mismo.

Dentro de este último, se contempla la formación de alumnos de la carrera de Ingeniería en Sistemas de Información.

Por otro lado, se prevé la capacitación de alumnos becarios que formarán parte del equipo mientras el proyecto dure. De esta manera, se generará un grupo de trabajo importante que promocióne este tema.

Vemos también la posibilidad de colaborar con el crecimiento profesional de los integrantes del grupo, los cuales se ven predisuestos a perfeccionarse continuamente.

El grupo está compuesto por: Director, Co-Director, dos profesores investigadores de apoyo, tres profesores aspirantes a incorporarse a la carrera de investigador y un estudiante investigador de la carrera de Ingeniería en Sistemas de Información.

#### 5. Conclusión

Con el desarrollo e implementación de esta herramienta se espera obtener un producto de suma utilidad, tanto para el LabSis como para empresas privadas y/o públicas que permita tener el control inmediato del

hardware activo en la entidad y colaborar en la toma de decisiones y en la prevención, detección, sustracción y fallos de componentes de hardware.

Es importante destacar, que el sistema busca la integración de sistemas ya desarrollados e implementados y de sistemas que se encuentran en su etapa de desarrollo.

#### 6. Referencias

[1] Shyyunn Sheran Lin, Gregory S. Thompson, Viren Malaviya, "Embedded approach for device inventory collection utilizing OS programmability" SSTG, Cisco Systems 170 W Tasman Drive, San Jose, California, U.S.A. (sheranl@cisco.com, gst@cisco.com) IEEE 1M 2011 Cisco Systems Cisco.com

[2] N. D. Arnold and D. A. Dohan, "Connection-Oriented Relational Database of the APS Control System Hardware"

Argonne National Laboratory, Argonne, IL 60439, USA 2003

[3] Seguridad en Ambientes Informáticos (SAI).  
<http://www.jidis.frc.utn.edu.ar/papers/e7c362c8b5427c807ee23beab34d.pdf>

[4] Raydel Montesino and Stefan Fenz

"Automation possibilities in information security management"

Information Security Department  
University of Informatics Sciences (UCI)

Havana, Cuba, raydelmp@uci.cu

SBA Research and Vienna University of Technology, Vienna, Austria, sfenz@sba-research.org IEEE 2011



- [5] Álvaro Herrero, Emilio Corchado, Paolo Gastaldo, Francesco Picasso, Rodolfo Zunino, “Auto-Associative Neural Techniques for Intrusion Detection Systems”. Department of Civil Engineering University of Burgos C/ Francisco de Vitoria s/n, 09006 Burgos, Spain. Dept. of Biophysical and Electronic Engineering (DIBE)Genoa University Via Opera Pia 11a, 16145 Genoa, Italy. 2007
- [6] Detección de Apertura de Gabinetes  
<http://www.cneisi.frc.utn.edu.ar/papers/b998c93b46bb857450dfc6a89a03.pdf>
- [7] Sistemas alternativos de video vigilancia  
<http://www.cneisi.frc.utn.edu.ar/papers/b998c93b46bb857450dfc6a89a03.pdf>
- [8] Sistema Distribuido de Seguimiento Local y Remoto.  
[http://www.frsf.utn.edu.ar/cneisi2010/archivos/10-Sistema\\_Distribuido\\_de\\_Seguimiento\\_Local\\_y\\_Remoto.pdf](http://www.frsf.utn.edu.ar/cneisi2010/archivos/10-Sistema_Distribuido_de_Seguimiento_Local_y_Remoto.pdf)  
<http://laboratorios.fi.uba.ar/lie/Revista/Articulos/020205/A2ago2005.pdf>
- [9] Barchini, G. Métodos “I+D” de la Informática. Universidad Nacional de Santiago del Estero, Argentina. 2005.  
<http://laboratorios.fi.uba.ar/lie/Revista/Articulos/020205/A2ago2005.pdf>
- [10] OCS Inventory Reports  
<http://www.ocsinventory-ng.org/>
- [11] OCS Inventory  
<http://www.ocsinventory-ng.org/en/download/download-server.html>  
<http://wiki.ocsinventory-ng.org/index.php/Documentation:Main>
- [12] VirtualBox  
[www.virtualbox.org](http://www.virtualbox.org)
- [13] Uso de herramienta libre para la generación de reglas de asociación, facilitando la gestión eficiente de incidentes e inventarios  
[http://www.41jaiio.org.ar/sites/default/files/7\\_JSL\\_2012.pdf](http://www.41jaiio.org.ar/sites/default/files/7_JSL_2012.pdf)

## Optimización de un esquema “Occupancy Problem” orientado a E - Voting

Jeroen van de Graaf<sup>1</sup> ; Germán Montejano<sup>2,3</sup>; Pablo García<sup>3</sup>

<sup>1</sup>Departamento de Ciência da Computação  
Universidade Federal de Minas Gerais  
Av. Antonio Carlos, 6627 – 31270-010 - Belo Horizonte – Minas Gerais - Brasil  
Tel.: +55-3409-5836  
[jvdg@dcc.ufmg.br](mailto:jvdg@dcc.ufmg.br) – web: <http://www.dcc.ufmg.br/~jvdg>

<sup>2</sup>Departamento de Informática  
Universidad Nacional de San Luis  
Ejército de los Andes 950 – (5700) San Luis – San Luis – Argentina  
Tel.: +54-2652-424027 – Int. 251  
[gmonte@unsl.edu.ar](mailto:gmonte@unsl.edu.ar) – web: <http://www.unsl.edu.ar>

<sup>3</sup>Departamento de Matemática  
Universidad Nacional de La Pampa  
Av. Uruguay 151 – (6300) Santa Rosa – La Pampa – Argentina  
Tel.: +54-2954-425166– Int. 125  
[pablogarcia@exactas.unlpam.edu.ar](mailto:pablogarcia@exactas.unlpam.edu.ar) – web: <http://unlpam.edu.ar>

### RESUMEN

Este trabajo tiene el objetivo de presentar un enfoque alternativo al clásico “problema de la ocupación”, el cuál consiste en analizar las probabilidades relacionadas con disponer de una serie de bolillas que se colocarán, de manera aleatoria, en un conjunto de ranuras. El problema se generaliza con facilidad y resulta de sumo interés en múltiples aplicaciones prácticas. Desde una implementación informática, se piensa en un vector capaz de almacenar un dato en cada posición.

En particular, si pensamos en esquemas de voto electrónico, muchos modelos actuales basan su nivel de seguridad, en lo referente a la privacidad, en que la elección de la posición exacta en la que un voto se almacenará sea auténticamente aleatoria.

Tal exigencia, sin embargo, genera la imposibilidad de evitar que dos o más votos se alojen en una misma posición, lo cuál deriva en la pérdida de todos los sufragios coincidentes en una posición determinada del vector. Por lo tanto, minimizar la probabilidad de colisiones es un punto fundamental.

La aproximación teórica más conocida se conoce como “Birthday Paradox” y demuestra que para obtener niveles razonables de seguridad, es necesario implementar un vector de un tamaño significativamente mayor que la cantidad de votos que se deba administrar.

El presente trabajo entonces, propone un modelo alternativo, consistente en la aplicación de  $n$  vectores paralelos, con  $n \geq 2$ . Cada sufragio se replicará en todos los vectores, en posiciones aleatorias y potencialmente diferentes en cada caso. En tal esquema, un voto específico sólo se perderá si colisiona en todos los canales paralelos.

Para dejar clara la mejora que el esquema implementa, se realizan comparaciones en las que la cantidad total de slots implementados es la misma, pero distribuidos en dos o más vectores paralelos.

Se presentan fundamentos teóricos existentes y algunos resultados obtenidos con la utilización de un simulador, programado ad hoc, el cuál genera procesos eleccionarios, manejando los parámetros necesarios y entrega resultados sobre el comportamiento de las variables que se desea observar.

### PALABRAS CLAVE

*Colisión, slot, seguridad incondicional, anonimato, voto electrónico, Dining Cryptographers, Birthday Paradox, Occupancy Problem.*

### CONTEXTO

Este trabajo se enmarca en un convenio de intercambio entre la Universidade Federal de Minas Gerais (Brasil) y la Universidad Nacional de San Luis (Argentina). El mismo posibilita el intercambio de alumnos y docentes entre ambas instituciones, promoviendo, además, el desarrollo de proyectos comunes. En particular, corresponde a la articulación de la Maestría en Ingeniería de Software (UNSL) y el Mestrado em ciência da Computação (UFMG), ambos acreditados con la máxima categoría en Argentina y Brasil, respectivamente.

## 1. INTRODUCCIÓN

Los esquemas tradicionales de voto electrónico proveen, mayoritariamente, un nivel de seguridad computacional en lo referente al anonimato. Al mismo tiempo, proporcionan seguridad de tipo incondicional, durante el proceso de votación. En [8] se plantea que tal enfoque no resulta apropiado y debería invertirse, dado que la protección del proceso se lleva a cabo durante un tiempo limitado, ya que apenas termina la elección, los resultados se hacen públicos. En cambio, el anonimato debe protegerse indefinidamente. Es fácil imaginar las consecuencias de que se conozca la trayectoria, como votante, de un candidato que se postula en un proceso electoral actual.

Un esquema que busca dar respuesta a tal planteo se basa en implementar un vector de slots, capaces de almacenar votos. En los productos de software que adhieran a tal modelo, el anonimato podrá mantenerse de manera incondicional si la posición en la que se almacena cada voto es auténticamente aleatoria. En consecuencia, la probabilidad de que se produzcan colisiones (es decir, que dos o más votos se alojen en el mismo slot) presenta un valor mayor que cero.

Un antecedente teórico relacionado es conocido como *Occupancy Problem* y es profusamente abordado en los textos relacionados con la teoría de probabilidades. Por ejemplo, [3] lo describe detalladamente y [2] propone una fórmula matemática de aproximación basada en la distribución de Poisson. En cualquier caso, los valores obtenidos muestran que la relación slots – votantes debe mantenerse alta si se desea obtener niveles razonables de seguridad.

Por ejemplo, la muy conocida paradoja de cumpleaños (Birthday Paradox) se refiere al hecho, poco intuitivo, de que dado un grupo de 23 personas, la probabilidad de que dos de ellas cumplan años el mismo día es cercana a 0,5. Esto parece indicar que si se pretenden obtener niveles razonables de seguridad para un esquema de voto electrónico, la cantidad de almacenamiento que deberá implementarse será muy elevado en comparación con el tamaño de la muestra.

En particular, las aplicaciones de voto electrónico exigen altísimos niveles de seguridad. Debe tenerse en cuenta que un sistema de voto electrónico no resultará aceptable si no presenta ventajas significativas con respecto a los esquemas tradicionales, con boletas de papel. Y debe reconocerse que tales esquemas no presentan problemas severos en ese punto en particular.

En [8] se describe un esquema basado en un vector de slots para almacenar votos. Sin embargo, la aplicación de tal solución, tal como se observa en Birthday Paradox, resulta muy desfavorable si se desea generalizar el esquema para su aplicación en E-Voting, dado que el número de slots a implementar para proporcionar los niveles de

seguridad requeridos en tal aplicación crece de manera muy significativa. Sobre todo teniendo en cuenta los enormes niveles de exigencia de la aplicación, que define relaciones de poder muy importantes en las sociedades actuales.

Por lo tanto, el presente trabajo propone una variante que optimiza el almacenamiento que se destine a los votos. Consiste en implementar múltiples vectores iguales, de manera tal que la suma de los slots coincida con la del vector único que se implementa en el modelo original. La idea es que cada voto se replique en todos los vectores, pero en posiciones aleatorias potencialmente diferentes en cada caso.

Ante ese planteo, es evidente que un voto específico solamente será perdido si colisiona en todos los vectores implementados. Luego, se desea establecer si la probabilidad de pérdida es mejor en este escenario que en el original. Aparecen dos situaciones contradictorias:

1. El número de colisiones va a aumentar, dado que la totalidad de los sufragios se replicará en cada uno de los vectores implementados.
2. Por propiedad de los sucesos independientes, la probabilidad de que un voto se pierda efectivamente, es el producto de las probabilidades de que colisione en cada uno de ellos.

Se busca demostrar que la segunda propiedad tiene un peso mucho mayor y que, manteniendo ciertos límites en la relación entre la cantidad de sufragios y el tamaño de cada vector individual se obtendrán resultados muy superiores en términos de eficiencia.

## 2. EXPERIMENTOS REALIZADOS

Inicialmente se analiza la situación si se implementa un único vector. Se definen los siguientes parámetros:

$m$ : Cantidad de slots que se implementan.

$n$ : Cantidad de votantes.

Luego definimos el evento:

$X = \text{"se pierde el voto 1"}$ .

La probabilidad de  $X$  se define de la siguiente manera:

$$P(X) = 1 - P(\bar{X})$$

$$P(X) = 1 - P(x_i <> x_j \quad \forall j > 1)$$

$$P(X) = 1 - (1/m) (1/(m-1))^{n-1}$$

Se busca definir la probabilidad de que no se pierda ningún voto. Obviamente, la probabilidad de que un voto no sea efectivizado en el recuento final es similar para todos los casos.

Se define la siguiente variable aleatoria:

$Y =$  "Cantidad de votos perdidos".

La variable aleatoria  $Y$  responde a una distribución binomial, cuyos parámetros serán  $(m, P(X))$ . En consecuencia, la probabilidad de que al menos un voto no sea contabilizado es:

$$P(Y \geq 1) = 1 - P(Y=0)$$

$$P(Y \geq 1) = 1 - (1 - P(X))^m$$

$$P(Y \geq 1) = 1 - (1 - (1/m) \cdot (1/(m-1))^{n-1})^m$$

Sin embargo, tales desarrollos presentan la dificultad de que, aplicados a valores que puedan resultar de interés para la problemática de voto electrónico, no son calculables, dado que algunos de los valores involucrados son demasiado pequeños como para ser representados en los sistemas tradicionales.

Lo mismo ocurre con una aproximación propuesta en [2]. La aproximación, basada en la distribución de Poisson, muestra que para valores grandes de  $m$  y  $n$ , la siguiente ecuación es una aproximación aceptable.

$$P(n; \lambda) = e^{-\lambda} (\lambda^n / n!)$$

Tal resultado corresponde a una distribución de Poisson, en la que  $\lambda$  toma el siguiente valor:

$$\lambda = me^{-n/m}$$

Tal como se describió para fórmulas previas, aparece un inconveniente: no resulta accesible el cálculo del factorial involucrado en la fórmula si se desea aplicar a valores que resulten de interés. Por ejemplo, la versión de LibreOffice 3.5.7.2, en su planilla de cálculo Calc, sólo puede calcular factoriales hasta 170. Cabe mencionar que un recinto de votación electrónica en Brasil es de alrededor de 500 votantes.

Por lo tanto, el enfoque seleccionado para verificar el interés del enfoque consiste en la aplicación de simulaciones. Se implementa un simulador con las siguientes características:

Inputs:

- $m$ : cantidad de slots.
- $n$ : cantidad de votantes.
- $c$ : cantidad de canales paralelos.
- $r$ : cantidad de procesos electorarios que se simulan en la presente sesión del simulador.

Outputs:

- Votos efectivos acumulados ( $vea$ ): indica la cantidad exacta de votos que no se perdieron, teniendo en cuenta la totalidad de las corridas.
- Votos perdidos acumulados ( $vpa$ ): expresa la cantidad de votos perdidos en total. Obviamente:

$$nr = vea + vpa$$

- Cantidad de corridas con Pérdidas ( $cccp$ ).
- Cantidad de corridas sin pérdidas ( $ccsp$ ).

Es evidente que :

$$r = cccp + ccsp$$

- Media del generador aleatorio ( $mga$ ): este valor se obtiene a los efectos de verificar la calidad de la muestra generada.
- Porcentaje de Votos Perdidos ( $pvp$ ), referido a la totalidad de simulaciones de una sesión determinada.
- Mejor caso ( $mc$ ): indica cuál es el menor número de votos perdidos en algún acto electorario, considerando todas las corridas.
- Peor caso ( $pc$ ): idem al anterior para indicar el mayor número de votos perdidos en alguna corrida.

El programa utiliza un generador aleatorio que responde a una distribución uniforme. Su funcionamiento consiste en generar, para cada voto, una posición al azar para cada uno de los canales paralelos. Luego, al momento de colocar un voto se sigue una de las dos conductas siguientes:

- Si el slot está vacío, se guarda allí un número, correspondiente al número de orden del votante.
- Si el slot está ocupado, se suma a lo que ya contiene, una constante mucho mayor que el número de votantes, lo cuál permite detectar el número exacto de votos que colisionaron en un slot específico. Para todos los experimentos utilizados, dicha constante se fijó en 1.000.000, valor mucho mayor que el número de votantes utilizado (480).

### 3. RESULTADOS OBTENIDOS/ESPERADOS

Los resultados preliminares, basados en simulaciones, muestran una marcada optimización en la utilización del almacenamiento. A los efectos de ilustrar tal situación se muestran algunos resultados obtenidos. Para todos ellos se eligieron dos parámetros fijos:

- Cantidad de votantes: 480.
- Cantidad de corridas: 1.000.000.

En el primer caso, se implementaron 4800 slots y se vigila la variable "cantidad de corridas sin pérdidas". La figura 1 grafica los resultados obtenidos con la implementación de 1, 2, 4, 5, 6, 8, 10 y 12 canales paralelos. En la misma puede observarse que el valor óptimo se obtiene con la utilización de 6 u 8 canales.



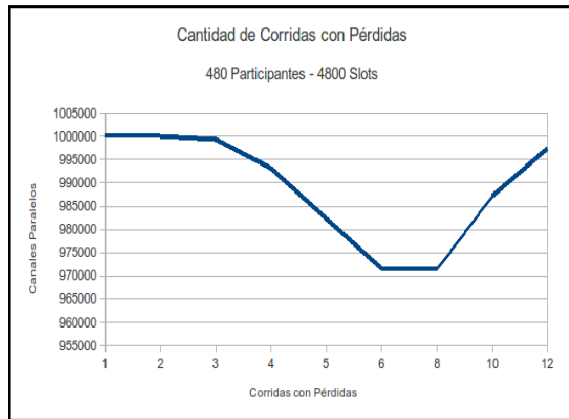


Fig1: Cantidad de Corridas con Pérdidas (4800 slots)

La mejora progresiva que se obtiene aumentando desde 1 a 6 canales es producto de la propiedad mencionada previamente relacionada con la independencia de los sucesos. Tal optimización se pierde cuando se utilizan más de 8 canales porque la cantidad de colisiones en cada canal unitario aumentan significativamente porque el tamaño de cada vector unitario es igual a la cantidad de votantes en el caso de utilizar 10 canales e incluso menor si se usan 12 canales, en cuyo caso cada canal tendrá sólo 400 slots, para 480 votantes.

El segundo ejemplo es similar, pero se implementan 9600 slots. Como muestra la figura 2, la eficiencia crece de manera continua con el agregado de canales. Por ejemplo, con un sólo canal, sobre un millón de corridas, no se perdieron votos solamente en 5. En cambio, con 12 canales, se registraron exactamente 971901 corridas en las que no se perdieron votos. La forma de la curva se modifica con respecto a la figura anterior, porque aún con 12 canales, la medida de cada uno, es siempre mayor que la cantidad de votantes (800 slots para 480 votantes si se usan 12 canales).

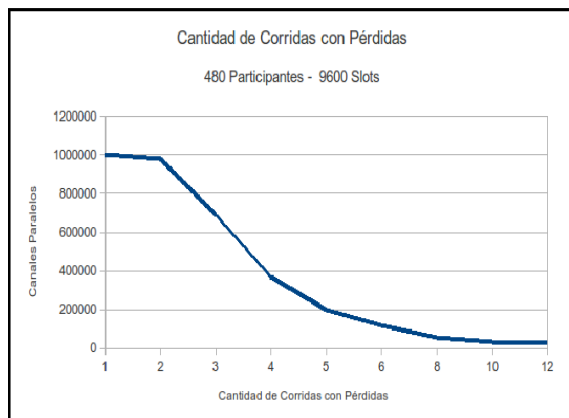


Fig 2: Cantidad de corridas con pérdidas (9600 slots)

Obviamente, los resultados mejoran si se aumenta el número total de slots. Sin embargo, el punto que se desea resaltar es la optimización obtenida, al incorporar más canales paralelos, con un mismo número total de slots.

Otra variable que se ha investigado es el comportamiento del peor caso, es decir, si se

realizan 1.000.000 de corridas, cuál es el mayor número de votos que se perdió en alguna de ellas. La figura 3 muestra significativos progresos al implementar vectores paralelos.

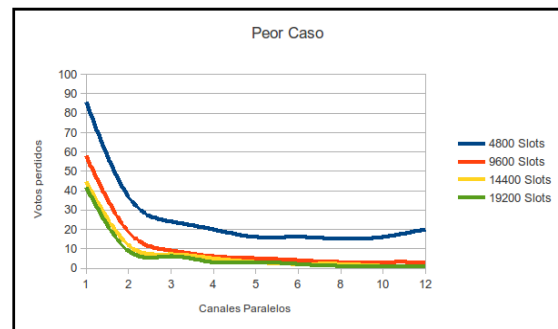


Fig. 3 – Evolución del peor caso

Otro valor que demuestra el nivel de optimización que proporciona el nuevo esquema, pasa por el total de datos efectivos acumulados tras 1.000.000 de simulaciones. Este magnitud implica cuantos votos se perdieron en total. En una simulación con 12 canales paralelos, 480 participantes y 19200 slots sólo se perdieron 38 votos, totalizándose 479.999.962 votos efectivos. Dependiendo de los valores de seguridad requeridos, los valores observados para 9600 y 14400 slots también podrían ser aceptables. La situación se ilustra en la figura 4.

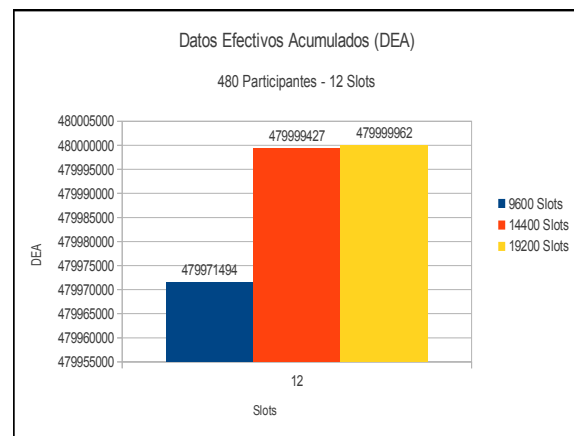


Fig. 4 – Datos efectivos acumulados

En definitiva, los resultados obtenidos muestran una mejora sustancial en el manejo de las colisiones, utilizando la misma cantidad total de slots.

Con respecto a los pasos futuros, mientras se está desarrollando este documento, se continúa investigando con el objetivo de obtener conclusiones matemáticas que permitan cuantificar de manera precisa el nivel en que mejora el esquema con la incorporación de vectores paralelos. Si bien las experiencias realizadas hasta el momento permiten predecir una optimización sustancial en la utilización del almacenamiento, se busca definir de manera formal cuáles son los niveles exactos de optimización que el enfoque aporta.

En definitiva, se desea, por ejemplo, responder de manera matemática formal preguntas como las siguientes:

- ¿Cuál es el número óptimo de vectores para una cantidad de votantes específico para una cantidad determinada de slots?
- ¿Cuál es el número mínimo de slots que debe implementarse y cómo deben organizarse los mismos para que la probabilidad de perder al menos un voto sea menor que un valor deseado?

También se desea obtener conclusiones definitivas relacionadas con posibles herramientas de recuperación post colisión. Por ejemplo, si en cada réplica de un voto determinado se coloca información relacionada con las ubicaciones del mismo voto en los demás slots paralelos, se puede aplicar una simple operación XOR para recuperar votos que hayan colisionado en todas sus instancias, siempre que al menos una de tales colisiones sea simple y con otro sufragio del cuál no se hayan perdido todas las instancias.

Se espera obtener conclusiones formales de todos los tópicos mencionados.

#### 4. FORMACIÓN DE RECURSOS HUMANOS

El tema presentado está siendo desarrollado con mayor nivel de detalle en la tesis de maestría de uno de los autores (García), en la cuál los otros dos (van de Graaf y Montejano) son director y codirector, respectivamente. La misma se titula "Optimización de un Esquema de Dinig Cryptographers Asincrónico" y sera defendida durante el presente año en el ámbito de la Universidad Nacional de San Luis (UNSL) para obtener el título de Magister en Ingeniería de Software. Dicha maestría está calificada como "A" en el ámbito de la CONEAU. Además de eso, García realizó un intercambio durante todo 2012 en el ámbito de la Universidade Federal de Minas Gerais (UFMG), en Belo Horizonte, Brasil, a través de una beca CAPES - CAFPBA, obteniendo allí los créditos necesarios para una doble titulación, por lo cuál dicha institución otorgará el título de Mestre em ciência da Computação, carrera de posgrado que también tiene la máxima calificación en el ámbito de la educación superior brasileña. Se espera mantener y profundizar la relación entre ambas instituciones.

También se busca derivar otras tesis de maestría que presenten de manera rigurosa esquemas que apliquen estas técnicas en un esquema concreto de voto electrónico.

La tesis mencionada abarca otra problemática que deja abierta una línea de investigación: la obtención de métodos probadamente eficientes en esquemas que aseguren la seguridad incondicional del anonimato de quienes intercambian mensajes en una red. Ese punto debería derivar en una tesis doctoral que aborde de manera rigurosa enfoques alternativos en ese sentido.

#### 5. BIBLIOGRAFÍA

1. Chaum D.: "The Dining Cryptographers Problem: Unconditional Sender and Recipient Untraceability". *Journal of Cryptology*. 1988.
2. Feller W.: "An Introduction to Probability Theory and its Applications". Volumen I. Tercera Edición. John Wiley and Sons. New York, 1957.
3. Flajolet P., Gardy D., Thimonier L.: 'Birthday paradox, coupon collectors, caching algorithms and self-organizing search'. *Discrete Applied Mathematics* 39, ps. 207-223. North-Holland. 1992
4. Fuster Sabater A., De La Guía Martínez, D., Hernández Encinas L., Montoya Vitini F., Muñoz Masqué J.: "Técnicas Criptográficas de Protección de Datos". 3a Edición actualizada. ISBN: 978-84-7897-594-5. Editorial Ra - Ma. 2004.
5. Mao W.: "Modern Cryptography: Theory and Practice". Prentice Hall -ISBN: 978-0132887410. 2003.
6. Meyer P: "Probabilidad y Aplicaciones Estadísticas". Addison Wesley Iberoamericana. Segunda edición.
7. Trappe W., Washington L.: "Introduction to Cryptography with Coding Theory". Prentice Hall. ISBN: 0-13-061814-4. 2002.
8. Van de Graaf J.: "Anonymous One Time Broadcast Using Non Interactive Dining Cryptographer Nets with Applications to Voting". Publicado en: "Towards Trustworthy Elections". Ps 231-241. Springer-Verlag Berlin, Heidelberg. ISBN:978-3-642-12979-7. 2010.

# APLICACIÓN DE SISTEMAS BIOMÉTRICOS EN LA ADMINISTRACIÓN PÚBLICA LOCAL PARA PROTECCIÓN DE LA INFORMACIÓN

Graciela Etchart<sup>1</sup>, Carlos Alvez<sup>1</sup>, Marcelo Benedetto<sup>1</sup>, Miguel Fernández<sup>2</sup>

<sup>1</sup>Facultad de Ciencias de la Administración - Universidad Nacional de Entre Ríos

Av. Tavella 1424, Concordia, Entre Ríos - CP 3200

[getchart@fcad.uner.edu.ar](mailto:getchart@fcad.uner.edu.ar), [caralv@fcad.uner.edu.ar](mailto:caralv@fcad.uner.edu.ar), [marben@fcad.uner.edu.ar](mailto:marben@fcad.uner.edu.ar)

<sup>2</sup> Municipalidad de Concordia – provincia de Entre Ríos

Mitre 76 - Concordia, Entre Ríos (3200) - [migfer@concordia.gov.ar](mailto:migfer@concordia.gov.ar)

## Resumen

Los entes estatales tienen la necesidad de proteger tanto bienes materiales como información de diverso tipo, ya sea en formato digital o de otra índole. Por esto es menester, contar con un alto nivel de seguridad a través de mecanismos eficientes y eficaces de control de acceso a las zonas restringidas donde se encuentran los bienes a proteger.

En el proyecto se trabajó el caso de la Municipalidad de Concordia, y en particular los sectores destinados al funcionamiento de la Dirección de Informática y al área de Tesorería ya que debido a los riesgos a los que está expuesto el sistema informático de estas áreas, ha sido considerado como prioritario por parte de la Gestión Política.

En lo que a acceso a ambos sectores se refiere, se realizará un análisis comparativo entre dos tecnologías de control de acceso biométrico para la identificación de las personas que accedan al sector de servidores y a Tesorería.

**Palabras clave:** Seguridad, Sistemas Biométricos, Administración Pública Local, Identificación de Personas.

## Contexto

Este trabajo está encuadrado en el acta acuerdo entre la Municipalidad de Concordia y la Facultad de Ciencias de la Administración de la UNER, suscripto en el marco del proyecto de investigación PID UNER 7035 “Identificación de personas mediante sistemas biométricos. Estudio de factibilidad y su implementación en organismos estatales”.

La Dirección de Informática, a través de su personal, se encargará de la adecuación del ambiente físico y del hardware y el personal de la UNER, del análisis de diferentes tecnologías de control de acceso a utilizar. Entre las acciones del personal municipal, se encarga de

tareas como instalación de equipos y medidas de seguridad ambientales (control de incendios, control de temperatura, etc.).

El personal de la UNER realizará un análisis comparativo entre dos tecnologías de control de acceso biométrico para la identificación de las personas, un relevamiento y un diagnóstico de las dificultades de seguridad que se presentan en el personal que accede al sector de servidores. Uno de ellos es un sistema de reconocimiento biométrico basado en la geometría de la mano y, la otra alternativa es un sistema de autenticación basado en dos modalidades biométricas: huella y rostro.

## Introducción

La importancia de utilizar tecnologías biométricas radica en el hecho de que los mecanismos tradicionales como las claves de acceso y tarjetas magnéticas son muy falibles. Por ejemplo, las tarjetas magnéticas se pueden extraviar o sustraer, las claves se pueden olvidar (o ser observadas por alguien más), etc. Por esto, es importante emplear sistemas de reconocimiento biométricos para identificar a las personas y de esta manera reforzar la seguridad de la institución.

## Líneas de Investigación y Desarrollo

El proyecto desarrolló un estudio comparativo de los sistemas biométricos. El mismo, fue realizado tomando como base a un dispositivo mono-biométrico (geometría de la palma de la mano) contra un dispositivo multi-biométrico adquirido en el segundo año de ejecución del proyecto (de Rostro y Huella Dactilar).

Para la elección de los rasgos a utilizar, se tuvieron en cuenta principalmente las características de Fiabilidad, Resistencia a ataques, Aceptabilidad y Costo.

## Objetivos y resultados

- Efectuar un estudio comparativo de los sistemas biométricos. El mismo, fue realizado tomando como base a un dispositivo mono-biométrico (geometría de la palma de la mano) contra un dispositivo multi-biométrico (de Rostro y Huella Dactilar).
- Adecuar las instalaciones con las dimensiones y características necesarias para instalar los servidores y los distintos dispositivos de seguridad que posibiliten brindar un marco de resguardo y protección de la información.
- Implementar un sistema confiable de control de acceso de personas a la oficina de servidores y al área de Tesorería, que permita la identificación y el registro de los accesos.

Del análisis realizado entre las distintas alternativas, desde el punto de vista técnico y financiero, se consideraron apropiados y se adquirieron los siguientes dispositivos:

- HANDPUNCH 1000: Lector de geometría de la mano, con capacidad para el registro de 512 usuarios y 5120 transacciones en memoria. Permite el reconocimiento 1 a 1 (verificación).
- ZKSOFTWARE IFACE 202: Equipo de reconocimiento de huella dactilar y facial, con capacidad de 700 rostros, 3000 huellas dactilares y hasta 100.000 fichajes de capacidad autónoma. Permite el reconocimiento 1 a 1 (verificación), como 1 a n (identificación). También permite el uso de tarjeta de proximidad.

Este equipamiento fue instalado en las dependencias de Tesorería y la Dirección de Informática, áreas estas que por su función ya fueron previamente definidas como estratégicas en el proyecto en lo que a Identificación de personas se refiere.

Ambos dispositivos fueron sometidos a pruebas de acceso durante un mes por el personal involucrado en dichas oficinas y, en base a las características analizadas, se pudieron establecer los siguientes resultados comparativos:

Sistema Biométrico	Geometría de la mano (mono-biométrico)	Huellas y (multi-biométrico)
Característica		
Fiabilidad	Alta	Muy alta
Resistencia a ataques	Alta	Muy alta
Aceptabilidad	Alta	Muy alta
Costo	Alto	Medio

Por lo expuesto, en el caso estudiado, con los dispositivos involucrados y con las pruebas realizadas, se concluye que el sistema multi-biométrico presenta actualmente mayores beneficios para su implementación.

La razón de ello, radica en que al incorporar dos sistemas de identificación en un mismo dispositivo, los parámetros de medición mejoran notablemente el rendimiento y la exactitud de la comparación.

Cabe destacar además y de acuerdo a los relevamientos efectuados durante el desarrollo del proyecto, que los dispositivos de geometría de la mano en un futuro podrían perder en la industria, el mercado que actualmente poseen. Las razones, radican fundamentalmente en que los mismos tienen un costo alto, una permanencia media del rasgo de lectura y que los factores ambientales lo afectan en forma negativa respecto a otros dispositivos.

## Entrevistas realizadas

A los efectos de relevar datos significativos sobre la visión del personal vinculados con los temas de seguridad, y los riesgos a los que se encuentran expuestos a situaciones que pudieran comprometer el normal funcionamiento de sus tareas; se conceptuó importante abarcar dos sectores considerados claves para los objetivos del proyecto: el área de informática y el sector de tesorería.

Para dicho relevamiento se confeccionó un formulario de entrevista, con 10 ítems que incluyeron temáticas tales como gestión de la información, protección de los datos, normas de seguridad, y cuestiones relacionadas con los sistemas y dispositivos existentes para la identificación de las personas con acceso a dichos sectores.

Fueron entrevistadas doce (12) personas del área informática ( 90% de la planta total) y seis (6) personas del sector tesorería ( 75% del total).

Se detalla a continuación los resultados de las entrevistas, producto de la compilación efectuada, resaltando en cada caso los aspectos más importantes de las opiniones vertidas por los encuestados.



### Resultados de las entrevistas.

*¿En qué soporte se encuentra la información que maneja su área (digital, papel, etc.)?*

En la oficina de Informática se manifestó que el 80% de los datos se encuentran soportados en dispositivos digitales y el resto consiste en documentación impresa, particularmente aquella que respalda tareas sensibles y que requiere de autorizaciones de niveles superiores.

En la oficina de Tesorería los encuestados manifestaron que en promedio el 50% de la información es provista mediante acceso a terminales de computación (medio digital), pero que el 50% restante conforman un archivo donde se almacenan los comprobantes de respaldo necesarios para el funcionamiento de dicha oficina.

*¿ Conoce los procedimientos o normas de seguridad que posee el organismo?*

En el sector informático solamente una persona manifestó desconocer normas de seguridad, por lo que el 92% manifestaron estar informados sobre algunas normativas y/o procedimiento en este sentido. Seis encuestados, de los once en total, expresaron que las normas no son suficientemente completas y que no se encuentran debidamente sistematizadas, ya que han sido emitidas por diferentes niveles jerárquicos y aparecen un tanto desarticuladas.

En el sector Tesorería el 100% de los encuestados expresaron conocer procedimientos y/o normas de seguridad, aunque con distintos grados de opinión sobre si las mismas eran suficientes. Es interesante destacar como dato complementario, que la mitad de los encuestados (50%) opinan que las normativas no contemplan en forma integral los aspectos de seguridad.

*¿Existe un manual de procedimiento para la seguridad de la organización?. ¿Se encuentra disponible y es de fácil acceso?.*

La totalidad de los encuestados del área informática manifiesta que no existe un manual específico que sistematice las diferentes normas de seguridad del sector. Remiten su opinión a la pregunta anterior.

En el sector Tesorería todas las personas coincidieron en expresar que no existe un manual. Las normativas se encuentran en

carpetas de “Resoluciones” firmadas por Directores y Secretarios. Solamente una persona manifestó no conocer el lugar de archivo mientras que las cinco restante expresaron su conocimiento y facilidad de acceso.

*¿Qué perjuicio ocasionaría una eventual pérdida de equipamiento o información?.*

En el sector de Informática todos los encuestados manifestaron diversos tipos de acontecimientos que podría originar la pérdida de los equipos instalados y particularmente de la información contenida en ellos. Existe suficiente conciencia sobre la magnitud de los riesgos considerando que, por ejemplo, en los servidores se encuentra soportada toda la información del municipio. El 83% expresó que si bien existen sistemas de respaldo de todos los datos, la caída de los sistemas puede originar en algunos casos la paralización de las actividades de todo el municipio por un tiempo variable en función de las características del hecho.

En el sector Tesorería, las seis personas encuestadas expresaron que la pérdida de equipamiento informático retrasaría las tareas por el tiempo que demanda su reemplazo. En idéntica magnitud respondieron que la pérdida de la documentación soportada en papel, podría originar inconvenientes al momento de ser necesario respaldar fehacientemente algún acto administrativo.

*¿Existe algún control para las personas que acceden al lugar físico de trabajo?.*

En el caso del Sector de informático hubo coincidencia en las respuestas a esta pregunta separando dos sectores:

Sala de servidores. Posee una sola puerta de acceso con una cerradura común como única medida de protección, que permanece cerrada y solo se la habilita para las tareas de mantenimiento. La llave de la puerta está en poder de los responsables del sector quienes controlan el ingreso de las personas.

Sala de Desarrollo y Operaciones. Posee dos puertas con cerraduras comunes que permanecen abiertas solamente durante el horario de trabajo ( de 07:00 a 13:00 hs.). Todos los encuestados manifestaron que no existe ningún tipo de control dentro de dicho horario.

En el sector tesorería existe una puerta única de acceso de madera con vidrio esmerilado. Los

seis encuestados manifestaron conocer el sistema de control con sistema de apertura eléctrica que es accionado desde el interior de la oficina, ante la solicitud de ingreso de una persona. No existe normativa alguna para autorizar los ingresos, ya que el mismo personal de tesorería acciona la cerradura y permite el ingreso de personas que son de su conocimiento personal.

*¿Existe alguna medida de seguridad para acceder a la información?*

En la oficina de informática, el 100% de las personas responden afirmativamente para el caso de la información digitalizada. Existen medidas que exigen la identificación del usuario para acceder a cualquier dato informatizado, como así también las claves de identificación. Respecto de la información soportada en papel, ocho casos de los encuestados expresan que la documentación archivada en carpetas y biblioratos comunes se ubican físicamente en armarios comunes, sin cerraduras, a los que cualquier persona.

En el sector tesorería todos los encuestados señalaron que para acceder a la información “grabada en la computadora” debían previamente acceder con una clave que solamente era conocida para cada una de las personas. Tres personas (el 50%) manifestó que resultaba “tedioso” la exigencia de cambiar las claves en forma periódica, tal como se les exige. Para el caso de la documentación guardada en archivos, alguna información se encuentra bajo llave, pero mucha otra no tiene medidas de control de acceso (dentro del personal afectado al sector).

*Conoce que son los métodos biométricos para identificación de personas?*

En el sector informático todos los encuestados expresaron conocer algún tipo de metodología biométrica para la identificación de personas.

Por su parte, en el sector tesorería, el 100% expresó que conocía el sistema que tiene instalado el municipio para el control de ingreso y egreso del personal (huella dactilar).

*¿ Cuáles conoce ?:*

En esta pregunta se detallaban algunos sistemas y los encuestados debían marcar los que conocían. Se resume las respuestas de cada sector:

Sector informático: todos los encuestados manifestaron conocer los sistemas consignados. En el ítem “Otros” un 50% detalló sistemas multi-biométricos.

En el sector tesorería manifestaron conocer: Huella dactilar: 6 personas (100%); Voz: 1 persona (17%); Palma: 2 personas (33%); Iris: 1 persona (17%).

*¿Cree que un sistema de identificación de las personas basado en biometría contribuiría a dar seguridad a su área?*

A esta pregunta, en ambas oficinas respondieron afirmativamente la totalidad de los encuestados.

*A su criterio, qué cualidades o características debería tener un sistema de identificación de personas?*

En esta pregunta se detallaban algunas características y los encuestados debían marcar las que les parecían pertinentes. Se resume las respuestas de cada sector:

Sector informático: Universalidad: 10 personas (83%); Fiabilidad: 12 personas (100%); Facilidad de administración: 8 personas (67%); permanencia de rasgos: 8 personas (67%).

En el sector tesorería: Facilidad de operación: 6 personas (100%), Comodidad: 4 personas (67%).

En base observaciones realizadas en las instalaciones, relevamiento efectuado a través de entrevistas a informantes claves y al personal de Tesorería y Dirección de Informática del Municipio de Concordia respecto a la identificación de personas que acceden a su oficinas, se detectó la siguiente problemática:

- Carencia e los niveles de protección de los equipamientos informáticos afectados a la Dirección de Informática del Municipio. El equipamiento actualmente no cuenta con un ambiente físico adecuado como tampoco con mecanismos de acceso seguros. Adecuación del ambiente físico y del hardware como ser la instalación de equipos y medidas de seguridad ambientales (control de incendios, control de temperatura, etc.) y el análisis de diferentes tecnologías de control de acceso a utilizar.
- Desconocimiento y falta de normas referidas al acceso e identificación de personas a dependencias claves en la administración pública local por parte de sus agentes.

- Falta de inversión en tecnología de control de acceso biométrico para la identificación de las personas que accedan a los sectores antes mencionados. La importancia de utilizar tecnologías biométricas radica en el hecho de que los mecanismos tradicionales como las claves de acceso y tarjetas magnéticas son muy falibles. Por ejemplo, las tarjetas magnéticas se pueden extravíar o sustraer, las claves se pueden olvidar (o ser observadas por alguien más), etc. Por esto, es importante emplear sistemas de reconocimiento biométricos para identificar a las personas para reforzar la seguridad.

### Formación de Recursos Humanos

Se brindó a los integrantes del proyecto, formación en lo que se refiere a sistemas de identificación de personas basados en la biometría.

Los integrantes del proyecto, se desempeñan en cátedras relacionadas directamente o indirectamente con el tema central de la investigación, por lo que este trabajo tendrá impacto directo e inmediato en la docencia.

Se procederá a dirigir becarios de investigación, así como también tesis finales de grado, dirigidos por el director del proyecto de investigación y/o por los docentes-investigadores del mismo. Para estos casos, también se prevé la presentación a convocatorias de becas ante organismos provinciales y nacionales.

Los integrantes participarán de reuniones científicas y técnicas que permitan actualizar los conocimientos en el tema de interés. También se trabajará en la presentación de trabajos en congresos nacionales e internacionales relacionados con el área del proyecto. Estos trabajos servirán para divulgar los conocimientos obtenidos durante el trabajo de investigación.

Uno de los principales objetivos del proyecto es que el personal docente de la UNER dedicado al mismo avance y/o concluya con sus estudios de posgrado, así como también se incorporen becarios realizando investigaciones en temas afines a la temática del proyecto.

### Referencias

1. Wing B. ANSI/NIST-ITL 1-2011. Information Technology: American National Standard for Information Systems Data Format for the Interchange of Fingerprint, Facial & Other Biometric Information. November, 2011.
2. Carlos E. Alvez, Marcelo G. Benedetto. "Los Sistemas Biométricos y su Factibilidad de Aplicación en los Organismos Estatales". XII Workshop de Investigadores en Ciencias de la Computación (WICC 2010), El Calafate, Santa Cruz Argentina, 5 y 6 de Mayo de 2010, Universidad Nacional de la Patagonia San Juan Bosco. Páginas 247-251.
3. Graciela Etchart, Lucas Luna, Carlos Leal, Marcelo Benedetto, Carlos Alvez. Sistemas de reconocimiento biométricos, importancia del uso de estándares en entes estatales. CGIV - XIII Workshop de Investigadores en Ciencias de la Computación (WICC 2011), 5 y 6 de Mayo de 2011. Universidad Nacional de Rosario. Rosario - Argentina. Páginas 339-343.
4. Graciela Etchart, Lucas Luna, Rafael Leal, Marcelo Benedetto, Carlos Alvez. "Sistema adecuado a estándares de reconocimiento de personas mediante el iris". CGIV - XIV Workshop de Investigadores en Ciencias de la Computación (WICC 2012), 25 y 26 de Abril de 2012. Universidad Nacional de Misiones. Posadas - Argentina. Páginas 321-325.
5. Alvez Carlos E. Modelos para la recuperación de imágenes por similitud en Bases de Datos Objeto-Relacionales. Tesis Doctoral. Santa Fe, Argentina, Abril de 2012. ISBN 978-987-33-2249-5.
6. Casal Gabriel, Rovolta Mercedes. Biometrías. Herramientas para la Identidad y la Seguridad Pública. Jefatura de Gabinete de Ministros. Presidencia de la Nación. Noviembre de 2010.
7. Julio Fuoco, Tendencias Biométricas, desafíos y oportunidades. En Biometrías 2. Jefatura de Gabinete de Ministros. Presidencia de la Nación. Octubre de 2011.
8. Alvez C., Benedetto M., Berón G., Etchart G., Luna L. y Leal C. Desarrollo de un sistema multi-biométrico mediante reconocimiento de iris y voz, adecuado a estándares, para su aplicación en organismos públicos. SIE 2011 - Simposio de Informática en el Estado. Córdoba, 31 de Agosto, 01 y 02 de Septiembre de 2011. 40° JAIIO. Páginas: 206 - 220.

## Reconocimiento flexible de formatos de archivos para el descubrimiento de evidencia digital

Gabriel Casullo<sup>(1)</sup>, Erika Givaudant<sup>(1)</sup>, Leandro Tami<sup>(1)</sup>, Leopoldo Sebastián Gómez<sup>(2)</sup>

<sup>(1)</sup> Facultad de Ingeniería  
Universidad FASTA

info@casu-net.com.ar – erikalaura@gmail.com – leandro.tami@gmail.com

<sup>(2)</sup> Laboratorio Pericial Informático  
Poder Judicial del Neuquén  
sebastian.gomez@jusneuquen.gov.ar

### CONTEXTO

Esta línea de investigación es definida a partir de una necesidad detectada en el área de especialidad de informática forense. Se pretende contribuir a profundizar el análisis forense para el descubrimiento de evidencia digital mediante el reconocimiento flexible del formato de archivos a partir de sus características estructurales y de contenido.

El proyecto en curso contribuye a fortalecer los vínculos de cooperación entre la universidad y el ámbito profesional. Concretamente se apoya un trabajo académico de grado impulsado desde la Facultad de Ingeniería de la Universidad FASTA en colaboración con el Laboratorio Pericial Informático del Poder Judicial de la Provincia del Neuquén.

### RESUMEN

Se presentan avances en el estudio de un método computacional robusto para la identificación del formato de archivos, con capacidades especiales para la detección de aquellos formatos que puedan tener daños mínimos por fallas en el medio de almacenamiento o transmisión, o bien cuya estructura sea modificada intencionalmente con el objeto de obstruir una investigación digital. En una primera etapa se compara el encabezado del archivo a identificar con una base de datos de firmas de formatos de archivo conocidos utilizando una métrica de distancia y confeccionando un ranking mediante una lista de formatos candidatos con aquellos que

obtuvieron las menores distancias. La segunda etapa tiene por objeto profundizar el análisis del contenido de archivos, extrayendo metadatos y características para una posterior aplicación de técnicas de inteligencia computacional que permitan aumentar la precisión de la identificación. Los métodos utilizados son implementados y evaluados mediante experimentación sobre diferentes corpus digitales.

*Palabras clave:* magic numbers, header obfuscation, transmogify, file types, file fragments, content-based file type identification.

### 1. INTRODUCCION

Los archivos digitales tienen el potencial de servir de evidencia inculpatória o exculpatória y pueden aportar elementos de prueba relevantes para el esclarecimiento de un hecho delictivo. Resulta de vital importancia poder identificar el formato de los archivos que se someten a análisis forense para conocer preliminarmente mayores detalles sobre su estructura y contenido. Por ello, durante el desarrollo de actividades periciales esta tarea es uno de los primeros exámenes que el investigador realiza sobre el material probatorio.

En condiciones normales la identificación consiste simplemente en observar la extensión de un archivo para obtener a priori una primera descripción general de su contenido, pero dichos atributos son poco fiables ya que



pueden ser fácilmente modificados. Las herramientas forenses actuales permiten la identificación automática del formato de un archivo mediante un procedimiento habitualmente reseñado como "análisis de firmas". Este método computacional es independiente de la extensión del archivo, y consiste en verificar si un conjunto de bytes pertenecientes al encabezado del mismo está presente en una base de datos de firmas ya conocidas para diversos formatos de archivo.

A pesar de ser de gran ayuda para la detección de ocultamientos de archivos mediante modificaciones en la extensión, las firmas de archivo son también vulnerables a cambios accidentales o intencionales en el encabezado, y en estos casos dichas herramientas forenses no tolerantes a errores fracasan en su identificación. El "pattern matching" puede fallar en aquellos casos que el encabezado haya sido alterado aunque sea por pequeños cambios a nivel de bit. Por ello resulta necesario el desarrollo de un método de reconocimiento más flexible que permita obtener un espectro de formatos "candidatos", posibilitando al investigador intentar el acceso a la información digital contenida en el archivo a través de la aplicación que sea más apropiada para la manipulación de los mismos.

## 2. LINEAS DE INVESTIGACION y DESARROLLO

Actualmente existen diferentes enfoques científicos para el reconocimiento automático de formatos de archivos. Se han estudiado métodos de identificación basados en el contenido [1], incluso cuando el repositorio digital está corrupto, utilizando funciones de extracción jerárquicas que sirvan al PCA (Principal Component Analysis) juntamente con un entrenamiento de redes neuronales auto-asociativas, obteniendo porcentajes muy favorables en el reconocimiento.

Otros investigadores [2] han optado abordar esta problemática mediante un método computacional de clasificación de fragmentos de archivos, considerando aspectos de su estructura interna para una identificación efectiva. En algunas situaciones esto

demuestra no ser una tarea sencilla, ya que el reconocimiento suele basarse en la aparición de patrones pero a veces los archivos carecen de dichas señas identificadoras por la aplicación de métodos de compresión o cifrado sobre su contenido. En el recorrido que llevan a cabo de los métodos computacionales disponibles se destaca la librería *libmagic*, base del comando *file* de UNIX. Dicha utilidad es efectiva para la identificación de archivos completos, pero no tiene los resultados esperados en el tratamiento de fragmentos. Una aproximación a la solución es el análisis de distribución de secuencias de bytes, también conocido como "análisis de 1-gramo". Este método consiste en construir un histograma con cada carácter ASCII posible detectado en el archivo. Estas frecuencias constituyen una huella o "fingerprint", y la pertenencia a un tipo específico de archivo se determina según la distancia entre la huella generada y las de una base de datos de huellas conocidas.

Desde otra perspectiva de análisis estadístico se han propuesto dos algoritmos para predecir el tipo de archivo a partir del análisis de un fragmento perteneciente a dicho contenedor digital [3]: el primero está basado en el discriminante lineal de Fisher mientras que el segundo utiliza las subcadenas y subsecuencias comunes de caracteres de un fragmento. El discriminante lineal de Fisher es un método estadístico utilizado para clasificar individuos en grupos. Hay una fase de entrenamiento en la cual los datos de individuos pertenecientes a grupos conocidos se utilizan para armar un modelo de clasificación. Este modelo está compuesto por un conjunto de funciones lineales, una para cada grupo. A partir de esta información, el modelo es capaz de predecir el grupo al que pertenece un individuo de acuerdo a la función de clasificación que devuelva el valor más alto. El algoritmo basado en subcadenas y subsecuencias de caracteres comunes presupone que dos archivos del mismo tipo probablemente tengan en común cadenas de caracteres más largas comparados con otros archivos de diferente estructura o contenido. Los resultados empíricos corroboran tales

afirmaciones aún en archivos con ausencia de metadatos.

Se han presentado trabajos científicos [4] que promueven el uso de la métrica de similaridad del coseno como evaluador de distancia en las funciones de extracción de patrones de bytes y el uso de un esquema que aplica pasos recursivos en clusters para la identificación de archivos, obteniendo resultados satisfactorios.

### 3. RESULTADOS OBTENIDOS/ESPERADOS

Como inicio de las actividades de investigación se intentó un abordaje inicial a través del reconocimiento del *header* de los archivos para obtener una mejora sobre el ya conocido “análisis de firmas”. En una primera instancia se procuró la localización de bases de datos de firmas identificadoras de archivos. Se utilizó como estructura básica de información la base de firmas *TrID* [5] que cuenta con 4000 firmas de distintos formatos de archivo. Se hicieron pruebas de medición de distancias entre los *headers* de archivos de prueba y las contenidas en la base de firmas anterior utilizando la distancia de Levenshtein. Esta métrica es de gran utilidad ya que permite medir la distancia de entradas de longitudes desiguales. Pese a ello, existe una gran proporción de falsos positivos debido a que la base de datos de *TrID* contiene muchos tipos de archivos comparten los mismos encabezados.

FIRMA	TIPO	EXTENSION
21 42 44 4E	Outlook Exchange Offline Storage	OST
21 42 44 4E	Microsoft Personal Address Book	PAB

Luego de una depuración se obtuvieron 2734 firmas únicas y se implementó un analizador básico para el reconocimiento del formato de archivos. Los *headers* repetidos corroboran que dicho análisis no es suficiente para la identificación del formato de un archivo desconocido y que son necesarias nuevas fuentes de información, como por ejemplo la presencia o ausencia de metadatos de formatos

de archivo conocidos. La identificación de archivos con la herramienta forense se comportó satisfactoriamente aunque es intolerante a errores, equivalentemente a lo que sucede con la utilidad *file* de sistemas UNIX.

Para las tareas de experimentación se generaron lotes de prueba mediante la implementación de una aplicación informática complementaria. Dicha aplicación utiliza un diccionario y realiza una selección de palabras en forma aleatoria para luego realizar consultas automáticas sobre los buscadores *Google*, *Bing* y *FindThatFile* con un formato requerido pasado como parámetro. Para cada resultado de la consulta la aplicación accede desde el enlace al sitio donde se aloja el archivo y lo descarga. Finalmente se genera una lista de las extensiones de los archivos obtenidos. Para contar con un corpus digital sin falsos positivos, se corroboró que la extensión de los archivos descargados se correspondiera con sus tipos verdaderos. Para ello, se usó como valor de referencia la salida retornada por la herramienta *file* y se estableció una paridad entre la salida de dicha herramienta y los tipos MIME correspondiente a cada formato de archivo.

La experiencia aportada desde la práctica profesional ha determinado la necesidad de concentrar la efectividad de la aplicación forense para la identificación de formatos de archivos hacia aquellos que habitualmente son más proclives a contener evidencia digital. Se han priorizado formatos específicos de imagen (png, jpg, bmp), de video (avi, mpg), audio (wav, wma, mp3), de procesamiento de texto y planillas de cálculo (pptx, ppt, doc, docx, xls, xlsx), de base de datos (mdb) y otros compuestos en formatos propietarios (pdf, psd).

A partir del corpus digital de archivos a identificar y la lista depurada de firmas de formato de archivos que han sido determinadas como prioridad para el análisis forense, se desarrollaron herramientas para medir la distancia entre los *n* primeros bytes de un archivo y cada una de las firmas de la base. Si A y B son dos firmas de longitudes diferentes

y A es un prefijo de B, se considerará más representativa a la firma de mayor longitud. Para ello, se define el 'ratio' como la relación entre distancia y longitud, y se utiliza este criterio para el ordenamiento de los resultados de la identificación.

Complementariamente se implementó una herramienta forense que efectúa dicho reconocimiento sobre grandes lotes de archivos, incorporando errores aleatorios en los encabezados. Cada identificación produce un tipo MIME que es comparado con el que se obtiene al utilizar el comando *file*, y se determina automáticamente si el resultado ha sido correcto o si no se logra el reconocimiento del formato.

De los resultados obtenidos de la experimentación se verifica que: a) Un daño de 1 byte en una firma corta, puede ser demasiado significativo como para permitir la identificación de un formato. Por ejemplo, un daño de 1 byte en la firma de un archivo BMP, representa un 50% de daño, lo cual invalida la identificación con éste método. b) Existen tipos de archivo que comparten la misma firma, por ejemplo ZIP y JAR. El método utilizado por el analizador básico solo acota el conjunto de posibles formatos a la hora de identificarlos.

Ambas aseveraciones hacen necesaria una segunda etapa de identificación que tendrá que basarse en el contenido del archivo, considerando la extracción de características particulares y la búsqueda de presencia de metadatos de formatos conocidos.

Utilizando la herramienta *exiftool* [6] es posible acceder a los metadatos disponibles en una amplia variedad de tipos de archivos. Si bien esta herramienta se especializa en EXIF (EXchangeable Image Format), un estándar para especificación de formatos de archivos de imagen, es viable su utilización para extraer metadatos de otros tipos de archivo que no son imágenes. Se trabaja para integrar un filtro adicional en el analizador básico a través de metadatos de formatos de archivos conocidos. Asimismo se prevé verificar la presencia de datos en posiciones específicas dentro de cada

archivo, comprobando que estos se encuentran dentro del rango de valores válidos y tengan el tipo adecuado para cada formato soportado.

Como corolario de las primeras experimentaciones ha de mencionarse que el análisis del *header* como única métrica de evaluación es insuficiente para lograr resultados satisfactorios en la identificación del formato de archivos. Se ha comprobado mediante el uso de una base de datos de firmas utilizada como esquema básico de validación que un mismo *header* puede repetirse en distintos tipos de archivos. Para subsanar los problemas de ambigüedad sobre algunos tipos de archivos mediante evaluaciones sobre del *header* se ha decidido hacer uso de MIME como valor estándar de comparación.

La investigación en curso se ha planteado en etapas que conllevan la implementación de métodos computacionales que actúan como clasificadores combinados para la identificación del formato de archivos. La estrategia propuesta consiste en integrar las siguientes líneas de análisis forense: a) Identificación mediante el *header*, b) Identificación mediante metadatos, c) Identificación mediante la tasa de compresión. Finalmente se prevé integrar dichos predictores mediante métodos de inferencia con lógica difusa. Se procura que el analizador reconozca el formato de un archivo con un porcentaje de verosimilitud, y los métodos implementados en las etapas subsiguientes de la presente investigación reafirmen ese reconocimiento.

El objetivo principal de este proyecto de investigación es el estudio de un método computacional flexible capaz de poder indicar con un grado de verosimilitud el formato de un archivo desconocido o sospechoso, eventualmente sin extensión y/o con alguna pequeña modificación en su estructura interna, a fin de poder acceder a su contenido mediante la aplicación nativa que lo generó.

Concretamente se desea implementar una herramienta que asista a un profesional forense en el reconocimiento del formato de un archivo, para aquellos casos en que el mismo

tenga pequeños cambios internos o en la extensión, como producto de un daño lógico eventual o bien provocado intencionalmente para intentar obstruir una investigación digital.

Para realizar pruebas de performance en el reconocimiento de formatos de archivos, preliminarmente se planteó utilizar un corpus digital generado a partir de archivos extraídos aleatoriamente desde Internet con consultas automáticas sobre diversos buscadores. Se pretende complementariamente realizar una experimentación sobre archivos extraídos desde discos rígidos en uso, en aras de simular escenarios más próximos a la realidad de trabajo de una pericia informática.

#### 4. FORMACION DE RECURSOS HUMANOS

La línea de investigación presentada sirve de apoyo a una tesis de grado de tres alumnos pertenecientes a la carrera Ingeniería en Informática de la Facultad de Ingeniería de la Universidad FASTA bajo la dirección funcional de un especialista en informática forense del Laboratorio Pericial Informático del Poder Judicial de la Provincia de Neuquén

#### 5. REFERENCIAS

- [1]. Amirani, M.C.; Toorani, M.; Beheshti, A., "A new approach to content-based file type detection", *Computers and Communications, 2008. ISCC 2008. IEEE Symposium on*, pp.1103,1108, 6-9 July 2008.
- [2]. Roussev, V.; Garfinkel, S.L., "File Fragment Classification - The Case for Specialized Approaches", *Systematic Approaches to Digital Forensic Engineering, 2009. SADFE '09. Fourth International IEEE Workshop on*, pp.3,14, 21-21 May 2009.
- [3]. Calhoun, W., Coles, D, "Predicting the types of file fragments", *Digital Investigation: The International Journal of Digital Forensics & Incident Response*, vol. 5, pp.14-20, September 2008.
- [4]. Amhed, I., Lhee, K., Shin, H. and Hong, M., "On Improving the Accuracy and Performance of Content-Based File Type Identification", *Information Security and*

*Privacy, Lecture Notes in Computer Science*, vol. 5594, pp 44-59, 2009.

[5]. TrID by Marco Pontello,  
<http://mark0.net/soft-trid-deflist.html>

(Último acceso: 09/03/2013)

[6]. ExifTOOL by Phil Harvey,  
<http://www.sno.phy.queensu.ca/~phil/exiftool/>

(Último acceso: 09/03/2013)



# Detección de Anomalías en Oráculos Criptográficos tipo RSA por medio de análisis probabilísticas y estadísticos.

Ing. Antonio CASTRO LECHTALER<sup>1</sup> Msc.  
Esp. Marcelo CIPRIANO<sup>2</sup>  
Eduardo MALVACIO<sup>3</sup>

Laboratorio de Investigación en Técnicas Criptográficas y Seguridad Teleinformática.  
Escuela Superior Técnica "Gral Div. Manuel N. Savio".  
Facultad de Ingeniería.  
Instituto de Educación Superior del Ejército Argentino.

[acastro@iese.edu.ar](mailto:acastro@iese.edu.ar)

[marcelocipriano@iese.edu.ar](mailto:marcelocipriano@iese.edu.ar)

[edumalvacio@gmail.com](mailto:edumalvacio@gmail.com)

## 1. Resumen.

Esta línea de investigación persigue la elaboración de herramientas matemáticas susceptibles de ser sistematizadas en un software capaz de detectar anomalías y mal funcionamiento en servicios de infraestructura de clave pública (PKI) que utilicen el esquema RSA.

Esta herramienta se puede aplicar en redes Públicas o Privados, Lan's, o Wan's e incluso Internet; en sistemas militares como del ámbito civil.

Dada la complejidad de los sistemas actuales, se hace cada vez más complicada la detección de determinado tipos de errores. Los mismos pasan de las versiones de pruebas a las versiones definitivas. Dichos errores provocan debilitamiento de los sistemas que la PKI intenta proteger. Incluso se podría sospechar la inoculación intencional de estos bugs, aunque no en todos los casos [01].

El esquema RSA utiliza números primos grandes para los actuales avances en la factorización de números: 1024, 2048 e inclusive 4096 bits de longitud o mayores

aún. Un sesgo en la selección de estos valores primos podría provocar una brecha en la seguridad.

Llamaremos Oráculo al un servicio que ofrezca a sus clientes claves públicas, privadas y módulos en un esquema de cifrado/descifrado RSA<sup>1</sup> en una Infraestructura de Clave Pública.

Esta herramienta informática que estamos elaborando permitiría la detección de mal funcionamiento en la distribución probabilística de los factores primos. Sin embargo dada la astronómica cantidad de números posibles, la detección es estadística, mediante el estudio de muestras suministradas por el Oráculo en observación.

## 2. Palabras Claves.

Seguridad en Redes, Infraestructura de Clave Pública, PKI, Detección de Anomalías, Open-SSL, RSA.

---

<sup>1</sup> La 3-tupla  $(n,e,d)$ :  $n$  (módulo) es el producto de 2 primos,  $e$  (clave pública) y  $d$  (clave privada) son inversos entre sí mód  $\phi(n)$ .

### 3. Contexto.

El Laboratorio de Investigación en Técnicas Criptográficas y Seguridad Teleinformática pertenece a la Escuela Superior Técnica –Facultad de Ingeniería– del Ejército Argentino en el área del Posgrado en Criptografía y Seguridad Informática que se dicta en esta institución, junto a otros posgrados y carreras de grado en ingeniería.

El desarrollo científico y tecnológico es relevante a nivel estratégico y es por ello que tanto las Fuerzas Armadas en general como el Ejército en particular.

El Instituto de Investigaciones Científicas y Técnicas para la Defensa (CITEDEF) realizó aportes a través de Proyectos de Investigación Científico Tecnológicos Orientados (PICTO) para la realización durante 6 años del proyecto recientemente finalizado sobre “REDES PRIVADAS COMUNITARIAS”, -dentro del cual se llevó adelante gran parte de esta investigación-, cuyos resultados parciales o finales han sido presentados en varios CACIC para su difusión a la comunidad científica.

### 4. Introducción.

#### 4.1. Generalidades

La Criptografía dejó de pertenecer a la esfera de los secretos militares y diplomáticos y se ha volcado al ámbito civil en general. No han dejado de crecer sus aplicaciones: ingreso a redes sociales, servidores de mails, home banking, compras online, redes WI-FI, almacenamiento de información confidencial o sensible, etc.

Todo lo que implique intercambio de información entre dos equipos informáticos, debería estar al resguardo de ojos indiscretos, sobre todo los intercambios inalámbricos.

Los sistemas se encargan de entablar todos los servicios requeridos por los estándares y protocolos de manera automática: tickets de ingreso a sistemas, intercambios de claves, autenticación de usuarios y equipos, inicio de sesión, y muchos otros son controlados en forma automática. Los usuarios por lo general no cambian las opciones por defecto: por ejemplo las personas que instalan sus routers con servicios inalámbricos sin autenticación de usuarios ni protocolos de encriptación.

¿Cómo comprobar si estos procesos y servicios contienen errores que pueden alterar la seguridad de lo que pretenden proteger?

#### 4.2. Código abierto no necesariamente aumenta la Seguridad de los Sistemas de Información.

El uso de Software Abierto (open source) en el ámbito de la Seguridad de la Información, es mucho más atractiva que la de software cerrado dado que se puede “ver” lo que el sistema hace y cómo lo hace. La ausencia de “puertas traseras” o claves no documentadas u otras técnicas como la inoculación de código virulento. Aunque estén abiertos sus códigos no significa que se encuentren libres de errores.

No todos los usuarios pueden leer el código e interpretar su significado y de esa manera corroborar el correcto funcionamiento del servicio. Luciano Bello ha descubierto un error en

OpenSSL<sup>2</sup>. Que fue enmendado 20 meses después que la versión defectuosa. [02].

Según la Ley de Linus “dado un número muy grande de ojos, los errores se convierten en evidentes” [03].

Está claro que al tiempo que aumenta la sofisticación y complejidad de los códigos, la cantidad de ojos disponibles para evidenciar los errores decrece. Es por ello que automatizar el análisis del funcionamiento del software en busca de anomalías se hace cada vez más necesario.

### 4.3. Planteamiento del problema.

Un oráculo tipo Open-SSL que ofreciera protección criptográfica (en sus diferentes formas antes mencionadas: claves de sesión, autenticación de equipos, etc.) es vulnerable cuando:

- 1) Oráculo genera un número relativamente pequeño de primos<sup>3</sup>.
- 2) Oráculo está diseñado con una directiva errónea para la selección de los primos o la generación de los módulos.

En ambos casos Oráculo se distancia del comportamiento equiprobable, dentro de ciertos parámetros asumidos, a un comportamiento sesgado. Susceptible de ser estudiado por un atacante y vulnerar así el sistema.

- 3) Un atacante tiene forma de construir *un subconjunto P' de Primos con los que trabaja* tal que el conocimiento de P' le permita vulnerar la factorización de RSA

<sup>2</sup> Una mala inicialización de una variable provocó una predictibilidad en el generador de números, abriendo una vulnerabilidad inimaginable.

<sup>3</sup> Vulnerabilidad de OpenSSL de Devian descubierta por L. Bello.

para una cantidad significativa de módulos.

Vemos que se deben hallar los factores primos del esquema RSA, proceso complejo pues esa es la fortaleza que permite a esta criptosistema proteger la información: la complejidad en la factorización de un número enorme.

## 5. Resultados y Objetivos.

### 5.1. Resultados intermedios de la línea de Investigación.

En el año 2008 nuestra investigación se orientó a la elaboración una herramienta informática que permite hallar los primos que componen un módulo RSA con la información que aportan su clave pública y su clave privada, resultado presentado y publicado en CACIC 2008 [04] realizado en La Rioja, Argentina.

Las posteriores pruebas de codificación e implementación demostraron que este procedimiento corría muy veloz presentándose estos resultados en CUBA [05].

Seguimos aún estudiando el comportamiento de este algoritmo y lo comparamos con el procedimiento que aborda el mismo problema y lo resuelve, existente en la bibliografía tradicional para la enseñanza de la criptografía [06].

Los análisis indicaron que la complejidad computacional de nuestro algoritmo era del orden  $O(\log n)$  mientras que el de la bibliografía de referencia tenía un orden  $O(\log^3 n)$ . Resultado presentado en Chile y que nos sorprendió gratamente [07].

Hallada la herramienta matemática que permitiría detectar anomalías (en caso que

las hubiere) el resultado publicado en CACIC 2011 [08] en La Plata.

Con el abordaje probabilístico del problema presentado en 41 JAIIO [09] y por último el diseño final de la herramienta probabilística y estadística presentado en CACIC 2012 [10] en Bahía Blanca se terminó de dar forma definitiva al sustento teórico.

Simultáneamente al avance matemático se ensamblaron y codificaron todas las herramientas matemáticas antes mencionadas, en una plataforma de software programado en C++.

## 5.2. Etapa actual

Se llevaron adelante pruebas en las que, por lotes, se le solicita a OPEN-SSL la entrega de módulos RSA a efectos de buscar colisiones de primos. Es decir, se analizan los factores primos de cada uno de ellos y se los compara con los de los otros módulos del lote en busca de factores repetidos.

En primer lugar estamos buscando las causas por las cuales el proceso informático agota la memoria de la computadora y el sistema operativo interrumpe la ejecución del programa. Aún no encontramos la causa y es por ello que la exploración sólo se ha realizado en lotes de 1000 módulos cada uno y para 64 bits de tamaño de los mismos.

Pese a esta dificultad se han podido evaluar 9.000.000 de módulos agrupados en 9000 lotes de 1000 cada uno. Los resultados obtenidos no responden a las predicciones teóricas.

Actualmente estamos revisando todo en búsqueda de respuestas a las

inconsistencias halladas. Habría 2 posibilidades a tener en cuenta:

a) la detección de errores en la implementación de la herramienta matemática y su depuración.

b) la detección de una Anomalía en la distribución de factores primos del Oráculo, justo lo que la herramienta busca hallar.

## 5.3. Etapa final.

Una vez que aseguremos la ausencia de errores, en caso que los hubiera, manipularemos a Oráculo para que presente vulnerabilidades. Correremos el programa y de manera aleatoria repetiremos el experimento con oráculos vulnerables como así también sin manipular.

Se espera que el programa sea capaz de detectar los oráculos vulnerables y alertar acerca de ellos. Asimismo informar también sobre los oráculos seguros.

El experimento se repetirá gran cantidad de veces, registrando los falsos positivos como los negativos si los hubiere. De esa forma estudiar la eficacia y eficiencia de esta herramienta.

Cabe aclarar que dada la magnitud en tamaño y la enorme cantidad de valores con las que trabajan estos oráculos, el abordaje es probabilístico-estadístico.

Publicaremos estos resultados en los próximos meses. Es nuestra intención poder hacerlo en el CACIC 2013.



## 6. Formación de Recursos Humanos.

Algunos algoritmos fueron codificados y probados en el contexto de la Cátedra de Computación I a cargo del Ing. Mg. Alejandro Repetto, que posee nuestra facultad en la carrera de Ingeniería Informática,

Si bien aún el laboratorio no posee ningún becario doctorando o post-doctorando, el resto de los integrantes del equipo son investigadores categorizados en el programa de Incentivos del Ministerio de Ciencia, Tecnología e Innovación Productiva y en el Sistema Científico Tecnológico de la Defensa en el régimen de Régimen para el personal de Investigación y Desarrollo de las Fuerzas Armadas (RPIDFA).

## 7. Referencias y Bibliografía

[01] Young A and Yung M. An Elliptic Curve Asymmetric Backdoor in OpenSSL RSA Key Generation. Chapter 10. Cryptovirology. 2006.

<http://www.cryptovirology.com>.

[02] Bello L, Bertacchini M. “Generador de Números Pseudo-Aleatorios Predecible en Debian”. III Encuentro Internacional de Seguridad Informática. Manizales, Colombia. Octubre 2009.

[03] Glass, Robert “Facts and Fallacies of Software Engineering”. Addison-Wesley Professional, 2003.

[04] Cipriano, M. “Factorización de  $N$ : recuperación de factores primos a partir de las claves pública y privada.” Anales del XIV Congreso Argentino de Ciencias de la Computación CACIC 2008. Chilecito, La Rioja, Octubre 2008.

[05] Castro Lechtaler, C; Cipriano, M; Benaben A; Quiroga, P. “Study on the effectiveness and efficiency of an algorithm to factorize  $N$  given  $e$  and  $d$ ” Anales del IX Seminario Iberoamericano en Seguridad de las Tecnologías de la Información, La Habana, CUBA. 2009.

[06] Menezes, A; van Oorschot, P and Vanstone, S. *Handbook of Applied Cryptography*. CRC Press. 5th Edition, 2001.

[07] Benaben, A; Castro Lechtaler, A; Cipriano, M; Foti, A. “Development, testing and performance evaluation of factoring algorithms whit additional information” XXVIII Conferencia Internacional de la Sociedad Chilena de Computación. Santiago de Chile

[08] Castro Lechtaler, A; Cipriano, M. “Detección de anomalías en Oráculos tipo OpenSSL por medio del análisis de probabilidades” Anales del XVII Congreso Argentino de Ciencias de la Computación CACIC 2011. La Plata, Buenos Aires, Octubre 2011.

[09] Castro Lechtaler, Antonio, Cipriano Marcelo; Malvacio Eduardo; Cañón, Sebastián; *Procedure for the Detection of Anomalies in Public Key Infrastructure (RSA Systems)*. Anales del XIII Simposio Argentino de Tecnología, 41 Jornadas Argentinas de Informática e Investigación Operativa JAIIO – SADIO. La Plata, Buenos Aires, Agosto 2012.

[10] Castro Lechtaler, Antonio; Cipriano, Marcelo; Malvacio, Eduardo. *Experimental detection of anomalies in public key infrastructure* . Anales del XVIII Congreso Argentino de Ciencias de la Computación CACIC 2012. Bahía Blanca, Buenos Aires, Octubre 2011.

## Inicio de la Línea de Investigación “Ingeniería de Software y Defensa Cibernética”

Roberto Uzal<sup>2</sup>, Jeroen van de Graaf<sup>1</sup>, Germán Montejano<sup>2,3</sup>, Daniel Riesco<sup>2</sup>, Pablo García<sup>3</sup>

<sup>1</sup> Departamento de Ciencia da Computacao, Universidade Federal de Minas Gerais - Brasil  
jvdg@dcc.ufmg.br – <http://www.dcc.ufmg.br/~jvdg>

<sup>2</sup> Facultad de Ciencias Físico-Matemáticas y Naturales, Universidad Nacional de San Luis – Argentina –  
ruzl@uolsinectis.com.ar, gmonte@unsl.edu.ar, driesco@unsl.edu.ar - <http://sel.unsl.edu.ar>

<sup>3</sup> Facultad de Ciencias Exactas y Naturales, Facultad de Ingeniería, Universidad Nacional de La Pampa – Argentina -  
pablogarcia@exactas.unlpam.edu.ar – <http://www.unlpam.edu.ar>

### Resumen

Se presenta una nueva línea de investigación que estará inserta en el proyecto “*Ingeniería de Software: Aspectos de alta sensibilidad en el ejercicio de la profesión de Ingeniero de Software*” de la UNSL [1]. Esta línea estará incluida en trabajos realizados en cooperación con universidades de Brasil [2] y comprende a la prevención de la intrusión y/o mitigación de los efectos de software malicioso sofisticado destinado a causar daños a la infraestructura crítica del país, a paralizar el funcionamiento de los servicios esenciales o para acceder a información gubernamental de carácter secreto [3] ...[14]. Están incluidos en esta línea: la prevención y detección de software malicioso sofisticado, la ingeniería reversa del malware detectado [15] [16] [17] [18], la identificación del emisor mediante “análisis de flujo de redes” [19] [20] y la neutralización de ataques provenientes desde otros estados naciones. Esta presentación incluye ejemplos reales de agresiones devastadoras entre países utilizando “malware” extremadamente complejo y la enumeración de los conceptos y habilidades a ser desarrollados en el contexto descripto. Se cita el consenso internacional que la UNSL ha logrado respecto de la necesidad de un “Nuevo Paradigma de Seguridad Informática” [21] y se describen los temas de investigación ya encarados. Finalmente se enumeran las referencias.

**Palabras clave:** Defensa Cibernética, Seguridad Informática, Análisis de Flujo de Redes, Nuevo Paradigma en Seguridad Informática

### Contexto

El proyecto “*Ingeniería de Software: Aspectos de alta sensibilidad en el ejercicio de la profesión de Ingeniero de Software*” de la UNSL, con ligeras variantes en su denominación, viene desarrollándose, muy exitosamente, desde hace más de quince años. Tres de los integrantes de proyecto han alcanzado la Categoría I en el ámbito del Programa de Incentivo a la Investigación en Universidades Nacionales y los otros miembros del equipo han alcanzado diversas Categorizaciones en el mismo Programa. El proyecto está íntimamente asociado al desarrollo de la carrera de grado de Ingeniería Informática y de las carreras de post grado Doctorado en Ingeniería Informática, Maestría en Ingeniería de Software, Maestría en Calidad del Software y Especialización en Ingeniería de Software.

Este proyecto mantiene efectivos esquemas asociativos [1] con el Instituto de Tecnología del Software de las Naciones Unidas (Macao), con la Universidad Estatal de San Francisco (EEUU), con la Universidad Federal de Minas Gerais (Brasil), con la Universidad de Minho (Portugal) y con las Universidades de Castilla La Mancha, Vigo y Politécnica de Valencia (España). Asimismo el proyecto “*Ingeniería de Software: Aspectos de alta sensibilidad en el*

*ejercicio de la profesión de Ingeniero de Software*” trabaja en permanente contacto con el Parque Industrial del Software de San Luis, el de Córdoba y con centros de excelencia de elaboración de Software de Europa.

El financiamiento del proyecto proviene regularmente de la UNSL, de programas nacionales e internacionales específicos, de emprendimientos como el CAPG – BA [2] y de cada uno de las actividades que han venido realizándose como resultado de la interacción con el entorno social y productivo.

## 1. Introducción

En el contexto del proyecto *“Ingeniería de Software: Aspectos de alta sensibilidad en el ejercicio de la profesión de Ingeniero de Software”* de la UNSL se habían venido estudiando detenidamente los aspectos tecnológicos correspondientes a incidentes entre estados naciones del tipo:

- La masiva y devastadora agresión cibernética de Rusia a aeropuertos, sistemas ferroviarios, hospitales, sistema financiero y medios periodísticos de Estonia en el 2007. El ataque provocó la reacción de Alemania en ayuda de Estonia y luego la intervención de la NATO. A partir de este conflicto la NATO (Organización del Atlántico Norte) consolida su política y estructura de Defensa Cibernética [22].
- La alteración, mediante Armas Cibernéticas, del software de un Sistema de Radar de origen Ruso en el Norte de Siria, a orillas del Éufrates, en el 2007, impidiéndole detectar a cazas bombarderos de Israel que atacaron y destruyeron construcciones realizadas, aparentemente, por norcoreanos [3].
- La intrusión de China en sistemas satelitales de Estados Unidos [23].
- El prácticamente confirmado acceso, también por parte de China, a información del área Defensa altamente sensitiva, residente en la Intranet del Jet Propulsion Laboratory

(California Institute of Technology – NASA) [24].

- La voladura, utilizando “virus de red” o “gusanos”, de las baterías de centrífugas en la planta de enriquecimiento de uranio en Natanz, Irán [25] [26].
- La presencia de la ultra sofisticada Arma Cibernética, Flame, en las plataformas de explotación petrolera de Irán [27]
- La Guerra Cibernética de carácter “sine die” entre Pakistán e India [28] [29].

Asimismo, en el contexto del proyecto *“Ingeniería de Software: Aspectos de alta sensibilidad en el ejercicio de la profesión de Ingeniero de Software”* de la UNSL, en relación con lo expuesto más arriba, se habían venido encarando los aspectos antes de esta formalización de la nueva línea de investigación:

### 1.1. Acciones de Comunicación / Difusión realizadas por la UNSL:

- 1.1.1. Conferencia en el Centro de Oficiales de las FFAA: “Nuevos escenarios, nuevos conflictos, nuevas armas, nuevos soldados: Cyber War”; 24 de abril de 2012
- 1.1.2. Artículo de opinión sobre Cyber War en la Revista DEF de Agosto de 2012
- 1.1.3. Reportaje al Dr. Roberto Uzal sobre Cyber War - Canal de TV C5N el 19 de agosto a las 23 horas.
- 1.1.4. Reportaje al Dr. R. Uzal sobre Guerra Cibernética realizado por la Revista DEF ON LINE – Agosto de 2012 <http://www.defonline.com.ar/?p=9064>
- 1.1.5. Consejo Profesional de Ingenieros Electrónicos y en Telecomunicaciones – COPITEC – “Guerra Cibernética” (conferencia), 4 de julio de 2012
- 1.1.6. Consejo Argentino de Relaciones Internacionales, seminario sobre "Crimen Organizado Transnacional y

- Ciberamenazas", 11 de noviembre de 2012
- 1.1.7. Sociedad Científica Argentina. Seminario sobre Guerra Cibernética dirigido a los Oficiales del Ejército de la Especialidad Informática; 21 de noviembre de 2012.
  - 1.1.8. Exposición en la Universidad Nacional de San Luis en el contexto de las Jornadas por un Ciberespacio más seguro y confiable: "Ciberamenazas, escenarios y propuestas"; 18 de diciembre de 2012 <http://webfmn.unsl.edu.ar/boletines/boletin418/boletin418.html>
  - 1.1.9. Artículo "Guerra Cibernética: ¿Un desafío para la Defensa Nacional", Revista Visión Conjunta (Escuela Superior de Guerra Conjunta – Ministerio de Defensa), Año 4, Nro 7, 2012
- 1.2. Tesis de Maestría de Walter Agüero: El alumno – graduado de la Maestría en Ingeniería de Software de la Universidad Nacional de San Luis Walter Agüero ha manifestado su interés en encarar su trabajo de tesis según el siguiente esquema general:
    - 1.2.1. Desarrollo de un módulo adicional, para un determinado sistema operativo de computadora personal, tal que le posibilite a dicha PC, bajo determinadas condiciones, sustraerle / copiarle la lista de contactos a teléfonos celulares que se encuentren dentro de determinado distancia (de la PC).
    - 1.2.2. Desarrollo de un módulo adicional, para determinados sistemas operativos de teléfonos celulares, tal que impida la sustracción / copia mencionada en el párrafo anterior.
    - 1.2.3. Desarrollo de otro módulo adicional, para determinados sistemas operativos de teléfonos celulares,

que, cuando se produzca un intento de sustracción / copia de la lista de contactos, le advierta tal anomalía al usuario indicando la URL de la computadora que realizó tal intento.
  - 1.3. Tesis de Maestría de Pablo García: El alumno graduado de la Maestría en Ingeniería de Software de la Universidad Nacional de San Luis Pablo García desarrolló un enfoque algorítmico que incrementó significativamente las condiciones de confidencialidad del sistema de voto electrónico (extranjero). Los avances de Pablo García pueden ser utilizados en el contexto de la Defensa Cibernética.
  - 1.4. Tesis Doctoral de Mario Berón: El Dr. Mario Berón ha realizado un muy valioso aporte a la Ingeniería Reversa a nivel programa y a la "comprensión" del funcionamiento y prestaciones de software del que se carece de su documentación de diseño
  - 1.5. Convocatoria de ENIGMA: Integrantes del proyecto "*Ingeniería de Software: Aspectos de alta sensibilidad en el ejercicio de la profesión de Ingeniero de Software*" han sido invitados a desempeñarse como evaluadores de los trabajos que se presenten a "ENIGMA" The Brazilian Journal of Cryptography and Information Security, publicación científica auspiciada por el Ministerio de Defensa de Brasil
  - 1.6. Un incidente en el Parque Industrial: Se contactó, a fines del año pasado, con el Gerente de una Planta de Tratamiento de Efluentes de una importante industria. De estudiado surgió que dicha Planta de Tratamientos de Efluentes había sido atacada (los PLC de la misma) mediante



un malware del mismo tipo del utilizado en el ataque a las instalaciones iraníes de Natanz. Esto no es extraño pues está casi verificada la existencia de un “mercado negro” de Ciberarmas. <http://www.businessweek.com/magazine/cyber-weapons-the-new-arms-race-07212011.html>

- 1.7. Posibilitar al LaCIS el Análisis de Flujos de Redes: El Laboratorio de Calidad e Ingeniería de Software del Área Metodologías y Programación del Departamento Informática de la UNSL tiene por objetivo promover la transferencia tecnológica y la asistencia técnica al entorno productivo. Un aspecto esencial que deberá encarar también el LaCIS es el de capacitar a Equipos de Respuesta a Incidentes en el Ciberespacio en lo que hace a la capacidad de detectar Botnet(s) y “Command and Control Servers” mediante “Análisis de Flujo de Redes”. Se han adoptado previsiones en el sentido señalado.
- 1.8. Asistencia al Ministerio de Defensa – Estado Mayor Conjunto de las Fuerzas Armadas de Argentina en el Planeamiento Estratégico de la Defensa Cibernética de Argentina: Integrantes de esta línea de investigación están materializando esta asistencia en el momento de la elaboración de este documento. Marzo de 2013.
- 1.9. Una síntesis de las ideas desarrolladas en las actividades detalladas más arriba se plasmó en un "paper" titulado "Trust in Cyberspace: New Information Security Paradigm" el cual fue presentado en la Conferencia IEEE que se realizará los días 22 y 23 de marzo de 2013 en Hangzhou, China. El trabajo ya fue aceptado por el Comité de Programa (arbitraje internacional mediante)

## 2. Ejes de la nueva línea de investigación

Ingeniería Reversa a nivel programa, Criptografía y criptoanálisis, Detección de malware inmune a los productos “anti malware” de disponibilidad comercial, Análisis de Flujo de Redes, Enfoques innovadores en Auditoría de Seguridad Informática

## 3. Objetivos y resultados esperados

- 3.1. Ingeniería Reversa a nivel programa: Lograr la Ingeniería Reversa, partiendo de lenguaje de máquina, de productos de software elaborados con un enfoque multiparadigma.
- 3.2. Criptografía y criptoanálisis: Desarrollar herramientas de criptografía y criptoanálisis de alto rendimiento con “know how” propio.
- 3.3. Detección de malware inmune a los productos “anti malware” de disponibilidad comercial: Desarrollar los conceptos y herramientas correspondientes.
- 3.4. Análisis de Flujo de Redes: Desarrollar herramientas de alta efectividad.
- 3.5. Enfoques innovadores en Auditoría de Seguridad Informática: Proponer el perfeccionamiento de los estándares actualmente vigentes (del tipo ISO 27001 / 27002)

## 4. Formación de Recursos Humanos

Además de la Tesis doctoral de Mario Berón, la Tesis de Maestría de Pablo García y la Tesis de Maestría de Walter Agüero, está estudiándose la viabilidad de elaboración de nuevas tesis doctorales y nuevas tesis de maestría, abordando temas relacionados con la Defensa Cibernética, a ser encaradas en el futuro próximo. También se está poniendo un particular énfasis en el desarrollo de temas de Seguridad Informática en el contexto del Doctorado en Ingeniería Informática, de las Maestrías en Ingeniería de Software y Calidad del Software y en la Especialización en Ingeniería de Software. Asimismo se está

negociando la posibilidad de que la UNSL se haga cargo de la capacitación en Defensa Cibernética de personal del Ministerio de Defensa de Argentina. El equipo inicial de la línea de investigación que se inicia está constituido por el Dr. Roberto Uzal, el Dr. Jeroen van de Graaf<sup>1</sup>, el Dr. Germán Montejano, el Dr. Daniel Riesco, el Dr. Mario Berón y el Lic. Pablo García (tesis de maestría ya elaborada).

## 5. Referencias

- [1]. <http://www.sel.unsl.edu.ar/>
- [2]. <http://webfmn.unsl.edu.ar/boletines/boletin416/noticia5-brasil.htm>
- [3]. Richard Clarke, "Cyber War", Harper Collins, 2010
- [4]. <http://www.eluniverso.com/2009/11/12/1/1361/a-pagon-brasil-genera-temores.html>
- [5]. Wall Street Journal, June 15, 2011
- [6]. <http://www.youtube.com/watch?v=XgnRvntfdo>
- [7]. [http://www.nytimes.com/2012/06/06/books/confront-and-conceal-by-david-sanger.html?\\_r=0&adxnml=1&adxnmlx=1362427401-0atlzUnTtJgjKOpj6e1Qvw](http://www.nytimes.com/2012/06/06/books/confront-and-conceal-by-david-sanger.html?_r=0&adxnml=1&adxnmlx=1362427401-0atlzUnTtJgjKOpj6e1Qvw)
- [8]. [http://www.youtube.com/watch?v=EvP\\_x09cfU0](http://www.youtube.com/watch?v=EvP_x09cfU0)
- [9]. <http://www.in.gov.br/visualiza/index.jsp?data=27/12/2012&jornal=1&pagina=11&totalArquivos=304>
- [10]. <http://www.defesanet.com.br/cyberwar/noticia/5954/CDCiber---Centro-de-Defesa-Cibernetica-inicia-em-Junho->
- [11]. <http://revistaepoca.globo.com/Revista/Epoca/0,,EMI249428-15223,00-GENERAL+JOSE+CARLOS+DOS+SANTOS+PODEMOS+RECRUTAR+HACKERS.html>
- [12]. <http://www.eluniverso.com/2009/11/12/1/1361/a-pagon-brasil-genera-temores.html>
- [13]. <http://cespe.espe.edu.ec/2012/10/page/2/>
- [14]. <http://sseguranca.blogspot.com.br/2012/12/politica-cibernetica-de-defesa.html>
- [15]. Berón M. y Henriques P. y Varanda Pereira M. y Uzal R.; "Comprensión de programas"; XI Workshop de Investigadores en Ciencias de la Computación., 1, 2009, ISBN: 978-950-605-570-7
- [16]. R. Berón M. y Henriques P. y Varanda M. y Uzal R." Inspección de código para relacionar los dominios del problema y programa para la comprensión de programas". X Workshop de Investigadores en Ciencias de la Computación., 1:549-553, 2008
- [17]. Beron M. y Cruz D. y Pereira M. y Henriques P. y Uzal R. "Evaluation criteria of software visualization system used for program comprehension", 3a Conferencia Nacional em Interacção Pessoa-Máquina, 3:81-86, 2008. Estado: Publicado. Editorial: Universidade de Évora. ISBN: 972-9464-9-9.
- [18]. Berón M. y Henriques P. y Varanda Pereira M. y Uzal R.; "Simplificando la comprensión de programas a través de la interconexión de dominios". XIV Congreso Argentino de Ciencias de la Computación., 1, 2008. Estado: Publicado. Editorial: Universidad Nacional de la Rioja. ISBN: 978-987-24611-0-2.
- [19]. [http://www.sba-research.org/wp-content/uploads/publications/acsac12\\_disclosure.pdf](http://www.sba-research.org/wp-content/uploads/publications/acsac12_disclosure.pdf)
- [20]. <http://www.cert.org> V. Krmíček, T. Plesník Detecting Botnets with NetFlow FloCon 2011, January 12, Salt Lake City, Utah
- [21]. [http://www.atlantispress.com/publications/aisr/iccsee-13/index\\_iccsee-13.html?http%3A//www.atlantispress.com/php/paper-details.php%3Fid%3D4471](http://www.atlantispress.com/publications/aisr/iccsee-13/index_iccsee-13.html?http%3A//www.atlantispress.com/php/paper-details.php%3Fid%3D4471)
- [22]. <http://news.bbc.co.uk/2/hi/europe/6665145.stm>
- [23]. <http://www.defensenews.com/article/20111028/DEFSECT01/110280301/Report-Cyber-Attacks-Targeted-U-S-Satellites>
- [24]. <http://www.foxnews.com/scitech/2012/03/01/chinese-hackers-nasa-jpl-lab/>
- [25]. <http://www.dailymail.co.uk/news/article-2178781/Iran-nuclear-facilities-hit-cyber-attack-plays-AC-DCs-Thunderstruck-volume.html>
- [26]. <http://www.reuters.com/article/2013/02/26/us-cyberwar-stuxnet-idUSBRE91P0PP20130226>
- [27]. <http://www.youtube.com/watch?v=vrRj-kRofRg> (documental de la TV Iraní)
- [28]. <http://cyberleaks.org/the-cyber-war-still-continues-between-india-and-pakistan-since-1954-a-report/>
- [29]. <http://www.indianexpress.com/news/cyber-war-blaming-pakistan-is-not-enough/990637>

# Identificación temprana de características transversales en el lenguaje de la aplicación capturado con el Léxico Extendido del Lenguaje

Leandro Antonelli

Director: Gustavo Rossi

Co-director: Julio Cesar Sampaio do Prado Leite

Asesor técnico: Alejandro Oliveros

**Tesis presentada para la obtención el grado de Doctor en Ciencias Informáticas,  
Facultad de Informática, Universidad Nacional de La Plata**

**Fecha de exposición: 27 de Abril 2012**

## Abstract

En este trabajo de tesis se presenta una estrategia para identificar características transversales del software utilizando el Léxico Extendido del Lenguaje (LEL). Debido a que la construcción del LEL se realiza tempranamente en el proceso de desarrollo del software, esto redundará en beneficios al evitar el retrabajo que pudiera ocasionar la detección a mitad del desarrollo. El lenguaje que utiliza para representar el conocimiento de la aplicación (LEL) posee buena expresividad, pero por sobre todo, utiliza el lenguaje conversacional sin utilizar ningún tipo de formalismo, con lo cual puede ser utilizado por todas las personas que participan del desarrollo de software. En esta tesis se muestra tanto la aplicabilidad como la efectividad de la estrategia propuesta. La aplicabilidad se ilustra a través de 3 ejemplos del mundo real. Además, la presente tesis también describe un experimento que se llevó a cabo con el fin de mostrar la efectividad de la estrategia. Por último, la tesis describe una herramienta que permite asistir en la aplicación de la estrategia propuesta.

## 1 Introducción

El software tiene que enfrentarse a muchas características (*concerns*). Con el fin de hacer frente a esta complejidad, la ingeniería de software tradicional tiene como estrategia la descomposición, esto es la estrategia de *divide y vencerás*: dividir al problema en piezas más pequeñas las cuales pueden ser resueltas cada una por separado. Luego de obtenida la solución de cada una de las piezas, todas las soluciones parciales se deben integrar para conformar la solución al problema original. Sin embargo, esta estrategia falla cuando es necesario trabajar con piezas que están dispersas en otras piezas (porque es una pieza común y se vincula con todas ellas). Esto generalmente ocurre cuando se inyectan atributos de calidad, ya que la implementación de estos atributos es común a las diferentes piezas. Un ejemplo recurrente es el de autorización: diferentes piezas de software pueden requerir operaciones que garanticen que la autorización adecuada está presente en la ejecución de cada pieza. Como tal, autorización es un ejemplo de característica transversal, esto es, una característica que atraviesa diferentes piezas que conforman al todo.

Las características transversales se caracterizan por ser enmarañadas (*tangled*) y dispersas (*scattered*). Cuando una misma pieza posee diferentes características (*concerns*) se denomina enmarañada (*tangled*), y por otro lado, cuando una misma característica está distribuida por diferentes piezas, se dice que está dispersa (*scattered*). La comunidad de Desarrollo de Software Orientado a Aspectos (*Aspect-Oriented Software Development, AOSD*) se ocupa de las características transversales con el fin de atacar los problemas de enmarañamiento (*tangling*) y dispersión (*scattering*).

El desarrollo de software es una sucesión de descripciones en diferentes lenguajes donde a partir de una descripción se produce la siguiente. Por lo cual, si se incorporan cambios en una descripción deben ajustarse tanto las descripciones previas como la siguiente para que se mantenga la consistencia entre todas ellas. Bohem [Boehm 1997] indica que si un error se desliza en la descripción de requerimientos y recién en la etapa de codificación es detectado, la corrección del mismo costaría entre 100 y 200 veces más que si se lo hubiera corregido en la etapa de requerimientos. Por este motivo, es necesario tomar las decisiones lo antes posible en el desarrollo del software.

Yu et al. [Yu 2004] se enfocan en descubrir aspectos desde los objetivos. Los modelos basados en objetivos proveen la descripción y el análisis de las intenciones que subyacen a un nuevo sistema de software. Algunos modelos orientados a objetivos como i\* también modelan los actores que poseen tales intenciones. Estos objetivos son “groseramente hablando, precursores de los requerimientos”. Nuseibeh [Nuseibeh 2004] argumenta que el mundo del problema, el cual está habitado por clientes y usuarios, es un terreno fértil para identificar características (*concerns*). En verdad, él indica que el mundo real es a menudo la fuente más apropiada para identificar características transversales. Rashid et al. [Rashid 2006] argumentan que, es posible identificar características transversales mientras se aplica la estrategia de análisis del dominio en la etapa de ingeniería de requerimientos cuando se desarrolla software orientado al reuso.

El Léxico Extendido del Lenguaje [Leite 1993] es un glosario que tiene por finalidad describir el lenguaje del contexto de la aplicación. LEL es anclado en una idea bien simple “entender el lenguaje del problema sin preocuparse por entender el problema”. De esta forma, luego de comprender el lenguaje, el analista podrá escribir requerimientos utilizando como base el conocimiento adquirido a través del lenguaje capturado.

En este trabajo de tesis se presenta una estrategia para identificar las características transversales utilizando el Léxico Extendido del Lenguaje (LEL). La estrategia identifica características transversales en una forma similar a la que son identificados de los requerimientos. Sin embargo, mientras las técnicas tradicionales se basan en las acciones, la estrategia propuesta se basa en estados (aunque también utiliza sustantivos y acciones).

Este artículo está organizado de la siguiente manera. La sección 2 describe los trabajos relacionados. La sección 3 presenta la estrategia propuesta. La sección 4 discute aspectos de aplicabilidad, efectividad y usabilidad. La sección 5 presenta conclusiones. Finalmente, la sección 6 muestra nuevas líneas de trabajo.

## 2 Trabajos relacionados

Este capítulo describe distintos trabajos que principalmente se ocupan de identificar características transversales, ya que este es el objetivo de esta tesis: identificar características transversales en una etapa temprana del desarrollo de software. Sin embargo, en este capítulo también se describen técnicas de identificación de características transversales en el código fuente, puesto que por analogía se podría extrapolar las técnicas a una etapa temprana.

Baniassad [Baniassad 2006] presenta una forma de organizar los requerimientos descriptos en texto coloquial. El objetivo es factorizar los mismos sin repetir información, es por ello que los autores presentan una serie de actividades con este objetivo. La primera actividad es la identificación de las características repetidas y esta actividad puede ser realizada de varias formas: identificando términos aspectuales, identificando requerimientos que impactan o identificando características dispersas.



Baniassad [Baniassad 2004] describe una estrategia concreta para identificar aspectos. La estrategia se basa principalmente en analizar requerimientos identificando acciones y entidades. Las acciones son analizadas y luego de realizada una selección sobre las mismas son definidos *themes*. Luego, se organiza la información de los requerimientos guiados por los *themes*.

Sampaio [Sampaio 2005] propone una herramienta configurable para identificar características transversales a partir del texto. La herramienta es configurable porque permite identificar características transversales a partir de puntos de vistas, casos de uso, etc... Sin embargo, la identificación no comienza con estos productos, sino que la herramienta realiza un análisis del lenguaje coloquial, el que obtiene de documentos de requerimientos y partir de ellos comienza el análisis.

Rago [Rago 2009] considera que es necesario tomar en cuenta no sólo verbos para la identificación, así que su estrategia considera además los objetos sobre los cuales trabajan los verbos.

Rashid [Rashid 2003] plantea un modelo con puntos de vista que agrupan requerimientos. El agrupamiento lo realiza a través de puntos de vistas y luego identifican puntos de vista transversales que impactan en otros requerimientos.

Bounour [Bounour 2006] describe en su trabajo varias estrategias de minería de datos basadas en el código fuente. Modelan flujo de control o flujo de datos, estructura del código fuente o dispersión del mismo, y también analizan conceptos (entidades) en función de sus referencias.

### 3 Identificación de características transversales a través del Léxico Extendido del Lenguaje

La estrategia de identificación de características transversales se basa en un modelo que captura el lenguaje de la aplicación (LEL) y a partir de él realiza el análisis que determina las características transversales candidatas.

#### 3.1 Léxico Extendido del Lenguaje (LEL)

El Léxico Extendido del Lenguaje [Leite 1993] es un glosario que tiene por finalidad describir el lenguaje del contexto de la aplicación. Su objetivo es el describir ciertas palabras o frases peculiares a la aplicación y que su comprensión son vitales para poder comprender el contexto de la misma. LEL es anclado en una idea bien simple “entender el lenguaje del problema sin preocuparse por entender el problema”. De esta forma, luego de comprender el lenguaje, el analista podrá escribir requerimientos utilizando como base el conocimiento adquirido a través del lenguaje capturado.

El LEL es un glosario en el cual se definen símbolos (términos o frases). A diferencia del diccionario tradicional que posee sólo un tipo de definición por término (puede haber muchas acepciones, sin embargo, todas son significados del término que están describiendo), en el LEL cada expresión se define a través de dos atributos: la noción (*notion*) y los impactos (*behavioural responses*). La noción describe la denotación, es decir, describe las características intrínsecas y substanciales del símbolo. Por su parte, los impactos describen la connotación, es decir, un valor secundario que adopta por asociación con un significado estricto.

Se determinan cuatro categorías básicas de símbolos: sujeto, objeto, verbo y estado. A partir de las 4 categorías y a la descripción general de los símbolos, se desprende la descripción de cada categoría que se resume la

Tabla 3-1.

Tabla 3-1. Categorías de símbolos y su descripción

Categoría	Características	Noción	Impactos
Sujeto	Elementos activos que realizan acciones	Características o condiciones que los sujetos satisfacen	Acciones que el sujeto realiza
Objeto	Elementos pasivos con los cuales los sujetos realizan acciones	Características o atributos que los objetos poseen	Acciones que son realizadas con los objetos
Verbo	Acciones que los sujetos realizan con los objetos	Objetivo que el verbo persigue	Pasos necesarios para completar la acción
Estado	Situaciones en las cuales se pueden encontrar los sujetos y los objetos	Situación que representa el estado	Acciones que deben realizarse para cambiar a otro estado

El LEL es una herramienta adecuada y efectiva para modelizar el lenguaje de una aplicación por su adecuación a los mecanismos que posee el cerebro para organizar el conocimiento experto [Wood 1997]. La utilidad del LEL como modelo ha sido observada en los trabajos de Breitman [Breitman 2003] y Gruber [Gruber 1993]. En principio LEL es una herramienta muy conveniente para expertos sin habilidades técnicas, sin embargo, las personas con tales habilidades obtendrán un mayor beneficio con su uso. La conveniencia del LEL proviene de 3 características significativas: es fácil de aprender, es fácil de usar y posee buena expresividad. Hay experiencias en dominios complejos que validan la conveniencia del LEL. Gil et al. [Gil 2000] indican que: “la experiencia de construir un LEL de una aplicación completamente desconocida para los ingenieros de requerimientos y con un lenguaje altamente complejo, puede ser considerada exitosa, desde el momento en que los usuarios fueron los que notaron que los ingenieros de requerimientos habían desarrollado un gran conocimiento sobre la aplicación”. Por su parte, Cysneiros et al. [Cysneiros 2001] indican que: “el uso del LEL fue muy bien aceptado y comprendido por los interesados (*stakeholders*). Dado que los interesados (*stakeholders*) no eran expertos en los dominios tan complejos y específicos en los que trabajaron, los autores creen que el LEL puede ser adecuado para utilizarse en muchos dominios”.

### 3.2 Identificación de características transversales con LEL

La esencia del enfoque propuesto está basado en identificar como características transversales los símbolos que con mayor frecuencia son referenciados desde los impactos de otros símbolos [Antonelli 2010]. El enfoque es mucho más complejo, a continuación se detallan los pasos que se resumen en la Figura 3-1:

- (i) Construcción del LEL organizado en grupos. Se debe definir la noción y los impactos de cada símbolo y también se debe relacionar cada símbolo con un estado.
- (ii) Conteo de referencias. Se deben considerar (contar) las referencias desde los impactos de los símbolos que forman parte de un estado distinto al estado del símbolo donde llegan las referencias. Luego, se debe calcular las referencias para el grupo en su totalidad ya que el enfoque se basa en identificar grupos como candidatos a ser características transversales.

- (iii) Ranqueo de grupos. Se deben ordenar los grupos de símbolos (estados) de acuerdo a la probabilidad de que sean considerados características transversales.
- (iv) Análisis final. En esta etapa se deben identificar los grupos con mayor probabilidad de que sean consideradas características transversales. Además, se debe realizar un análisis a nivel de símbolo para determinar si hay símbolos que distorsionan el ranking del grupo en su totalidad. Incluso, se puede realizar un análisis más fino, para determinar si ciertos símbolos particulares del grupo son los que le dan la condición de característica transversal al estado. En este punto, puede verse que un análisis a dos niveles: a nivel estado y luego a nivel símbolo dentro del estado.

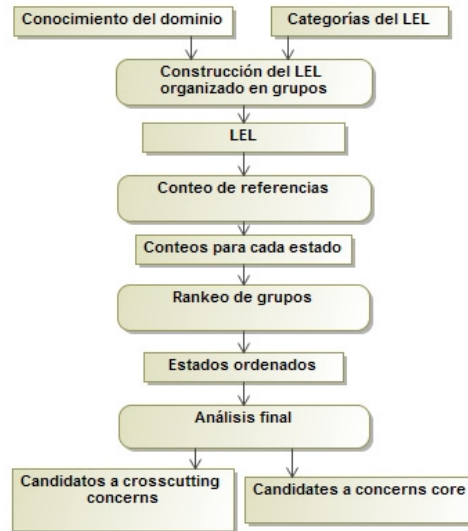


Figura 3-1. Resumen de la estrategia

### 3.2.1 Construcción de LEL organizado en grupos

La estrategia propuesta requiere que los símbolos del LEL identificados y descritos deben estar organizados en grupos determinados por los símbolos estados. Es necesario conformar grupos de símbolos porque se necesita organizar el conocimiento en algún bloque de trabajo con el fin de determinar que bloque representa funcionalidad *core* y que bloque representa funcionalidad *aspectual*. La organización en estados radica en que el software pueden ser descrito por máquinas de estado según Mahoney [Mahoney 2005] [Mahoney 2007]. El proceso de identificación y descripción de símbolos debe cumplir las siguientes etapas:

El primer paso es la construcción del LEL. El LEL debe ser construido de la forma tradicional siguiendo los siguientes pasos: (a) identificación de símbolos; (b) categorización de cada símbolo; (c) descripción de los símbolos de acuerdo a la categoría e (d) identificación de sinónimos.

El Segundo paso consiste en identificar al símbolo más representativo y significativo de la aplicación. Debe ser un sujeto o un objeto el cual sea el elemento más relevante en el contexto de la aplicación y que además posea una máquina de estados asociada. Luego de identificado el símbolo principal, es necesario identificar los estados vinculados con ese símbolo. Es decir, es necesario identificar y describir símbolos de categoría estados.

Como tercer paso, es necesario vincular a cada uno de los símbolos del LEL (de las categorías sujetos, objetos y verbos) con alguno de los estados determinados en el paso anterior. Un símbolo se

puede relacionar con un estado por diferentes razones. Se lo puede relacionar porque el símbolo aparece o es creado en el estado, o bien, porque el acoplamiento entre el símbolo y el estado implica una gran integración, a pesar de que el símbolo es creado en otro estado.

### 3.2.2 Conteo de referencias

El conteo de referencias consiste en calcular el número de referencias que cada grupo de símbolos posee desde los impactos de símbolos que se encuentran en otro estado. Es decir, para cada uno de los grupos de símbolos (determinado por cada estado del símbolo principal y todos los símbolos asociados a él) se deben revisar todos los impactos de todos los símbolos relacionados con otros estados, y computar aquellos en donde en la descripción de los impactos hacen referencia a símbolos del estado del que se están contando las referencias. Es importante el excluir las referencias entre los símbolos del mismo grupo, ya que se pretende medir el acoplamiento entre distintos grupos, es por ello que sólo hay que concentrarse en las referencias que trascienden desde un grupo a otro. Por otro lado, es importante notar que no se mide directamente la frecuencia absoluta de referencias, sino que se debe calcular el promedio. Sucede que los distintos grupos pueden estar conformados por distintos símbolos, por lo cual, grupos con más símbolos naturalmente tendrán más referencias, sin embargo, esta gran cantidad de referencias no se debe al nivel de acoplamiento entre los grupos, sino que se debe simplemente a la mayor cantidad de símbolos que un grupo puede tener. Es por ello que el análisis de acoplamiento se debe realizar el promedio de referencias que cada símbolo posee, y esto se debe realizar dividiendo la cantidad de referencias por la cantidad de símbolos que cada grupo posee. Además, es importante el contar las referencias desde los otros grupos por separado, es decir, en lugar de tener un único valor que represente la suma de todas las referencias, se deben acumular las referencias por cada uno de los distintos grupos. Sucede que esta es una medida de dispersión, puesto que determina el nivel de acoplamiento de cada grupo por separado, por lo cual, en definitiva indica cuando disperso está referenciado un grupo entre todos los demás. Finalmente, con el fin de obtener cierta precisión en el análisis, los promedios de referencias se deben calcular por cada categoría, es decir, se deben calcular las referencias producidas sólo por sujetos, sólo por objetos y sólo por verbos. Es decir, se debe determinar en que medida los sujetos de otros grupos referencian al grupo bajo análisis, y por separado como lo referencian los objetos y por otro lado los verbos. Esta distinción es importante ya que las metodologías tradicionales solo analizan los verbos (acciones, comportamiento) mientras que la estrategia propuesta además considera sujetos y objetos. Por lo cual, el hacer esta distinción permite comparar nuestra propuesta con otras.

### 3.2.3 Ranqueo de grupos

Van Den Berg et al. [Van Den Berg 2005] definen dos características que determinan la presencia de características transversales: *scattering* (dispersión) y *tangling* (enmarañamiento). Él determina que *scattering* ocurre cuando “un elemento origen es relacionado a múltiples elementos destinos”, mientras que *tangling* ocurre cuando “un elemento destino es relacionado a múltiples elementos fuentes”. Para el enfoque propuesto, los elementos (ya sean origen o destino) son los grupos de símbolos. Y se debe usar las referencias desde los impactos de los símbolos para medir *scattering* y *tangling*. Luego, los grupos candidatos a ser considerados características transversales deben satisfacer dos condiciones: (i) deben poseer muchos otros grupos que los referencian y (ii) deben poseer un alto promedio de referencias. Es importante maximizar ambas variables ya que (i) el número de grupos que los referencian indica cuan disperso (*scattered*) está el grupo, mientras que (ii) el promedio de referencias indica cuando acoplado (*tangled*, enmarañado) se encuentra el grupo.

Estas dos variables, determinan 3 situaciones posibles más. La Figura 3-2 muestra un diagrama xy donde el eje x representa el promedio de referencias (es decir, cuan acoplado el grupo es) mientras



que el eje y representa los grupos que referencian (es decir, cuando disperso el grupo está). El diagrama está sombreado de acuerdo a las dos variables, de forma tal que la porción más oscura indica la ubicación de las características transversales candidatas.

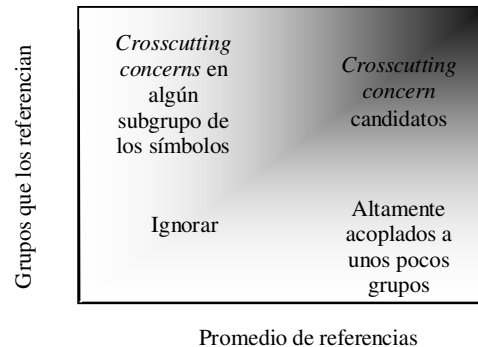


Figura 3-2. Nivel de dispersión (*scattering*) y acoplamiento (*coupling*)

### 3.2.4 Análisis final

Un análisis detallado debe ser realizado luego de que el conteo de referencias y el ranking se hayan realizado. El análisis detallado consiste en revisar los símbolos y determinar si algún símbolo podría alterar y sesgar el ranking de los grupos. Es necesario prestar atención a aquellos símbolos principales (sujetos, objetos y verbos) los cuales no se modularizarán como características transversales, sino que se hará como funcionalidad *core*. Estos símbolos podrían tener una gran cantidad de referencias y podrían generar que todo el grupo lo posea, por lo cual, el símbolo debe ser separado del grupo (no cuentan las referencias que recibe, sin embargo, si cuentan las referencias que posee). Como parte de este análisis final, es necesario analizar si dentro de un grupo además de haber símbolos *core* que no deben ser considerados características transversales también existen símbolos *aspectuales* los cuales surgen de un análisis más detallado y son más representativos de las características transversales que el mismo grupo.

## 4 Aplicabilidad, efectividad y usabilidad

La estrategia se aplicó a dos aplicaciones reales. Por un lado, una aplicación relacionada con un sistema antievasión de impuestos en Argentina y por otro lado, un portal web de una compañía que publica noticias en diferentes países de América Latina. Los casos de estudio tienen por objetivo mostrar la aplicabilidad de la estrategia en dos casos reales y no se pretende demostrar la efectividad, la cual se demuestra a través de un experimento. De todas formas, junto con cada caso de estudio se comparan los resultados obtenidos por la estrategia con la opinión de diferentes expertos que estuvieron involucrados con ambos sistemas.

La efectividad se ha probado comparando la estrategia del LEL contra la estrategia Theme/Doc [Baniassad 2004]. Con este fin, se definió una misma aplicación para la cual se aplicó ambas estrategias. Luego, se contrastaron los resultados en términos de las características transversales identificadas por cada una de las estrategias. La comparación se hizo a través de un experimento de tipo aleatorio simple entre sujetos (*simple between-subjects randomized*) con características realísticas.

Finalmente, la usabilidad está asegurada puesto que se desarrolló una herramienta que implementa la estrategia. La herramienta (llamada TICCWL) fue desarrollada como trabajo de grado [Aramayo 2012] e implementa los cálculos necesarios para identificar las características transversales, mientras que se apoya en otra herramienta (llamada C&L) para la edición del LEL. Es decir, TICCWL debe recibir el LEL construido a través de C&L [Felicísimo 2004] [Almentero 2009] [C&L 2009] y una vez importado, permite configurarlo, realizar los cálculos establecidos por la estrategia de identificación de características transversales del LEL y finalmente visualizar los resultados, para eventualmente hacer cambios de configuración y volver a realizar los cálculos.

## 5 Conclusiones

La presente tesis describe una estrategia para identificar características transversales del software (i.e. aspectos) en el lenguaje del contexto de la aplicación capturado a través del Léxico Extendido del Lenguaje. A continuación se describen los beneficios y limitaciones de la estrategia.

En primer lugar, la estrategia propuesta tiene la ventaja de ser muy fácil de usar. Para ello reúne varias características. La estrategia comienza con una descripción del contexto de la aplicación en lenguaje natural. Esta descripción se debe realizar a través del LEL. Esta tarea puede ser realizada por cualquier interesado (*stakeholder*) sin ninguna clase de experiencia y con poco esfuerzo. Hay experiencias en dominios complejos que validan la facilidad de utilización del LEL: Gil et al. [Gil 2000] y Cysneiros et al. [Cysneiros 2001]. Por otro lado, luego de construido el LEL la estrategia consiste en realizar sistemáticamente una serie de cálculos. Si bien los mismos son críticos y deben realizarse cuidadosamente, existe una aplicación que automatiza todos los cálculos y obtiene los resultados finales. Además, LEL que es una herramienta propuesta desde hace varios años por lo cual, la sobrecarga de trabajo necesario para identificar características transversales es mínimo y la ventaja de contar con esta información es por demás valiosa. Finalmente, la estrategia propuesta puede ser aplicada casi en cualquier etapa del desarrollo de software, incluidas las etapas de requerimientos y de codificación. El elemento clave que permite utilizar la estrategia en cualquier etapa es el LEL, ya que es él el que sintetiza el conocimiento de la aplicación, el cual puede originarse directamente del lenguaje de la aplicación, de la especificación de requerimientos o del código fuente. Este punto de la abstracción del LEL para identificar las características transversales, provee otra ventaja más. Las técnicas tradicionales generalmente se basan en las acciones para identificar características transversales, la estrategia propuesta se basa también en sujetos, objetos y estados.

A pesar de los beneficios, la estrategia presenta algunas limitaciones. La estrategia se basa en el LEL, por lo cual la descripción de los símbolos es determinante en la efectividad de la estrategia. En primer lugar, las descripciones deben ser lo más precisas como sea posible, ya que si el LEL no es correcto, las características transversales identificadas no serán de valor. Por otro lado, también es importante el nivel de detalle con que se describe cada uno de los estados. Los mismos se deben definir uniformemente, es decir, deben ser descriptos usando el mismo nivel de abstracción y detalle. Si un estado posee muchos símbolos, este estado potencialmente podría contener más referencias, y esto podría ocasionar que un estado sea identificado como característica transversal porque simplemente se lo describió con más detalle, mientras que otro estado que si debe serlo, se lo definió en forma muy sencilla. Es por ello que es necesario definir símbolos uniformemente con el mismo nivel y criterio de abstracción, incluso con el criterio de no sumar símbolos insignificantes que alteren los resultados. Y no sólo hay que considerar los símbolos a describir, también es necesario prestar atención a la descripción de los símbolos en sí. Ya que la descripción de los impactos afecta a la estrategia de la misma forma. Si un símbolo posee una descripción extensa con un gran nivel de detalle, mientras que el resto poseen descripciones más simples, el símbolo con el

mayor nivel de detalle afectará las referencias y con ello los conteos, y tal vez ocasionará que el estado al que pertenece sea considerado como característica transversal aunque no lo sea.

## 6 Trabajos futuros

Dado que la descripción del LEL es esencial para la efectividad de la estrategia, estamos trabajando en como asegurar que el LEL utilizado es de calidad. Para ello vamos a proponer guías bien precisas sobre como construir el LEL. Hemos realizado un experimento en donde 20 grupos de participantes construyen un LEL para una aplicación única y a partir de los inconvenientes que se presentan en el proceso de construcción y con ayuda de la literatura identificamos las guías. Por otro lado, la identificación de características transversales a partir del LEL es una de las componentes de la estrategia de desarrollo de software que estamos desarrollando. Estamos trabajando en obtener requerimientos funcionales, requerimientos no funcionales y cuestiones navegacionales a partir del LEL. De esta forma, con las características transversales, los requerimientos, atributos de calidad (requerimientos no funcionales) y cuestiones de navegación a partir del LEL se dispone de un núcleo importante de información para diseñar la arquitectura y planificar el desarrollo del software. Planificación que incluso será enriquecida con las métricas que también estamos diseñando estableciendo una relación con otras métricas como *use case points*.

## 7 Referencias

- [Almentero 2009] Kinder Almentero E. “Re-engenharia do software C&L para plataforma Lua-Kepler utilizando princípios de transparência” dissertação de Mestrado PUC Rio, Rio de Janeiro, 8 abril de 2009.
- [Antonelli 2010] Antonelli L, Rossi G, Leite JCSP “Early identification of crosscutting concerns in the domain model guided by states”, In: proceedings of the 2010 ACM Symposium on Applied Computing, Sierre, Switzerland, ISBN:978-1-60558-639-7, 2010.
- [Aramayo 2012] Aramayo A, Rossi J, “Una herramienta para identificar crosscutting concerns a través de LEL” tesis de grado en curso, Facultad de Informática, Universidad Nacional de La Plata, 2012.
- [Baniassad 2004] Baniassad E., Clarke S., “Finding Aspects In Requirements with Theme/Doc”, Position Paper for the Early-Aspects Workshop held as part of AOSD, March 2004.
- [Baniassad 2006] Baniassad E., Clements P.C., Araujo J., Moreira A., Rashid A., Tekinerdogan B., “Discovering early aspects, In: IEEE software”, ISSN:0740-7459, Volume 23, Issue 1, January, pp 61 – 70, 2006.
- [Boehm 1997] Boehm B. W., Software Engineering, Computer society Press, IEEE, 1997.
- [Bounour 2006] Bounour N., Ghoul S., Atil F., “A comparative classification of aspect mining approaches”, Journal of Computer Science Volume 2, Number 4, ISSN 1549-3636, 2005 Science publication, pp 322 - 325, 2006.
- [Breitman 2003] Breitman K.K., Leite J.C.S.P., “Ontology as a Requirements Engineering Product”, In: Proceedings of the 11th IEEE International Conference on Requirements Engineering (RE), IEEE Computer Society, Monterey Bay, California, USA, ISBN 0-7695-1980-6, 2003.
- [C&L 2009] C&L Tool, available at <http://pes.inf.puc-rio.br/cel/>, accessed in October, 2009.

- [Cysneiros 2001] Cysneiros L.M., Leite J.C.S.P., “Using the Language Extended Lexicon to Support Non-Functional Requirements Elicitation” in proceedings of the Workshops de Engenharia de Requisitos, Wer’01, Buenos Aires, Argentina, 2001.
- [Felicíssimo 2004] Felicíssimo C. H., Leite J.C.S.P., Breitman K. K., Fernandes da Silva L., “C&L: Um Ambiente para Edição e Visualização de Cenários e Léxicos”, in sessão de Ferramentas do Simpósio Brasileiro de Engenharia de Software Brasília, Brasil, October, pp 43 – 48, 2004.
- [Gil 2000] Gil D., Figueroa D. A., Oliveros A., “Producción del LEL en un Dominio Técnico.Informe de un caso.”, in proceedings of the Workshops de Engenharia de Requisitos, Wer’00, Rio de Janeiro, Brazil, 2000.
- [Gruber 1993] Gruber T.R., “A translation approach to portable ontology specifications”, In: Knowledge acquisition journal, volume 5, June 1993, pp 199 – 220, 1993.
- [Leite 1993] Leite J.C.S.P., Franco A.P.M., “A Strategy for Conceptual Model Acquisition”, In: Proceedings of the First IEEE International Symposium on Requirements Engineering, San Diego, California, IEEE Computer Society Press, pp 243-246, 1993.
- [Mahoney 2005] Mahoney M., “Modeling Crosscutting Concerns in Reactive Systems with Aspect-Orientation”, In: Models 2005, Doctoral Symposium, 2005.
- [Mahoney 2007] Mahoney M., Elrad T., “Generating code from scenario and state based models to address crosscutting concerns”, In: proceedings of the sixth international workshop on scenarios and state machines (SCESM’07), IEEE, 2007.
- [Nuseibeh 2004] Nuseibeh B., “Crosscutting requirements, In: proceedings of the 3rd international conference on Aspect-oriented software development”, ISBN:1-58113-842-3, Lancaster, UK, pp 3 – 4, 2004.
- [Rago 2009] Rago A., Marcos C., “Análisis Semántico para la Identificación de Aspectos”, in Jornadas Chilenas de Computación, 2009.
- [Rashid 2003] Rashid A., Moreira A., Araújo J., “Modularisation and composition of aspectual requirements”, In: proceedings of the 2nd international conference on Aspect-oriented software development, ISBN:1-58113-660-9, Boston, Massachusetts, pp 11 – 20, 2003.
- [Rashid 2006] Rashid A., Moreira A., “Domain Models are NOT Aspect Free”, In: Proceedings of MoDELS/UML, Springer, Lecture Notes in Computer Society, pp. 155-169, 2006.
- [Sampaio 2005] Sampaio A., Loughran N., Rashid A., Rayson P., “Mining Aspects in Requirements”, In Proceeding of the Workshop on Early Aspects (held with AOSD 2005) Illinois, Chicago, USA, 2005.
- [Van Den Berg 2005] Van Den Berg K., Conejero J.M., “A Conceptual Formalization of Crosscutting in AOSD”, In: Proceeding of Desarrollo de Software Orientado a Aspectos, Granada, España, 2005.
- [Wood 1997] Wood L.E., Semi-structured interviewing for user-centered design, Interactions of the ACM, april-may, pp 48-61, 1997.
- [Yu 2004] Yu Y., Leite J.C.S.P., Mylopoulos J., “From Goals to Aspects: Discovering Aspects from Requirements Goal Models”, In: International Requirements Engineering Conference, pp 38-47, 2004.



## EXTRACCIÓN DE CONOCIMIENTO EN GRANDES BASES DE DATOS UTILIZANDO ESTRATEGIAS ADAPTATIVAS

Waldo Hasperué

Directores: Armando De Giusti y Laura Lanzarini

III-LIDI. Facultad de Informática. Universidad Nacional de La Plata

Fecha de exposición: 28 de marzo de 2012

### 1. Resumen

Hoy en día, dada la existencia de enormes volúmenes de información digital, resulta de interés contar con técnicas que permitan, en una primera etapa, analizar tal información y obtener conocimiento nuevo a partir de ella. Además, como se espera que la información disponible se modifique o incremente a lo largo del tiempo, en una segunda etapa, sería relevante poder adaptar el conocimiento adquirido a los cambios o variaciones que ocurran en el conjunto de datos original.

El objetivo general de esta tesis es la definición de una nueva técnica de minería de datos capaz de generar conocimiento útil, produciendo resultados provechosos para el usuario final.

El aporte central es la definición e implementación de una técnica adaptativa que permite obtener modelo dinámico, formado por reglas de clasificación, capaz de adaptarse a los cambios de la información.

### 2. Hiper-rectángulos

La técnica presentada en esta tesis basa su funcionamiento en el manejo de hiper-rectángulos como descriptores de las diferentes clases de datos presentes en la base de datos estudiada.

Los hiper-rectángulos son polítopos  $D$ -dimensionales donde todas sus caras tienen forma de rectángulo. Así, si un problema dado se presenta en dos dimensiones se trabaja simplemente con rectángulos y si es en tres dimensiones se hace con los cuerpos en el espacio conocidos como ortoedros o cuboides.

Los límites de un hiper-rectángulo en cada una de las dimensiones del espacio son utilizados como descriptores de un determinado conjunto de datos. La principal ventaja de trabajar con hiper-rectángulos es que cada una de estas caras presenta un único valor en una dimensión determinada. En cada eje del espacio un hiper-rectángulo tiene dos caras ortogonales a dicho eje, por lo tanto cada cara tiene un valor distinto, así es posible utilizar estos dos valores para definir los límites del hiper-rectángulo en la dimensión correspondiente. De esta manera, cada dimensión queda definida dentro de un intervalo cerrado y los límites de todo el hiper-rectángulo queda definido por todos los intervalos de cada una de las dimensiones del espacio. A partir de un hiper-rectángulo con las características mencionadas es posible describir sus límites utilizando el lenguaje natural.

En el espacio  $D$ -dimensional dos o más hiper-rectángulos pueden presentar una superposición la cual no es deseable ya que, si ese caso ocurre, un mismo dato podría pertenecer a más de una clase al estar incluido en ambos hiper-rectángulos. El objetivo de la técnica propuesta en esta tesis es el de reducir las superposiciones existentes entre hiper-rectángulos de distintas clases y en el mejor de los casos eliminar todas las superposiciones existentes.

#### 2.1. Índices de superposición

La eliminación de una superposición entre dos hiper-rectángulos se hace modificándolos en una dimensión en particular, ya sea por la división de uno de ellos o por la reducción de volumen de uno o

ambos. La elección de la dimensión en la cual llevar a cabo la tarea de ajuste se basa en el cálculo de una serie de índices de superposición, donde cada uno de ellos mide un aspecto diferente de la superposición. Estos índices se combinan en un único valor para cada dimensión de manera que el índice con el valor más alto determina la dimensión por la cual llevar a cabo el ajuste.

Cada índice es calculado para todas las dimensiones. Para llevar a cabo el cálculo del índice se utilizan los valores mínimos y máximos de los hiper-rectángulos y los valores de los datos de ambas clases presentes en la superposición. De esta manera se define una superposición de área que queda determinada por los límites de la superposición y una superposición de datos que queda determinada por los datos incluidos en la superposición.

En esta tesis se proponen seis índices los cuales miden diferentes aspectos de una superposición. Cada uno de estos índices, denominados índices  $Z$ , son calculados para los dos hiper-rectángulos presentes en una superposición determinada y se calcula en cada una de las dimensiones del espacio del problema.

Los seis índices están diseñados para devolver un valor entre 0 y 1. El índice calculado en una dimensión vale 0 cuando en esa dimensión no exista un aporte importante para la decisión de dividir los hiper-rectángulos y vale 1 cuando esa dimensión es candidata para realizar la división.

Finalmente, se utilizan los valores calculados de los índices  $Z$  para el cálculo del índice  $\Omega$  en cada una de las dimensiones. Este índice  $\Omega$  será utilizado para determinar el grado de separabilidad de los hiper-rectángulos en dicha dimensión. Por lo tanto, la dimensión en la cual realizar el ajuste de hiper-rectángulos queda determinada por el índice  $\Omega$  con mayor valor.

En la técnica de clasificación presentada en esta tesis se propone utilizar como índice  $\Omega$  el promedio de los seis índices  $Z$ . Este índice  $\Omega$ , al ser calculado como un promedio, también dará un único valor en el intervalo  $[0, 1]$  con la particularidad de que el valor 1 significa que es una dimensión candidata para llevar a cabo la división y 0 lo contrario.

De esta manera, se calcula un índice  $\Omega_i$  para cada dimensión  $i$  el cual es calculado además para cada uno de los dos hiper-rectángulos presentes en la superposición, calculando así pares de índices  $\Omega$  en todas las superposiciones presentes en el modelo de datos. De todos los índices  $\Omega_i$  calculados, aquel con mayor valor determina el hiper-rectángulo a dividir y en que dimensión hacer la división.

### 3. CLUHR

La técnica propuesta en esta tesis, denominada CLUHR [1] (Clasificación utilizando hiper-rectángulos), permite, una vez que se consigue armar el modelo de datos, extraer como conocimiento reglas de clasificación, las cuales se obtienen de los hiper-rectángulos formados. Es posible encuadrar esta técnica en la categoría de los algoritmos de “divide y vencerás”, ya que dicho algoritmo comienza formando un único hiper-rectángulo para cada clase y luego mediante la división de estos logra armar el modelo de datos. Finalizado el modelo de datos es posible extraer un conjunto de reglas de clasificación, este conjunto representa a un conjunto de reglas por cobertura, es decir, no todo el espacio está cubierto por las reglas, e incluso y dependiendo de como se configure el algoritmo para el armado del modelo, algunas de ellas podrían cubrir una misma parte del espacio de los datos.

CLUHR es una técnica determinista y el resultado solo depende de la configuración elegida por el usuario al momento de armar el modelo. Esta es una de las principales ventajas de esta técnica ya que una misma entrada, con la misma configuración del algoritmo, produce siempre la misma salida.

La filosofía del algoritmo de CLUHR para el armado del modelo de datos es la siguiente: *mientras existan superposiciones de hiper-rectángulos, eliminarlas.*

El algoritmo tiene una etapa de inicialización en la cual crea, para cada clase de datos presente en la base de datos un hiper-rectángulo. De esta manera, el algoritmo comienza con tantos hiper-rectángulos como clases haya en la base de datos.

El proceso de clasificación y armado del modelo de datos consiste en un proceso iterativo en el cual se buscan todas las superposiciones existentes entre dos hiper-rectángulos de distintas clases. Luego se procede a realizar el cálculo de los índices  $\Omega$  para cada uno de los hiper-rectángulos presentes en las superposiciones encontradas y en todas las dimensiones del espacio del problema, para luego buscar aquella superposición en la cual el índice  $\Omega$  tenga el valor más alto. En la superposición hallada se realiza el ajuste de los hiper-rectángulos involucrados dividiendo al que tiene el valor  $\Omega$  más alto. Finalmente se ajustan los hiper-rectángulos que fueron modificados y los nuevos hiper-rectángulos que fueran creados a sus correspondientes datos formando así nuevos hiper-rectángulos.

El algoritmo vuelve a buscar superposiciones entre todos los hiper-rectángulos y se repite el proceso anterior. Este proceso iterativo continúa hasta eliminar todas las superposiciones existentes.

Al finalizar el algoritmo, cada clase tendrá uno o más hiper-rectángulos representativos y cada uno de ellos describe una regla de clasificación para la clase.

Es interesante mencionar que la división de los hiper-rectángulos puede resultar en un hiper-rectángulo el cual represente a un único dato, en cuyo caso se obtiene una regla de clasificación para ese único dato. Para lidiar con este problema, el algoritmo posee un parámetro  $\mu$  cuya tarea es la de determinar que superposiciones se analizan cuando la cantidad de datos involucrados de una o ambas clases es mayor que el valor especificado en  $\mu$ .

Con el uso de este parámetro, el algoritmo solo elimina aquellas superposiciones que tienen una cantidad mínima de datos participantes dejando sin modificar las que el experto o el usuario del modelo consideran insignificantes. Obviamente, el no eliminar estas superposiciones con pocos datos produce como resultado un modelo de datos menos preciso, por lo que el uso de este parámetro resulta clave para encontrar un equilibrio entre obtener un modelo de datos preciso versus una cantidad mínima de reglas de clasificación.

### 3.1. Extracción de las reglas

Cuando el algoritmo de clasificación finaliza su tarea, se obtiene como resultado un conjunto de hiper-rectángulos donde cada uno de ellos representa a un subconjunto de datos de una clase específica. De cada uno de los hiper-rectángulos se extrae una regla de clasificación del tipo IF-THEN, esta regla de clasificación tendrá la siguiente forma:

$$\text{IF } ((x_1 \geq Hn_1) \text{ AND } (x_1 \leq Hx_1) \text{ AND } \dots \text{ AND } (x_D \geq Hn_D) \text{ AND } (x_D \leq Hx_D)) \text{ THEN } C$$

donde  $Hn_i$  y  $Hx_i$  son los valores mínimos y máximos en la dimensión  $i$  del hiper-rectángulo  $H$  respectivamente.

Esta regla estará formada por  $2 \cdot D$  cláusulas y como es fácil imaginar la utilización de este tipo de reglas puede resultar muy engorroso para el usuario, en especial si la dimensión del problema a tratar es alta. Por lo tanto es necesario un proceso post-clasificación que simplifique al máximo la complejidad de estas reglas de clasificación. Este proceso consiste en eliminar las cláusulas que no son necesarias

para la descripción del problema. Así, se extrae un conjunto de reglas simplificadas que tienen la mínima cantidad de cláusulas necesarias para describir el conjunto de datos del propio hiper-rectángulo sin que se presente contradicción o ambigüedad con otra regla de otro hiper-rectángulo de otra clase.

### **3.2. Intervención del experto**

El algoritmo de CLUHR, para llevar a cabo el armado del modelo de datos, debe tomar muchas decisiones. La decisión de que superposición resolver es decidida mediante el cálculo de los índices  $\Omega$ . Es posible calcular este índice de muchas formas y una vez que se estableció que superposición eliminar surge la duda de si hay que dividir los hiper-rectángulos, disminuir su volumen, cual dividir o como llevar a cabo la disminución. En esta dirección, la técnica propuesta en esta tesis no determina ninguna decisión en particular, ya que esta depende fuertemente del problema.

De todas formas, es posible tomar todas estas decisiones previo al armado del modelo, configurar el algoritmo para que lleve a cabo el armado y ejecutarlo de manera totalmente automática para lograr el resultado. Aunque la técnica propuesta en esta tesis es completamente flexible en el sentido que un experto, previo a la ejecución de proceso, puede determinar que índices utilizar y como llevar a cabo las divisiones y así personalizar el proceso automático, en problemas reales nunca es posible armar un proceso automático que sea capaz de mejorar a las decisiones “on-line” que tomaría un humano, más cuando este es experto en el dominio del problema.

En esta dirección, la técnica presentada en esta tesis presenta la particularidad de ser utilizada como un proceso totalmente automático o un proceso interactivo mediante la intervención de un experto en el dominio del problema

### **3.3. Adaptabilidad del modelo**

Desde hace varios años se han propuesto varias técnicas en las cuales un modelo de datos ya armado actualiza su estructura interna ante los cambios producidos en la base de datos. Estas técnicas, incluidas en el área de la minería de datos incremental, se han aplicado a distintas tareas entre las que se incluyen clustering, frequent pattern mining y clasificación. La principal ventaja que presentan estas técnicas es que el modelo de datos armado es adaptado a los nuevos cambios sin necesidad de llevar a cabo todo el proceso desde cero, logrando así una técnica mucho más eficiente, ya que, en vez de re-entrenar el modelo de datos con la base de datos completa, sólo se modifica aquella parte de la estructura interna del modelo de datos que es afectada ante el cambio de la base de datos.

En esta dirección, la técnica propuesta en esta tesis es capaz de actualizar su estructura interna de manera eficiente ante la aparición de nuevos datos y ante la eliminación o modificación de los datos ya representados por el modelo.

Una técnica que es adaptable posee la capacidad de realizar con muy poco esfuerzo computacional las modificaciones pertinentes en su estructura interna para reflejar los cambios de los datos. La capacidad de adaptación logra un proceso de extracción de conocimiento más eficiente y transparente en la tarea de clasificación. Esto hace que resulte una técnica más eficiente y amigable para el usuario por las modificaciones mínimas necesarias que se realizan. Además es transparente, ya que el propio usuario puede ver qué parte del modelo sufre los cambios, y en el caso puntual de CLUHR que reglas de clasificación sufren modificación.



La adaptabilidad llevada a cabo por CLUHR es “on-line”, realiza solo mínimas modificaciones en su estructura interna y, por lo tanto, el conocimiento adquirido previamente “sufre” pequeños cambios.

#### 4. Resultados

CLUHR ha sido comparado contra otras técnicas de obtención de reglas utilizando bases de datos del repositorio UCI. Las técnicas contra las cuales se compararon los resultados fueron el clásico C4.5 [2], una estrategia que, al igual que la propuesta, utiliza hiper-rectángulos junto con un algoritmo evolutivo para obtener las reglas de clasificación [3] y una estrategia que utiliza PSO junto con ACO para obtener el conjunto de reglas de clasificación [4].

Se compara la precisión promedio alcanzada por el modelo de datos (tabla 1), la cantidad de reglas extraídas (tabla 2), la cantidad promedio de cláusulas de cada una de las reglas (tabla 3) y el número de veces que se recorre la base de datos durante el armado del modelo (tabla 4).

En la estrategia propuesta se cuenta como número de reglas a la cantidad de hiper-rectángulos del modelo de datos. En el algoritmo C4.5 de cada rama del árbol se extrae una regla de clasificación y la longitud de la rama determina el número de cláusulas. El método de PSO/ACO2 arroja como resultado el conjunto de reglas de clasificación.

Dado que la precisión de un modelo de datos y la cantidad de reglas son resultados inversamente proporcionales entre sí, ya que se puede lograr una muy buena precisión pero con una gran cantidad de reglas, y al mismo tiempo se puede obtener un número muy reducido de reglas pero con una mala precisión, en los ensayos realizados con CLUHR y con C4.5 se eligieron los parámetros correspondientes para lograr un equilibrio razonable entre precisión y cantidad de reglas y así poder comparar contra los resultados publicados en [4].

*Tabla 1. Exactitud del modelo logrados por cada una de las estrategias medidas y para cada una de las bases de datos ensayadas. Se presentan las medias y entre paréntesis el desvío estándar.*

	<i>C4.5</i>	<i>EHS-CHC</i>	<i>PSO/ACO2</i>	<i>CLUHR</i>
Ecoli	0,7964 (0,0141)	0,7948	-	0,7891 (0,0160)
Glass	0,6576 (0,0302)	0,6287	0,7095 (0,075)	0,6215 (0,0360)
Haberman	0,7103 (0,0202)	0,7122	-	0,7356 (0,0064)
Image	0,8586 (0,0155)	-	0,9667 (0,0117)	0,8538 (0,0135)
Ionosphere	0,9054 (0,0151)	-	0,8806 (0,0491)	0,8777 (0,0169)
Iris	0,9420 (0,0077)	0,9267	0,9467 (0,0526)	0,9300 (0,0079)
Liver	0,6418 (0,0300)	0,6167	-	0,5918 (0,0211)
Pima	0,7434 (0,0093)	0,7384	-	0,5595 (0,0191)
Sonar	0,7235 (0,0247)	0,7650	0,7505 (0,0911)	0,6666 (0,0283)
Vehicle	0,7111 (0,0099)	-	0,7305 (0,0445)	0,6819 (0,0171)
Vowel	0,6008 (0,0158)	-	0,8616 (0,0347)	0,7120 (0,0132)
Wine	0,9141 (0,0145)	0,9490	-	0,9530 (0,0113)
Wisconsin	0,9446 (0,0047)	0,9599	0,9487 (0,0253)	0,9251 (0,0102)
Forest covertype	0,7063 (0,0187)	-	-	0,6928 (0,0149)

Tabla 2. Cantidad de reglas extraídas por cada una de las estrategias medidas y para cada una de las bases de datos ensayadas. Se presentan las medias y entre paréntesis el desvío estándar.

	C4.5	EHS-CHC	PSO/ACO2	CLUHR
Ecoli	12,1 (1,45)	11,1	-	12,62 (1,44)
Glass	14,8 (0,79)	12,2	20,4 (1,35)	15,17 (1,30)
Haberman	10,7 (3,62)	4,4	-	4,29 (0,33)
Image	10,6 (0,70)	-	21,9 (0,99)	10,93 (0,47)
Ionosphere	10,2 (2,04)	-	3,6 (0,97)	3,98 (0,37)
Iris	4,0 (0,47)	3,4	3,0 (0,00)	3,21 (0,12)
Liver	23,9 (4,46)	9,8	-	17,79 (2,21)
Pima	13,2 (1,40)	11	-	10,45 (0,91)
Sonar	10,9 (1,60)	10,3	4,4 (1,58)	4,14 (0,20)
Vehicle	31,0 (2,31)	-	37,8 (1,2)	32,35 (2,03)
Vowel	32,8 (2,20)	-	29,0 (0,82)	31,74 (0,78)
Wine	5,1 (0,57)	3,6	-	3,18 (0,11)
Wisconsin	11,9 (1,79)	3,8	10,2 (1,87)	9,63 (1,39)
Forest covertype	39,7 (2,35)	-	-	41,25 (2,05)

Debido a que las ejecuciones del algoritmo C4.5 fueron llevadas a cabo para elaborar esta tesis y que los autores de [4] publican en sus resultados la media, desvío estándar y n para cada base de datos, se realiza una prueba *t*-student de doble cola con nivel de confianza del 95% para determinar si las diferencias logradas entre CLUHR y C4.5 y PSO/ACO2 son estadísticamente significativas o no. La tabla 5 muestra los resultados obtenidos. En dicha tabla se marca con un signo “+” cuando CLUHR es mejor estadísticamente, con un signo “-” cuando CLUHR es peor estadísticamente y con un signo “=” cuando no hay diferencias significativas.

En la tabla 4 las estrategias EHS-CHC y PSO/ACO2 no aparecen por desconocer la verdadera cantidad de veces que se recorre. Aunque, como se detalla en esta tesis, la estrategia EHS-CHC se estima que la recorre 10000 veces y en PSO/ACO2 se estima un promedio de 3000 veces. De esta misma tabla puede observarse que CLUHR recorre menos veces la base de datos comparado con C4.5. Se realizó un test *t*-student de doble cola al 95% y en la tabla figura un símbolo “+” cuando CLUHR es mejor estadísticamente, un símbolo “-” cuando CLUHR es peor estadísticamente y con un símbolo “=” cuando la diferencia no resulta estadísticamente significativa. Hay casos donde el número de veces es similar, pero en otras se recorre una, dos y hasta tres veces menos que las recorridas por C4.5. Solo en dos bases de datos CLUHR necesitó recorrer más veces la base de datos que C4.5.

Haciendo un estudio de los distintos resultados obtenidos no es posible determinar que CLUHR se destaque sobre el resto, tampoco lo contrario, que CLUHR sea una técnica mala comparada contra el resto. Comparado con C4.5 los resultados han sido muy similares mientras que PSO/ACO2 parecería lograr un número reducido no solo de reglas, sino también de cláusulas por regla. Esto último se debe a dos factores claves que presenta la propia estrategia. 1) al tener ordenado el conjunto de reglas permite eliminar muchas cláusulas de las mismas. 2) al ser una estrategia de optimización, las partículas de PSO recorren todo el espacio de búsqueda encontrando una solución óptima.

Comparando la precisión, el número de reglas y el promedio de cláusulas por regla podemos afirmar que CLUHR es equivalente al resto de las técnicas estudiadas produciendo resultados similares a los que arrojan C4.5 y PSO/ACO2. Sin embargo, si se analiza el costo computacional de cada método para alcanzar el resultado, puede afirmarse que CLUHR es el mejor.

En resumen, CLUHR demostró resolver todos los problemas estudiados ofreciendo resultados similares a las estrategias comparadas, pero con una utilización de recursos similar a la que presenta C4.5 y mucho menor a los utilizados por EHS-CHC y PSO/ACO2.

*Tabla 3. Número promedio de cláusula por regla por cada una de las estrategias medidas y para cada una de las bases de datos. Se presentan las medias y entre paréntesis el desvío estándar.*

	<i>C4.5</i>	<i>PSO/ACO2</i>	<i>CLUHR</i>
Ecoli	4,32 (0,30)	-	4,65 (0,15)
Glass	5,68 (0,75)	3,11 (0,18)	5,37 (0,18)
Haberman	4,54 (1,27)	-	2,54 (0,06)
Image	4,31 (0,58)	2,8 (0,27)	3,74 (0,10)
Ionosphere	5,36 (0,89)	3,33 (0,79)	5,17 (0,18)
Iris	2,25 (0,27)	0,93 (0,14)	2,08 (0,05)
Liver	6,80 (1,30)	-	5,01 (0,06)
Pima	4,55 (0,27)	-	5,27 (0,12)
Sonar	3,99 (0,43)	2,6 (0,63)	16,27 (0,72)
Vehicle	7,10 (0,34)	3,85 (0,18)	7,38 (0,33)
Vowel	5,69 (0,18)	4,2 (0,25)	8,13 (0,27)
Wine	2,46 (0,17)	-	4,08 (0,09)
Wisconsin	4,31 (0,39)	1,21 (0,07)	3,59 (0,11)
Forest	6,67 (0,82)	-	6,49 (0,48)
covertime			

*Tabla 4. Número de veces que se recorre la base de datos por CLUHR y C4.5.*

	<i>C4.5</i>	<i>CLUHR</i>	<i>Significancia</i>
Ecoli	4,19 (0,39)	3,53 (0,33)	+
Glass	5,64 (1,10)	3,97 (0,37)	+
Haberman	3,61 (1,26)	5,28 (0,32)	-
Image	3,84 (0,35)	1,67 (0,06)	+
Ionosphere	5,78 (0,73)	2,47 (0,14)	+
Iris	2,02 (0,13)	1,5 (0,06)	+
Liver	6,59 (1,48)	5,20 (0,50)	+
Pima	3,74 (0,24)	4,97 (0,39)	-
Sonar	4,03 (0,49)	2,41 (0,17)	+
Vehicle	5,98 (0,24)	5,06 (2,56)	=
Vowel	5,54 (0,13)	2,99 (0,11)	+
Wine	2,34 (0,10)	1,20 (0,01)	+
Wisconsin	3,19 (0,35)	3,02 (0,32)	=
Forest	5,71 (0,72)	5,24 (0,45)	=
covertime			
Total			+7

Tabla 5. Resultados de la prueba t-student para determinar si hay diferencias significativas entre los resultados obtenidos por CLUHR y los obtenidos por C4.5 y PSO/ACO2.

	<i>Exactitud</i>		<i>Número de reglas</i>		<i>Promedio de cláusulas por regla</i>	
	PSO/AC		PSO/AC		PSO/AC	
	C4.5	O2	C4.5	O2	C4.5	O2
Ecoli	=		=		-	
Glass	-	-	=	+	=	-
Haberman	+		+		+	
Image	=	-	=	+	+	-
Ionosphere	+	=	+	=	=	-
Iris	-	=	+	-	=	-
Liver	-		+		+	
Pima	-		+		-	
Sonar	-	-	+	+	-	-
Vehicle	-	-	=	=	-	-
Vowel	+	-	+	-	-	-
Wine	+		+		-	
Wisconsin Forest	-	-	+	+	+	-
covertype	=		-		=	
Total	-3	-6	+8	+2	-2	-8

En cuanto a la característica de estrategia incremental CLUHR es comparado contra la técnica ITI [5]. El principal problema que presenta la técnica ITI, como cualquier técnica basada en árboles de decisión es que la acumulación de datos y la re-evaluación de la función de decisión de los nodos provoca que tarde o temprano se requiera una re-estructuración de un sub-árbol. Cuando el sub-árbol a rehacer tiene como raíz un nodo de los primeros niveles, entonces esta re-estructuración es importante, ya que la reconstrucción de uno de estos sub-árboles representa recorrer un importante porcentaje de la base de datos.

En CLUHR, la aparición de nuevos datos, sólo causa que se modifique un hiper-rectángulo. Y si este tiene superposiciones con otros hiper-rectángulos entonces, se modifican los hiper-rectángulos involucrados.

Para comparar estos métodos se mide la cantidad de veces que se recorre la base de datos para la modificación del modelo. Se realizan 10 ejecuciones independientes para ambas técnicas, CLUHR e ITI. En cada ejecución se divide la base de datos en dos conjuntos al azar, el primero con el 70% de los datos y el segundo con el 30% restante. Con el primer sub-conjunto de datos se arma un modelo, tanto para CLUHR como para ITI y, con los datos del segundo-subconjunto se presentan uno por uno al modelo armado. Ante cada dato presentado se cuenta la cantidad de veces que se recorre la base de datos. Por lo tanto, para una ejecución, la cantidad total de veces que se recorre la base de datos es la suma de las veces que se recorre para cada dato.

La tabla 6 muestra los resultados comparativos entre el esfuerzo computacional que necesita CLUHR y el que necesita la técnica ITI. Se realizó un test t-student de doble cola al 95% y en la tabla figura un



símbolo “+” cuando CLUHR es mejor estadísticamente, un símbolo “-” cuando CLUHR es peor estadísticamente y con un símbolo “=” cuando la diferencia no resulta estadísticamente significativa. Como puede observarse CLUHR requiere mucho menos esfuerzo computacional que el que necesitan los árboles de decisión basados en el algoritmo ITI. CLUHR fue muy superior en 10 bases de datos.

*Tabla 6. Recursos utilizados (cantidad de veces que se recorre la base de datos) por CLUHR e ITI en cada una de las bases de datos.*

	<i>ITI</i>	<i>CLUHR</i>	<i>Significancia</i>
Ecoli	5,19 (0,95)	0,59 (0,45)	+
Glass	14,50 (2,56)	0,44 (0,11)	+
Haberman	12,84 (2,24)	0,99 (0,25)	+
Image	2,37 (0,44)	0,19 (0,15)	+
Ionosphere	1,71 (0,30)	1,59 (0,43)	=
Iris	0,25 (0,05)	0,90 (0,50)	-
Liver	14,58 (2,70)	0,61 (0,16)	+
Pima	21,53 (3,69)	1,31 (0,37)	+
Sonar	5,11 (0,90)	0,06 (0,05)	+
Vehicle	14,42 (2,48)	1,07 (0,32)	+
Vowel	31,20 (5,30)	0,11 (0,05)	+
Wine	1,32 (0,26)	0,26 (0,41)	+
Wisconsin	1,65 (0,30)	8,31 (2,11)	-
Total			+8

## 5. Conclusiones y Trabajo a futuro

CLUHR ha demostrado ser una poderosa estrategia para clasificación y extracción de conocimiento en donde se destacan las siguientes características:

- Ofrece similares resultados en cuanto a precisión, cantidad de reglas y cantidad de cláusulas por regla que se obtienen con otras estrategias de clasificación.
- Es una estrategia determinista que permite obtener el mismo resultado con la misma entrada.
- Consume mucho menos recursos que las estrategias de optimización y un número ligeramente menor a otras estrategias deterministas como C4.5.
- El algoritmo puede ser ejecutado de manera totalmente automática o totalmente supervisada pudiendo un experto en el dominio del problema participar en todas las decisiones que debe tomar la estrategia para lograr un modelo de datos.
- La estrategia es completamente personalizable pudiéndose elegir que índices de superposición utilizar a lo largo del procedimiento de armado del modelo.
- El algoritmo tiene un único parámetro lo que lo hace sencillo de utilizar. Además, mediante sus cálculos es posible obtener diferentes resultados con diferentes valores de este parámetro con el mismo costo computacional de una sola ejecución del algoritmo.
- Es una estrategia adaptativa, esto permite modificar su estructura interna con poco esfuerzo sin llevar a cabo una ejecución completa del armado del modelo. Esta adaptación puede ser usada de manera automática o supervisada por un experto.

Aún con todas las ventajas que presenta la estrategia propuesta quedan varios aspectos en los cuales se puede seguir investigando. Los mismos son detallados a continuación.

- En cuanto a los índices de superposición es posible seguir investigando el desarrollo de nuevos índices que midan otros aspectos de una superposición, como por ejemplo índices que midan características de más de dos clases al mismo tiempo o que midan aspectos de más de una dimensión a la vez.
- En ciertas situaciones es posible unir los hiper-rectángulos más pequeños de una clase en un único hiper-rectángulo, logrando así menos reglas en el modelo de datos. Es interesante investigar si las uniones de los hiper-rectángulos de una misma clase pueden ser llevadas a cabo durante el proceso de armado del modelo o es un tratamiento extra que debería hacerse en la etapa final de extracción de reglas simplificadas. El lograr mejorar este aspecto favorecería al procedimiento de extracción de reglas simplificadas, ya que este procedimiento podrá trabajar con una menor cantidad de reglas.
- Una de los puntos débiles de CLUHR es que el proceso final de extracción del conjunto de reglas simplificadas utiliza un procedimiento greedy del orden  $O(n^2)$ . Resulta interesante investigar de qué manera “marcar” aquellas caras de los hiper-rectángulos que representen un límite en el espacio de los datos. Si es posible identificar aquellas caras que representan los límites del espacio de datos se producirá la exclusión directa de la cláusula para la correspondiente regla sin necesidad de llevar a cabo el método de simplificación post-procesamiento.
- Otro aspecto débil en la estrategia propuesta es que solo es capaz de trabajar en dominios donde todos los atributos son numéricos. En esta dirección sería interesante poder trabajar de manera más confiable con atributos numéricos de pocos valores y con la posibilidad de adaptar la estructura de los hiper-rectángulos al dominio de datos nominales. Esto podría obtenerse quizás trabajando con conjuntos de datos nominales asociados a los hiper-rectángulos.
- Un experto en el dominio del problema puede supervisar el armado del modelo de datos participando en las decisiones que debe tomar la propia estrategia al momento de eliminar una superposición. Resulta útil contar con una herramienta que visualice las características de las superposiciones para que el experto pueda tomar una decisión acorde al problema.

### **Bibliografía**

- [1] Hasperué, W.; Lanzarini, L.; De Guisti. 2012. Rule Extraction on Numeric Datasets Using Hyperrectangles. *A Computer and Information Science*. Vol. 5, No 4, pp. 116-131.
- [2] Quinlan, J. R. 1993. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, Inc., 1993. ISBN 1-55860-238-0.
- [3] Garcia, S., Derrac, J., Luengo, J., & Herrera, F. 2009. A First Approach to Nearest Hyperrectangle Selection by Evolutionary Algorithms. *Intelligent Systems Design and Applications Ninth International Conference on*, 517-522. <http://dx.doi.org/10.1109/ISDA.2009.238>
- [4] Holden, N., & Freitas, A. A. 2008. A hybrid PSO/ACO algorithm for discovering classification rules in data mining. *Journal of Artificial Evolution and Applications*, 2008. 1-11.
- [5] Utgoff P. E. 1996. Decision Tree Induction Based on Efficient Tree Restructuring. *Machine Learning*. Vol. 29. - págs. 5-44.

# MODELO DE PROCESO DE CONCEPTUALIZACION DE REQUISITOS

Dr. Alejandro Armando Hossian

Tesis Doctoral en Ciencias Informáticas. Facultad de Informática. Universidad Nacional de La Plata

Directores de Tesis: Ramón García-Martínez (UNLP-UNLa) y Dr. Oscar Dieste (UPM)

Fecha de Defensa: 18 de Octubre de 2012

**Resumen.** El proceso de captura de requisitos constituye un proceso con connotaciones sociales relacionadas con diferentes personas (stakeholders), una circunstancia que hace que se presenten ciertos problemas cuando se lleva adelante la conceptualización de requisitos. En esta tesis se propone un Proceso de Conceptualización de Requisitos que se estructura en dos fases: (a) Análisis Orientado a al Problema: cuyo objetivo es comprender el problema dado por el usuario en el dominio en el que este se lleva a cabo, y (b) Análisis de Orientado al Producto: cuyo objetivo es obtener las funcionalidades que el usuario espera del producto de software a desarrollar, teniendo en cuenta la relación de estas con la realidad expresada por el usuario en su discurso. Se proponen seis técnicas que articulan cada una de las tareas que componen las fases de proceso propuesto.

## 1. Introducción

En estadios tempranos de la Ingeniería de Requerimientos, Alford [1977], Yeh y Zave [1980] y Davis [1993] identificaron la necesidad de obtener una representación intermedia de la información obtenida – conceptualización de los requisitos –, facilitando de esta manera una captura adecuada del problema a resolver por parte del profesional de ingeniería de software antes de pasar a la construcción de los modelos conceptuales, habida cuenta de que una correcta construcción de estos modelos es fundamental para el éxito en el desarrollo del proyecto software, mientras que su incorrección puede perjudicar seriamente a las organizaciones implicadas [Chen, 1990].

Asimismo cabe señalar, que la escasa existencia de trabajos referidos a la elaboración de representaciones intermedias de los caudales de información obtenidos a lo largo de la actividad de educación, tendientes a una búsqueda de reducción de la complejidad de la realidad y su problemática expresada por el cliente y/o usuario en su discurso, agravan aún más este problema.

En este sentido y en lo que se refiere a la gestión de requisitos en el campo de los sistemas de información, se pueden citar algunos principios fundamentales de estructuración de la información – “Partición, Abstracción y Proyección” –, los cuáles proporcionan una estructura de conocimiento a fin de contribuir a una visión simplificada de la realidad y su problemática [Juristo, 1991]. Los elementos que suelen utilizarse para este análisis de problemas son los objetos, las funciones y los estados, pudiendo éstos describirse en múltiples niveles de detalle. Dado que hay tantas relaciones que pueden existir entre todos los elementos, se hace necesario disponer de una estructura de conocimiento (colección estructurada de conceptos y sus interrelaciones) que permita la captura estas relaciones.

A partir de la partición, es posible capturar la relación estructural “agregación/parte de” entre objetos, funciones o estados en el dominio del problema.

A partir de la abstracción, es posible capturar las relaciones estructurales “general/específico” o “ejemplo de” entre objetos, funciones o estados en el dominio del problema.

A partir de la proyección, es posible capturar la relación estructural “visión de” entre objetos, funciones o estados en el dominio del problema.

Si bien estos principios ofrecen su aporte a los efectos de precisar un mejor entendimiento de sus requisitos, son de carácter muy general y de poco nivel de detalle.

De igual manera y en lo que respecta a la gestión del conocimiento dentro del campo de los sistemas basados en conocimientos (SBC), se puede citar una técnica de representación intermedia como el “Análisis de Protocolos” [Newell & Simon, 1972]. Este método es de gran utilidad a los fines de obtener heurísticas que el experto utiliza en la solución de problemas, pero que le resulta difícil explicar [Gómez et al., 1997].

En síntesis, esta técnica consiste en grabar en un protocolo el comportamiento del experto mientras este trabaja en la solución del problema. Luego ese protocolo se transcribe y se analiza para, finalmente, interpretarlo y convertirlo en un conjunto de razonamientos que convergen a la solución del problema. La reconstrucción de esta solución permite modelar los conocimientos del experto.

La forma más clásica de representar este conocimiento consiste en codificar el mismo en la forma de reglas de producción, las cuáles presentan una parte izquierda [PI] y una parte derecha [PD] (Si..[PI].. Entonces..[PD]..).

Cabe destacar, que si bien esta técnica permite poner en evidencia carencias y fallos en el documento de educción de conocimientos, también es cierto que determinados procesos no son reportados por el experto y que no todos los conocimientos son fáciles de representar en forma de reglas [García-Martínez y Britos, 2004].

Se ha propuesto como objetivo general de la tesis definir un marco metodológico que incorpore una actividad de conceptualización tendiente a mejorar la comprensión y captura de requisitos de usuario en la fase de análisis de la Ingeniería de Requerimientos del Software. Esta actividad de conceptualización buscará plasmar en un esquema de representación integrado la realidad descrita por el cliente y/o usuario en su universo de discurso, así como también la problemática embebida en ella que se intenta resolver mediante una solución software.

La tesis se enfoca a plantear un proceso de conceptualización que funcione a modo de puente vinculando las actividades propias de la educción y el modelado de requisitos en la Ingeniería del Software. Con base en problemas de comunicación e interpretación entre clientes y/o usuarios y desarrolladores, surge la necesidad de una representación intermedia que facilite la consistencia del proceso de convergencia de los requisitos planteados por el usuario hacia los respectivos modelos conceptuales.

La tesis ha buscado formular contribuciones sobre: [a] actividades de conceptualización de requisitos de usuario capaces de proporcionar un mecanismo de análisis del discurso que permita al desarrollador relevar aquellos aspectos significativos de la realidad y su problemática, [b] mecanismos de derivación de esquemas de representación que faciliten la comprensión del problema de usuario y su asociación con aspectos de la solución software a implementar, y [c] estrategias que fortalezcan los canales de comunicación entre clientes, usuarios y desarrolladores a los efectos de optimizar la validez de las representaciones elaboradas para modelar la realidad y su problemática en función de su grado de aproximación al entorno del usuario.

## **1. Delimitación del Problema**

La insuficiencia en el tratamiento de la complejidad contenida en el discurso del usuario en la literatura correspondiente, y la necesidad de cubrirla, ha sido resaltada por diversos autores: [Stucliffe ,1992; Yu et al., 1994; Wieringa, 1995; Holtzblatt et al, 1995; Beringer, 1996; Faulk, 1997; Jalote, 1997; Chatzoglou, 1999; Juristo et al, 2000; Davis et al, 2003] entre otros. Estos autores mencionan las dificultades para la construcción de los modelos conceptuales a partir de la información recogida en el proceso de educción y plasmada en el discurso de usuario. Asimismo cabe resaltar, que dichas dificultades dotan al proceso de Análisis de un grado tal de inmadurez que hace que sea difícil llevar a cabo en forma efectiva esta actividad, al mismo tiempo que dificulta la adopción de este enfoque en las organizaciones [Moreno, 1999].

Por consiguiente y en virtud de todo lo expuesto, el problema abierto que se aborda en este trabajo de Tesis, consiste en la existencia de una “brecha conceptual”, lo que se denomina un “gap”



[Stucliffe, 1992; Davis, 1993; Robertson, 1999] en la transición de un proceso (Educción de Requisitos) a otro proceso (Modelado Conceptual). La Figura 1 ilustra la idea expuesta:

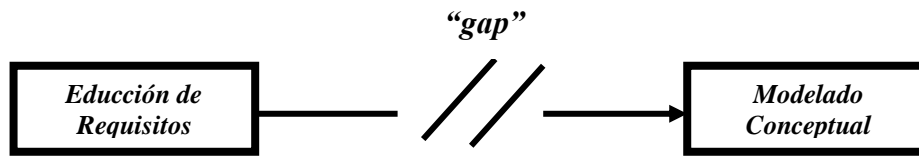


Figura 2. Representación del “gap” entre los procesos de Educción de Requisitos y Modelado Conceptual

A causa de lo expuesto, se manifiesta la necesidad de conceptualizar los requisitos manifestados por el usuario en su discurso antes de pasar a la construcción de los modelos conceptuales, con el objeto de reducir la complejidad mencionada y favorecer la comprensión del problema planteado por el usuario, contribuyendo así a la obtención de Modelos Conceptuales de mayor calidad [van der Vos, 1995; Chen, 1990].

Asimismo, es importante señalar la muy escasa cantidad de trabajos existentes en la literatura científica sobre la elaboración de representaciones intermedias de los caudales de información obtenidos por el IR en el proceso de educación. En otras palabras, trabajos que estén orientados a la búsqueda de reducción de la complejidad de la realidad y su problemática expresada por el usuario en su discurso.

El Modelo de Proceso de Conceptualización de Requisitos que se introduce, pretende realizar un aporte en este sentido. Con el soporte conceptual de tópicos pertenecientes a otras disciplinas, tales como el Análisis de Protocolo proveniente del campo de la Ingeniería del Conocimiento; y las Técnicas Cognitivas propias del campo de las Teoría Educativas, se analiza en detalle el Discurso del Usuario a los fines de estructurar y caracterizar el cuerpo de información presente en el mismo. De lo expuesto precedentemente surgen las siguientes preguntas de investigación:

- Pregunta 1: ¿Se puede plantear distinciones en el discurso del usuario que permitan diferenciar subdominios de análisis que minimicen la brecha conceptual entre la educación de requisitos y el modelado conceptual? En caso afirmativo: ¿Cuales?
- Pregunta 2: ¿De existir tales distinciones, se puede plantear un proceso que permita transformar el discurso del usuario en un conjunto de formalismos que lo sistematicen y lo documenten? De ser posible: ¿Cuáles son las fases de dicho proceso, las tareas vinculadas a cada fase y las técnicas asociadas a cada tarea?

## 2. Solución Propuesta

La solución que se propone en este trabajo de Tesis consiste en la inserción de una actividad de “Conceptualización de Requisitos”, la cuál tiene como finalidad actuar a modo de puente o enlace (“link”) entre las actividades de educación de requisitos y modelado conceptual, facilitando de esta manera la comprensión del problema manifestado por el usuario y, en consecuencia, la obtención de Modelos Conceptuales de mayor calidad [Chen ,1990; van der Vos, 1995; Chatzoglou, 1999; Juristo et al, 2000; Davis, 2003].

La ilustración de esta idea se puede visualizar en la Figura 2, en la cual se puede observar la ausencia del “gap”, el cual se sustituye por la actividad de “*Conceptualización de Requisitos*”.



Figura 2. Inserción de la actividad de “**Conceptualización de Requisitos**” entre las actividades de Educación de Requisitos y Modelado Conceptual

La idea de “conceptualizar” los requisitos de usuario por medio la actividad de “Conceptualización de Requisitos” antes de pasar a la confección de los modelos conceptuales, intenta cubrir esta “brecha conceptual” o “gap” [Stucliffe ,1992; Davis ,1993; Robertson, 1999] existente en la transición de un proceso (Educación de Requisitos) a otro proceso (Modelado Conceptual), actuando a modo de puente o enlace (“link”) entre dichos procesos. De este modo, se busca establecer una adecuada conexión entre los mismos a partir de la inserción de la actividad de Conceptualización de Requisitos. Esta tesis propone el Proceso de Conceptualización de Requisitos como instrumentación de dicha actividad.

A partir de la implementación de esta actividad de conceptualización de requisitos es posible la consecución de un conjunto de representaciones gráficas denominadas Representaciones Intermedias de los Requisitos de Usuario (RIRU), a partir de las cuales es posible “caracterizar” la información contenida en el discurso del usuario (por lo general en formato de “lenguaje natural” y es así como se la supone presentada en este trabajo), a los efectos de que sea mas sencillo su procesamiento para la construcción de los modelos conceptuales. Estas representaciones intermedias estarán conformadas, fundamentalmente, por un conjunto de representaciones gráficas: los Escenarios de Usuario Refinados (EUR), los cuales enlazados en forma adecuada a través del Mapa Unificado de Escenarios de Usuario Refinados (MUEUR) permiten caracterizar el discurso del usuario en una forma alternativa al lenguaje natural clásico.

La actividad de conceptualización de requisitos se lleva a cabo por medio de un proceso dual que se denomina Proceso de Conceptualización de Requisitos, el cual está estructurado en dos fases, a saber:

1. Una primera fase de Análisis Orientado al Problema, cuyo objetivo se focaliza en la comprensión del problema planteado por el usuario en el dominio en el cual este tiene lugar.
2. Una segunda fase de Análisis Orientado al Producto, cuyo objetivo consiste en la obtención de las funcionalidades que el usuario pretende obtener del producto software a desarrollar, teniendo en cuenta la vinculación de estas funcionalidades con la realidad manifestada por el usuario en su discurso.

Este proceso toma como punto de partida el Discurso de Usuario en Lenguaje Natural (DULN) y proporciona como salida el conjunto de Representaciones Intermedias de los Requisitos de Usuario (RIRU).

El soporte principal del Proceso de Conceptualización de Requisitos está compuesto por sus dos Fases, donde cada una de ellas está conformada por tres Tareas, y un conjunto de productos que pueden actuar como elemento de entrada y/o de salida de una determinada tarea. En otros términos, cada tarea precisa de ciertos productos para su realización, los cuales se procesan para proporcionar los correspondientes productos de salida. En la figura 3 se ilustra el modo de funcionamiento del proceso de conceptualización en base a la interdependencia conceptual existente entre la Fases, las Tareas y los Productos. En tal sentido, dicha figura muestra el flujo que siguen estos productos abasteciendo a determinadas tareas para su realización y/o ser procesados para constituirse en salida de las mismas.

En Figura 3 se puede observar que en la primera fase de Análisis Orientado al Problema la primera tarea que se lleva a cabo es la de Segmentación del Discurso de Usuario (SDU), la cual necesita del

Discurso de Usuario (DU) como producto de entrada y proporciona como producto de salida los correspondientes Segmentos de Texto (ST). Estos ST constituyen a su vez el producto de entrada para la realización de la tarea de Análisis Cognitivo de los Segmentos de Texto (ACST), la cual arroja como producto de salida los Tipos de Conocimiento (TC) embebidos en estos segmentos. A su vez, estos Tipos de Conocimiento (TC) junto con los Segmentos de Texto (ST) conforman el conjunto de productos de entrada necesarios para llevar a cabo la tarea de Construcción del Espacio Problema en Escenarios de Usuario (CEPEU), a partir de la cual se obtiene como producto de salida los correspondientes Espacio Problema en Escenarios de Usuario (EPEU).

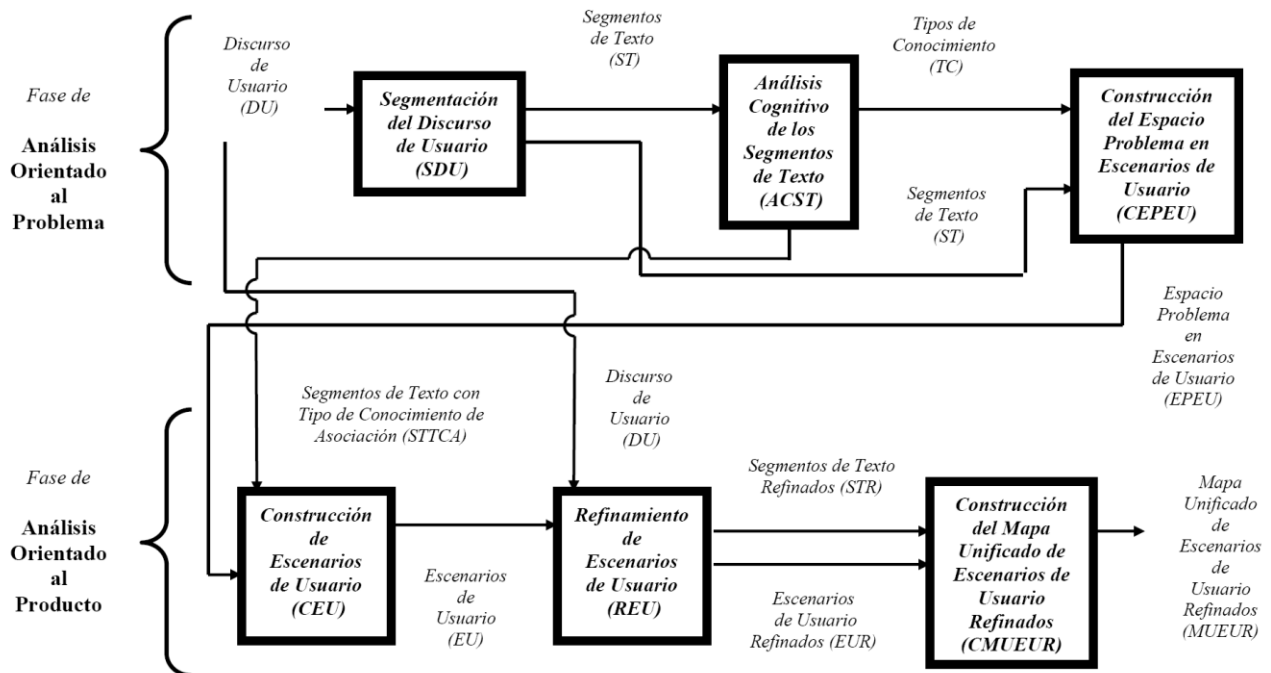


Figura 3. Interdependencia conceptual entre las Fases, Tareas y Productos

Luego se comienza con el desarrollo de la fase de Análisis Orientado al Producto donde la primera tarea que se realiza es la de Construcción de Escenarios de Usuario (CEU), la cual necesita como productos de entrada a los Segmentos de Texto con Tipo de Conocimiento de Asociación (STTCA) y los Espacio Problema en Escenarios de Usuario (EPEU), los cuales se procesan en el desarrollo de esta tarea y se obtienen los respectivos Escenarios de Usuario (EU). Estos Escenarios de Usuario (EU) junto con el Discurso de Usuario (DU) constituyen los productos de entrada para la realización de la tarea de Refinamiento de Escenarios de Usuario (REU), la cual proporciona como producto de salida los correspondientes Escenarios de Usuario Refinados (EUR). Finalmente, con los Escenarios de Usuario Refinados (EUR) y los Segmentos de Texto Refinados (STR) como productos de entrada, se realiza la tarea de Construcción del Mapa Unificado de Escenarios de Usuario Refinados (CMUEUR) y se obtiene el Mapa Unificado de Escenarios de Usuario Refinados (MUEUR).

### 3. Aportaciones de la Tesis

En esta tesis se ha corroborado que se pueden plantear distinciones en el discurso del usuario que permitan diferenciar sub-dominios de análisis que minimicen la brecha conceptual entre la educación de requisitos y el modelado conceptual. Estos sub-dominios son los relacionados con: [a] la descripción de los componentes de la realidad del ambiente de trabajo del usuario que su discurso pone en evidencia; y [b] los aspectos relacionados con las funcionalidades que el usuario espera que el artefacto software posea.

En este contexto, esta tesis ha propuesto:

- Un modelo de proceso de conceptualización de requisitos que se desarrolla en dos fases: una de Análisis Orientado al Problema y la otra de Análisis Orientado al Producto.
- Para la Fase de Análisis Orientado al Problema, se han propuesto las siguientes tareas: [i] *Segmentación del Discurso de Usuario*, la cual necesita del *Discurso de Usuario* como producto de entrada y proporciona como producto de salida los correspondientes *Segmentos de Texto*, [ii] *Análisis Cognitivo de los Segmentos de Texto*, que toma como producto de entrada a los *Segmentos de Texto* y proporciona como producto de salida los *Tipos de Conocimiento* embebidos en estos segmentos; y [iii] *Construcción del Espacio Problema en Escenarios de Usuario*, que tiene como insumos a los *Tipos de Conocimiento* y a los *Segmentos de Texto*, y proporciona como producto de salida los correspondientes *Diagramas de Espacio Problema en Escenarios de Usuario*.
- Para la Fase de Análisis Orientado al Producto, se han propuesto las siguientes tareas: [iv] *Construcción de Escenarios de Usuario*, la cual necesita como productos de entrada a los *Segmentos de Texto con Tipo de Conocimiento de Asociación* y los *Espacio Problema en Escenarios de Usuario*, los cuales se procesan en el desarrollo de esta tarea y se obtienen los respectivos *Escenarios de Usuario (EU)*; [v] *Refinamiento de Escenarios de Usuario*, que tiene como insumos a los *Escenarios de Usuario* y al *Discurso de Usuario*, que proporciona como producto de salida los correspondientes *Escenarios de Usuario Refinados* y [vi] *Construcción del Mapa Unificado de Escenarios de Usuario Refinados* que tiene como insumos los *Escenarios de Usuario Refinados* y los *Segmentos de Texto Refinados (STR)* para producir el *Mapa Unificado de Escenarios de Usuario Refinados*.
- Para la Fase de Análisis Orientado al Problema, se han desarrollado las técnicas: Técnica de Segmentación del Discurso de Usuario, Técnicas Cognitivas de Identificación de Conocimientos Factuales, Procedurales, Contextuales y de Asociación y la Técnica de Construcción del Diagrama de Espacio Problema de Escenarios de Usuario.
- Para la Fase de Análisis Orientado al Producto, se han desarrollado las técnicas: Técnica de Construcción del Diagrama de Escenarios de Usuario, Técnica de Refinamiento del Diagrama de Escenarios de Usuario y Técnica de Construcción del Diagrama del Mapa Unificado de Escenarios de Usuario Refinado.

La propuesta de modelo de proceso de conceptualización de requisitos, las tareas y técnicas asociadas han sido validadas en dos dominios de conocimiento con características bien diferenciadas: el primero sobre un Sistema de Abastecimiento de Combustible de Aeronaves en el Contexto de las Operaciones Aeroportuarias, el cual se circunscribe en el dominio de los sistemas de información clásicos; y el segundo correspondiente a un Sistema de Operaciones Bancarias por Cajero Automático, el cual se circunscribe dentro de los sistemas de información transaccional.

#### 4. Futuras Líneas de Investigación

Durante el desarrollo de este proyecto de tesis han surgido cuestiones que si bien no son centrales al tema abordado en la misma, constituyen temas concomitantes que (en opinión del tesista) darían lugar a las siguientes líneas de investigación futuras:

- En esta tesis se han utilizado técnicas de Análisis Cognitivo y técnicas de Ingeniería de Conocimiento. Estos conocimientos no forman parte de los estándares fijados para la disciplina con lo que cabe preguntarse:
  - ♦ ¿Qué conocimiento debería tener el ingeniero de requisitos para poder realizar conceptualización de requisitos?
  - ♦ ¿Debería buscarse una formación interdisciplinaria?

¿Qué disciplinas deberían estar involucradas?



- El proceso de conceptualización de requisitos suele ser subestimado al momento de asignar tiempos y recursos en la correspondiente planificación y presupuestación del proyecto. Poder concebir un proceso con fases y tareas como las propuestas en esta tesis acerca posiciones respecto de disponer de objetos conceptuales a los que asignarle recursos y prever su desarrollo en una línea de tiempo. En este contexto, se plantea el interés de trabajar en el desarrollo de herramientas que permitan estimar el esfuerzo que llevaría realizar este proceso de conceptualización.
- Si bien el proceso propuesto en la tesis aporta sistematicidad al proceso de conceptualización de requisitos y el mismo ha sido validado en dominios representativos, quedan como temas de trabajo abiertos:
  - ♦ La validación empírica más amplia del proceso de conceptualización de requisitos mediante la técnica de muestras apareadas basadas en grupos experimental y de control.
  - ♦ La validación empírica de las técnicas propuestas en un conjunto vasto y representativo de dominios de aplicación (sistemas de tiempo real entre otros).
- A los efectos de poder obtener modelos conceptuales de “alta calidad”, entre los cuales se pueden citar: diagramas de casos de uso, diagramas de clases, diagramas de objetos, diagramas de interacción (secuencia y colaboración) y diagramas de estado, entre otros; sería aconsejable considerar una línea de investigación orientada a estudiar la derivación estos modelos a partir de los diferentes productos que proporciona el proceso propuesto en esta tesis.

## 5. Producción Científica Derivada de Resultados Parciales de la Tesis

Durante el desarrollo de esta tesis se han comunicado resultados parciales a través de diversas publicaciones que a continuación se detallan:

### Capítulos de Libros:

Hossian, A., Dieste, O., García-Martínez, R. 2011. *A Process for Requirements Conceptualization*. En *Software Engineering, Methods, Modeling and Teaching*. Pág. 101-115. Sello Editorial Universidad de Medellín. ISBN 978-958-8692-32-6.

### Revistas con Referato:

Hossian, A.; Garcia-Martínez, R. 2013. *Proposal of Tasks and Techniques for a Requirements Conceptualization*. *International Journal of Computer and Communication Engineering*, 2(1): 36-40. ISSN 2010-3743.

### Congresos Internacionales:

Hossian, A., García-Martínez, R. 2012. *Phases, Activities, and Techniques for a Requirements Conceptualization Process*. *Proceedings 24th International Conference on Software Engineering and Knowledge Engineering*. Pág. 25-32. ISBN 978-1-891706-31-8.

Hossian, A., Dieste, O., García-Martínez, R. 2012. *Proposal of Tasks and Techniques for a Requirements Conceptualization*. *Proceedings 2012 International Conference on Software and Computer Engineering*.

### Congresos Regionales:

Hossian, A., Garcia-Martinez, R. 2011. *Problem-Oriented Analysis Phase within Process of Conceptualization of Requirements*. *Proceedings of II International Congress on Computer Science and Informatics (INFONOR-CHILE 2011)*. Pp. 95-103. ISBN 978-956-7701-03-2.

Hossian, A., Dieste, O., García-Martínez, R. 2012. *Conceptualización de Requerimientos: Propuesta de Proceso y Técnicas Asociadas*. Proceedings Latin American Congress on Requirements Engineering and Software Testing. Pág. 14-22. ISBN 978-958-46-0577-1.

Congresos Nacionales:

Hossian, A. Dieste, O., Garcia-Martinez, R. 2011. *Propuesta de Técnicas para un Proceso de Conceptualización de Requisitos*. Proceedings XVII Congreso Argentino de Ciencias de la Computación. Pág. 857-866. ISBN 978-950-34-0756-1.

## 6. Referencias

- Alford, M. 1977. *A Requirements Engineering Methodology for Real-Time Processing Requirements*. IEEE Transactions on Software Engineering, SE-3(1).
- Beringer, D. 1996. *The Goals of the Analysis Model*; Technical Report 96/216.
- Chatzoglou P., Soteriou A. 1999. *A DEA framework to assess the efficiency of the software requirements capture and analysis process*. Decision-Sciences. 30(2): 503-31.
- Chen, P. 1990. *Entity-relationship Approach to Data Modeling*. In System and Software Requirements Engineering, Thayer RH, Dorfman M (eds). IEEE. Computer Society Press.
- Chen, P. 1990. *Entity-relationship Approach to Data Modeling*. In System and Software Requirements Engineering, Thayer RH, Dorfman M (eds). IEEE. Computer Society Press.
- Davis, A. 1993. *Software Requirements: Objects, Functions and States*; Prentice-Hall International.
- Davis, A. 1993. *Software Requirements: Objects, Functions and States*; Prentice-Hall International.
- Davis, A. and Hickey, A. 2003. *Requirements Elicitation and Requirements Elicitation Technique Selection: A Model of Two Knowledge-Intensive Software Development Processes*. Proceedings of the Thirty-Sixth Hawaii International Conference on System Sciences, Los Alamitos, California: IEEE Computer Society Press.
- Faulk, S. 1997. *Software Requirements: A Tutorial*; In Software Engineering, IEEE Computer Society Press, pp 82-101.
- García Martínez, R. y Britos, P. 2004. *Ingeniería de Sistemas Expertos*. Editorial Nueva Librería. ISBN 987-1104-15-4.
- Gómez, A., N. Juristo, C. Montes, J. Pazos, Ingeniería del Conocimiento, Ed. Centro de Estudios Ramón Areces, (1997).
- Holtzblatt, K., and H. Beyer. *Requirements Gathering: The Human Factor*. Communications of the ACM, 38, 5 (May 1995), pp. 31-32.
- Jalote, P. 1997. *An Integrated Approach to Software Engineering*; Springer-Verlag.
- Juristo, N. 1991. *Método de construcción del núcleo de una base de conocimientos a partir de un modelo de clasificación documental*. Tesis Doctoral, Universidad Politécnica de Madrid.
- Juristo, N., Moreno, A. 2000. *Introductory paper: Reflections on Conceptual Modeling*. Data and Knowledge Engineering, 33(2): 103-117.
- Moreno Sánchez Capuchino, A., 1999. *Método Formal de Modelización Conceptual para Sistemas Software*. Tesis Doctoral Universidad Politécnica de Madrid, Madrid, 1999.
- Newell, A., y Simon, H. 1972. *Human Problem Solving*. Englewood Cliffs, NJ: Prentice Hall.
- Robertson, S., Robertson, J. 1999. *Mastering the Requirements Process*. Addison-Wesley.

- Sutcliffe, A., Maiden, N. 1992. *Analysing the Novice Analyst: Cognitive Models in Software Engineering*; International Journal of Man-Machine Studies, 36(5).
- Van der Vos, B., Gulla, J., Van de Riet, R., 1995. *Verification of Conceptual Models based in Linguistic Knowledge*. NLDB 1995.
- Wieringa, R. 1995. *Requirements Engineering: Frameworks for Understanding*; John Wiley.
- Yeh, R., Zave, P. 1980. *Specifying Software Requirements*, Proc. of the IEEE, 68(9): 1077-1085.
- Yu, E., Mylopoulos, J. 1994. *Understanding "Why" in Software Process Modelling, Analysis and Design*; Proceedings of the 16th International Conference on Software Engineering.

# Engineering Accessible Web Applications. An Aspect-Oriented Approach

**Author:** *Adriana E. Martín*

<sup>1</sup>Department of Exact Sciences, Caleta Olivia, University of Patagonia Austral (UNPA-UACO), Santa Cruz, Argentina

<sup>2</sup>GIISCo Research Group, Departamento de Ciencias de la Computación, Universidad Nacional del Comahue, Buenos Aires 1400, Neuquén, Argentina  
e-mail: [adrianaelba.martin@gmail.com](mailto:adrianaelba.martin@gmail.com)

**Supervisor:** *Dr. Alejandra Cechich*

GIISCo Research Group, Departamento de Ciencias de la Computación, Universidad Nacional del Comahue, Buenos Aires 1400, Neuquén, Argentina  
e-mail: [acechich@uncoma.edu.ar](mailto:acechich@uncoma.edu.ar)

**Co-Supervisor:** *Dr. Gustavo Rossi*

LIFIA, Facultad de Informática, Universidad Nacional de La Plata and Conicet, La Plata, Argentina  
e-mail: [gustavo@sol.info.unlp.edu.ar](mailto:gustavo@sol.info.unlp.edu.ar)

**Universidad Nacional de La Plata  
Facultad de Informática**

Thesis submitted in fulfillment of the requirements for the  
Doctor degree in Informatics Science  
Thesis defended on June 5, 2012

## ABSTRACT

*Every day more and more users with different abilities and/or temporally or permanent disabilities are accessing the Web, and many of them have difficulties in reaching the desired information. However, the development of this kind of software is complicated for several reasons. Though some of them are technological, the majority are related with the need to compose different and, many times, unrelated design concerns which may be functional as in the case of most of the application's requirements, or non-functional such as Accessibility. Even though, there is a huge number of tools and proposals to help developers assess Accessibility of Web applications, looking from the designer perspective, there is no such a similar situation. In this thesis, we present a novel approach to conceive, design and develop Accessible Web applications using concepts from Aspect-Oriented. In order to accomplish our goal, we provide some modeling techniques that we explicitly developed for handling the non-functional, generic and crosscutting characteristics of Accessibility. Specifically, we have enriched the UID technique with integration points to record Accessibility concerns that will be taken into account when designing the user interface. Then, by instantiating the SIG template with association tables, we work on an abstract interface model with Accessibility softgoals to obtain a concrete and accessible interface model for the Web application being developed. We explain deeply our ideas and point out the advantages of a clear separation of concerns throughout the development life-cycle. Thus, our proposal is based on recognized design techniques, which we embedded in a software tool to facilitate the transfer of the approach to the industry.*

## 1. INTRODUCTION

Since 1999, when the W3C<sup>1</sup> introduced the “Web Content Accessibility Guidelines 1.0” (WCAG 1.0) [18] as a set of guiding principles, the fact that Accessibility is a main topic in Web design upon which the success of a Web application depends, has become a landmark statement. However, developing accessible Web applications is usually hard for several reasons.

Firstly, there is a significant knowledge gap between developers and Accessibility specialists. Most developers do not have the necessary skills or training in designing and coding for Accessibility, and most Accessibility specialists have, in turn, limited developing practice. Thus, although there

---

<sup>1</sup> The World Wide Web Consortium at <http://www.w3.org/>



are many available tools and published sources of information on Web Application Accessibility, existing Web Accessibility guidelines and principles (and therefore, experts on these guidelines) do not address additional design issues that may typically arise when developing complex Web applications. To make matters worse, there is little evidence of design approaches dealing with Accessibility from the beginning of the design process. In most cases, Accessibility is regarded as a programming issue or even dealt with when the Web application is already fully developed and, consequently, the process of making this application accessible involves significant redesign and recoding, which might be out of the scope of the project and/or hardly affordable. As we will show next, the main problem with Accessibility is that it is a non-functional software concern, which affects (“crosscuts”) other application concerns. Moreover, Accessibility is a generic concern that may comprise dozens of specialized concerns and, therefore, many requirements associated with these. For example, at the application-level, Accessibility can be specialized according to the kind of Accessibility support given to the user, where specific requirements related to the user’s layout and the user’s technology supports are considered. As another example, at the meta-level, Accessibility can be specialized according to meta-features like compliance design and content order concerns. Finally, and as an example of the model-level, Accessibility can also comprise different concerns according to the methodological phase for the development of the Web application, where the Accessibility efforts are focalized.

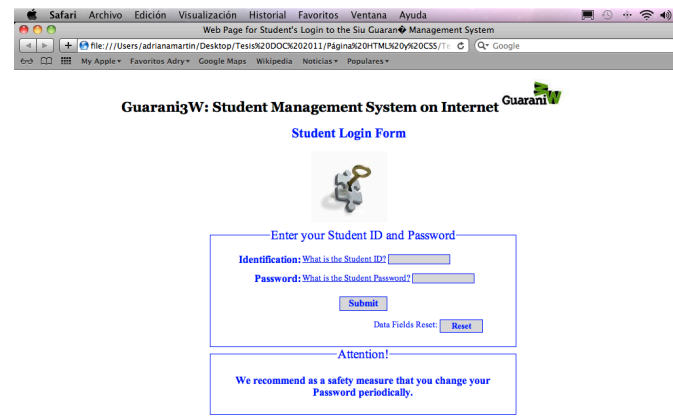
In this work we introduce our design approach, which proposes to include Accessibility concerns systematically within a methodology for Web application development. Firstly, to find out how Accessibility concerns should be introduced in the development life cycle, we analyzed how mature Model-Driven Web Engineering (WE) methods, such as UWE [7], OOHDM [14] or OOWS [6], face this cycle. We realized that all of them comprise several activities to focus on some specific design concerns; however, since OOHDM fulfill many of our expectations, we decided to join our modeling approach to this particular WE method. Secondly, since designing accessible Web applications involves the analysis of different interests, we proposed to use Aspect-Oriented Software Development (AOSD<sup>2</sup>) design principles to support the construction of accessible user interfaces. The fact that we choose Aspect-Oriented to develop our proposal ensures handling naturally the non-functional, generic and “crosscutting”<sup>3</sup> characteristics of the Accessibility concern. As a motivating example and to introduce properly the ideas behind our modeling approach, let us suppose a typical login Web page whose purpose is aiming a student’s identification at his/her university system, such as the SIU Guarani student registration system that is used by a number of Argentine universities. Figure 1 shows the page for the student’s login that provides a user interface composed of HyperText Markup Language (HTML) elements, such as labels and text fields. To help to an accessible interaction experience these HTML elements must fulfill some Accessibility requirements, which crosscut the same software artifact (the Web page for student’s login). For example, at the presentation level an HTML label element is a basic layout Accessibility requirement for many other HTML elements. Since a Web page for student’s login requires at least two text field elements (for student’s ID and password respectively), the presence and positioning of their respective label elements must be tested. So, to propitiate an accessible interaction experience on behalf of the student, these layout requirements must “crosscut” the same software artifact (the Web page) more than once, accordingly to the number of text field elements included in the presentation. Clearly this kind of behavior perfectly fits the “scattering” and “tangling” problems<sup>4</sup>, which motivate the main AOSD principles.

---

<sup>2</sup> Aspect-Oriented Software Development (AOSD) focuses on the identification, specification and representation of “crosscutting” concerns and their modularization into separate functional units as well as their automated composition into a working system.

<sup>3</sup> “Crosscutting” is a term used for certain type of functionality whose behavior causes code spreading and intermixing through layer and tiers of an application which is affected in a loss of modularity in their classes. Quality requirements (such as Accessibility) are examples of this common functionality that is usually described as “crosscutting concerns” and should be centralized in one location in the code where possible.

<sup>4</sup> “Scattering” and “Tangling” symptoms are typical cases of “crosscutting concerns” and they often go together, even though they are different concepts. A concern is “scattered” over a class if it is spread out rather than localized while a concern is “tangled” when there is code pertaining to the two concerns intermixed in the same class (usually in a same method).



**Figure 1:** A Student's Login Web page example

Since these two Accessibility requirements (presence and positioning of the label elements), are “scattered” in the Web page with a pair of label-text field HTML elements, the Web page is “tangled” with these Accessibility requirements. It seems natural therefore to address Accessibility using the Aspect-Oriented Software Development (AOSD) approach and, it is not just a coincidence that during this work we refer to Accessibility as a “concern”. The term “concern” from the AOSD perspective describes accurately the Accessibility features related to its nature. By using the AOSD paradigm we can avoid typical problems of “crosscutting” concerns, such as those shown in the previous Web page example. Our proposal applies these concepts by treating Accessibility as a first-class concern in the context of the OOHDM [14] WE approach.

The main objective of this work is to define a WE approach (process and techniques) to conceive, design and develop accessible Web applications using Aspect-Oriented concepts, which enable to address Accessibility early from requirements and through design to implementation.

The rest of the work is structured as follows: in Section 2, we offer an overview of our proposal describing the conceptual tools and model we envisage to deal with Accessibility concerns within a WE approach; while in Section 3, we briefly introduce the comparison between related work, focusing on the main contribution of ours. Finally, in Section 4 we conclude, present some further work and some of the contributions related to this work.

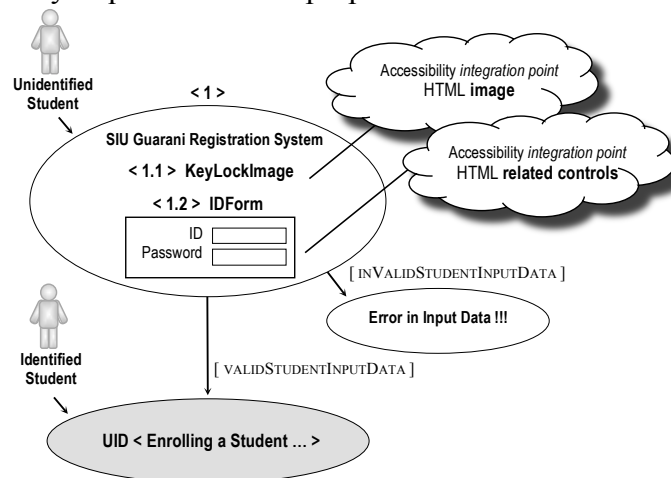
## 2. OUR APPROACH IN A NUTSHELL

In the spirit of modern Web Engineering approaches, we propose a model-driven development process in which the construction of a Web application consists of the specification of a set of conceptual models, each addressing a different concern (such as navigation or interface). We propose an iterative and incremental process, which uses, as input, a set of Web application's requirements as provided by any WE approach --e.g. a set of use cases, goals, etc. Our approach, proposes two conceptual tools working together to enable an early capture of the Accessibility concerns. These modeling techniques are the UID [17] with *integration points* and SIG [5] *template* for Accessibility, with which the interaction between OOHDM models links and reinforces Accessibility needs. Following, in Sections 2.1 and 2.2, we introduce these conceptual tools and then, in Section 2.3, we put all the pieces together to give a brief overview of our Aspect-Oriented approach for accessible design.

### 2.1 Accessibility through UIDs integration points

A User Interaction Diagram (UID) [17] is a diagrammatic modeling technique focusing exclusively on the information exchange between the application and the user. UIDs can be used to enrich the use cases models but they are also key graphical tools for linking requirements at later stages of a WE development process to obtain conceptual, navigational and user interface diagrams. With the traditional perspective given by techniques like [5] in mind, we introduce the concept of UIDs's *integration points* to model the Accessibility concerns of a user-system interaction. Particularly, we define two kinds of UIDs integration points as follows: (i) User-UID Interaction (U-UI) *integration*

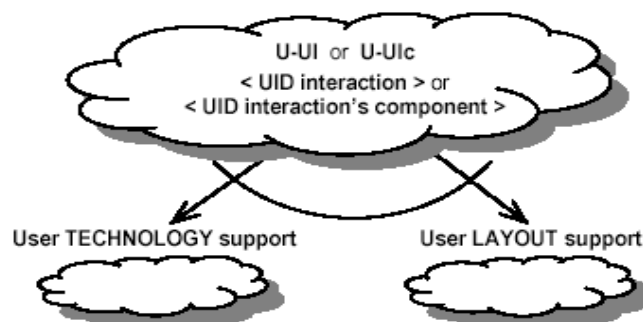
point to propitiate an accessible communication and information exchange between the user and a particular interaction of a UID interaction diagram; and (ii) User-UID Interaction's component (U-UIc) *integration point* to propitiate an accessible communication and information exchange between the user and a particular UID interaction's component of an UID interaction. These *integration points* with different granularity provide two alternatives for evaluating Accessibility during the interaction between the user and the system. For example, Figure 2 shows the resultant UID, corresponding to a use case "Login a student given the student's ID and password" (introduced in Section 1 by Figure 1), by applying our integration points technique. Notice that all the students (including those with disabilities) will need to interact with this online login Web page. As we can see in the example shown in Figure 2, we define two integration points at UID interaction <1> representing the student's login user-system interaction to consider, from the beginning, the Accessibility requirements that propitiate the access for all the students.



**Figure 2:** UID with Accessibility integration points: Login a Student given the Student's ID and Password

Basically, the UID with the *integration points* notation prescribes the inclusion of a cloud for every UID interaction or UID interaction's component, which Accessibility is essential to the user's task completeness. The first cloud establishes the <1.1> *integration point* to propitiate that the semantics of the KeyLockImage is correctly transmitted; while the second cloud establishes the <1.2> *integration point* to propitiate an accessible IDForm for user identification.

ACCESSIBILITY [ UID integration point ]



**Figure 3:** SIG Template for Accessibility

## 2.2 Applying the SIG template for Accessibility

After specifying the Accessibility *integration points* of the UIDs diagrams, we propose to develop a SIG diagram for WCAG [18] Accessibility requirements. Figure 3 shows our SIG *template* conceptual tool that we introduced taking into consideration proposals from the non-functional requirements [5] and user interface design literature [8] [14]. Figure 3 shows our SIG *template* where the Accessibility softgoal denoted with the nomenclature Accessibility [UID integration point] is the root of the tree. The kind of the UID *integration point* is highlighted into the root light cloud and related to a particular UID interaction or UID interaction's component number. From the

root node we identify two initial branches: (i) the user technology support, and (ii) the user layout support. The user technology support represents the Accessibility softgoal concerns helping to propitiate user's browsing and interaction by improving the Accessibility of user's current and earlier assistive devices and technologies (PDAs, telephones, screen readers, etc.); meanwhile, the user layout support represents the Accessibility softgoal concerns explicitly improving user's browsing and interaction focus on user's interface issues. The Accessibility softgoal concerns supply to their respective supports, prescribing on how to present and/or to logically organize the content we wish to convey to the user. They also warn about the Accessibility barriers as a consequence of an inappropriate choice of presentation and/or structural objects to user's interaction with the content. For example, returning to Figure 2, we establish the Accessibility softgoal for the interaction's components <1.1> KeyLockImage and <1.2> IDForm to propitiate accessible image and text input fields for all the students by defining two User-UID Interaction's components (U-UIc) integration points for the login process at UID interaction <1>. Finally, to instantiate the SIG *template* for specifying Accessibility concerns (shown in Figure 3) we work with the W3C-WAI WCAG recommendations (1.0 or 2.0) [18]. To facilitate the instantiation process of the SIG template we establish *association tables* for each of the following five groups of related HTML elements: (i) the HTML *control* elements group; (ii) the HTML *link* and *button* group, (iii) the HTML *text* and *non-text* group, (iv) the HTML *structural* elements group and, (v) the HTML *frame* and *style sheet* elements group, respectively. Basically, these *association tables* have the tasks of linking each abstract interface element present at a user interface model (ontology concepts from an Abstract Widget Ontology [14]) with their respective concrete HTML elements, and with the Accessibility concerns prescribed for those elements by the WCAG guidelines [18]. Before proceeding, we must also clarify that we have extended the original Abstract Widget Ontology [14], specifying that an abstract widget can be any of the following: (i) *SimpleActivator*, (ii) *ElementExhibitor*, (iii) *VariableCapture*, (iv) *LogicalStructuring* or (v) *ElementStyling*. We refer the reader to [9] for further details of our *association tables* and their relation with the extension proposed for the Abstract Widget Ontology. Returning to the explanation, the first step to obtain the *association tables* comes from a mapping between abstract interface widgets (ontology concepts from Abstract Widget Ontology) and concrete interface widgets (HTML elements). While the reason for HTML elements at the concrete interface model is completely clear, the purpose of the widget ontology is to provide an abstract interface vocabulary to represent the various types of functionality that can be played by interface widgets with respect to the activity carried out, or the information exchanged between the user and the application. Given these conceptual tools, the instantiation process of the SIG template is conducted as a refinement process over the SIG tree using the abstract interface model and the *association tables* as a reference.

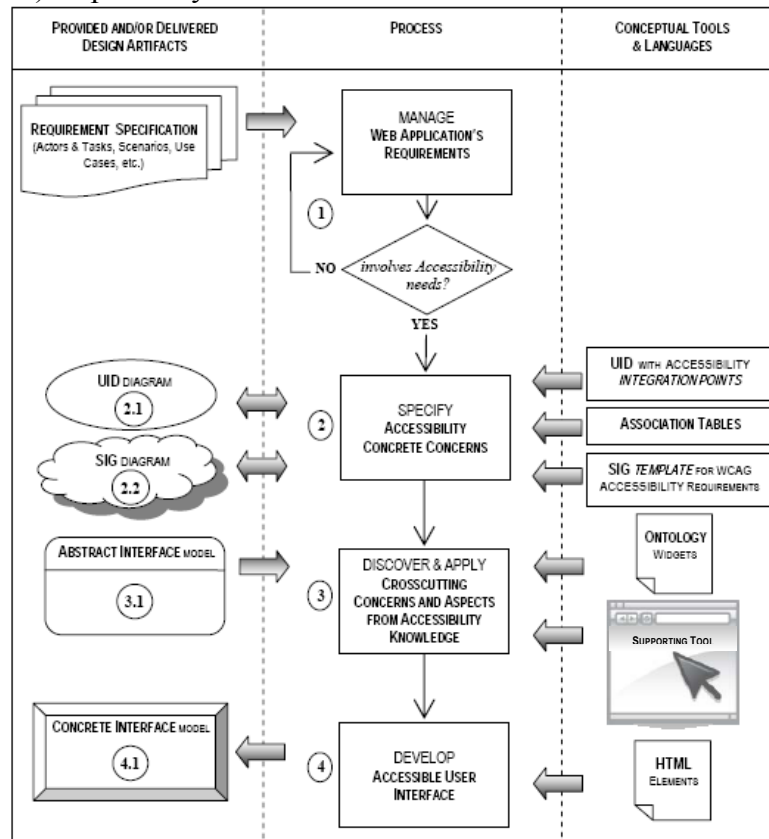
### 2.3 An Aspect-Oriented Approach for Accessible Design

The model we envisage to deal with Accessibility concerns within a Web engineering approach is illustrated in Figure 4, whose columns indicate: (i) the overall process with their main activities (in the middle), (ii) the conceptual tools and languages used (on the right) along with relations to the stage of the process where they are required, and (iii) the artifacts provided as input by the WE approach and/or delivered as output by our process (on the left). In order to ease reading, we need to recall here some previous explanations. In Figure 4, most arrows indicate an input or output, except for the UID and SIG diagrams as shown in Figure 4(2.1) and 4(2.2), where the arrows are input/output. This is because there are situations in which these artifacts could be developed once and then reused in different Web projects. For example, the Accessibility requirements of an image or a basic data entry form can be modeled once, and later reuse in new projects that require these interface elements. As highlighted in Figure 4(1), we propose a process that manages Web application requirements looking for those that involve Accessibility needs. This is because it is at the user's interface level where Accessibility barriers finally show, so we are particularly interested in discovering Accessibility requirements at the user interface design.

Then, as shown in Figure 4(2), we propose an early capture of Accessibility concrete concerns by developing two kinds of diagrams: the UID with Accessibility *integration points* and the Softgoal



Interdependency Graph (SIG) *template* for WCAG Accessibility requirements [18], as shown in Figure 4(2.1) and (2.2) respectively.



**Figure 4:** Overview of Our Approach

As we explained previously, we propose these conceptual tools basically to allow the representation of Accessibility requirements while executing a user's task.

As indicated in Figure 4(3), the Accessibility knowledge captured and organized by SIG diagrams at early stages aids designers making decisions through the abstract interface model, as shown in Figure 4(3.1), and then, as shown in Figure 4(4), toward its implementation through the concrete interface model with the desired Accessibility properties (conformance to the WCAG recommendations), as shown in Figure 4(4.1). The purpose here is to find out how WCAG Accessibility requirements "crosscut" interface widgets required for an IDForm. Since, and as we already explain in Section 1, applying the Accessibility concerns to be satisfied at the user interface causes typical crosscutting problems --i.e., "scattering" and "tangling" symptoms, it is clear that Aspect-Oriented is the natural approach to solve these crosscutting problems. The SIG diagrams not only provide Accessibility technology and layout support respectively for any of the HTML form components at the user interface, but also allow "aspects"<sup>5</sup> to be modeled and instantiated appropriately to avoid "scattering" and "tangling" symptoms. Finally, as highlighted in Figure 4(3), we propose a supporting tool to assist developers to discovering crosscutting concerns and applying aspects from the Accessibility knowledge capture at earlier stages. Basically, the type and the characteristics covered by the tool can be described as those normally provided by a Computer-Aided Software Engineering (CASE) tool.

### 3. MAIN CONTRIBUTIONS OF OUR WORK

In order to discuss and specify the contributions to the field of accessible design, we have developed an *evaluation framework* to carry out the comparison and evaluation between related work and ours. We highlight that we have studied and applied all related work to the same cases, to which we have applied ours.

<sup>5</sup> An aspect is a module that can localize the implementation of a crosscutting concern; the aspectual decomposition modularizes "scattering" problems --i.e. one concern in many modules, and "tangling problems" --i.e. one module, many concerns.

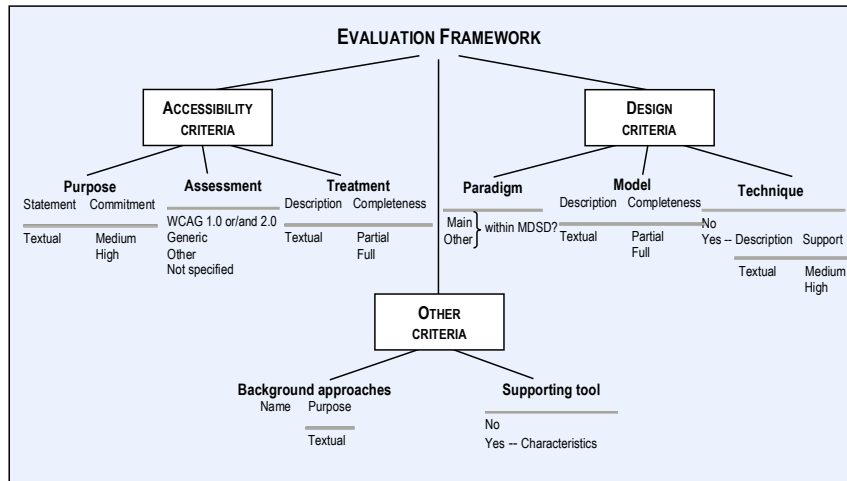


Figure 5: Evaluation Framework

As shown in Figure 5, our *evaluation framework* is divided into the following three main criteria: (i) Accessibility criteria, which assesses the degree of commitment with Accessibility by evaluating three topics: *purpose*, *assessment* and *treatment*, (ii) Design criteria, which evaluates design issues of the approaches under consideration by using three topics: *paradigm*, *model* and *techniques* and, (iii) Other criteria, which considers two additional topics: *background* and *supporting tool*. For brevity reasons, we refer the reader to [9] for further details of our *evaluation framework* and its instantiation for the state-of-the-art to the field of accessible design. Following, in Sections 3.1, we provide a synthesis of the comparison between related work and ours and then, in Section 3.2, we focus on the contributions of our approach.

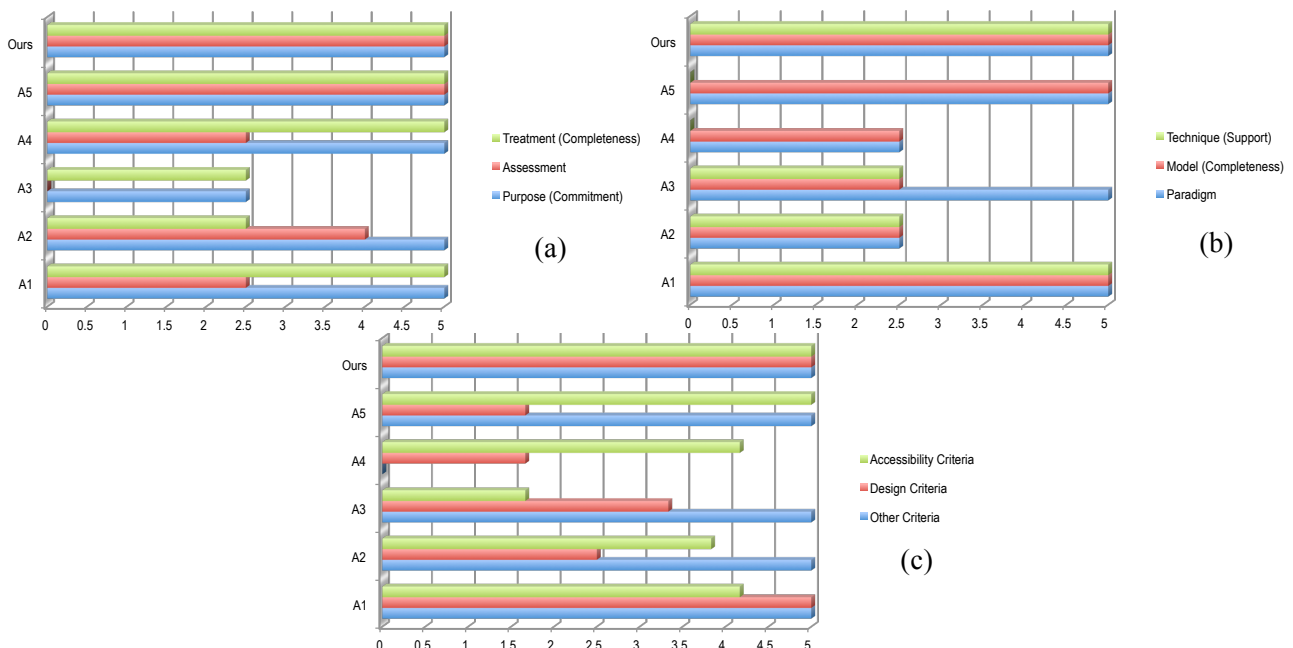


Figure 6: Scoring the Approaches for: (a) Accessibility Criteria, (b) Design Criteria and (c) Average by Approaches/Criteria

### 3.1 Summary of the Comparison and Evaluation Deliverables

For the purpose of facilitating the evaluation process and understanding the results, we identify each one of the approaches as follow: A1 [13], A2 [4], A3 [1][2][3], A4 [21], A5 [10][11] and Ours [9]. Again, for the sake of brevity, we refer the reader to [9], which provides a comprehensive description and implementation to cases of the approaches that comprise the state-of-the-art. Then, we score the topics related to the Accessibility and Design criteria from 0 to 5, as follows: (i) the scores “high” and “full” match to 5, while the scores “medium” and “partial” match to 2.5; (ii) at

the assessment topic, the option “WCAG 1.0 and 2.0” matches to 5, the option “WCAG 1.0” matches to 4, the option “generic” and “other” match to 2.5, and the option “not specified” matches to 0; and finally (iii) at the paradigm topic, the option “main within MDSD” matches to 5, while the option “other” matches to 2.5. Figures 6.2 and 6.3 show the scoring of the six approaches for the Accessibility and Design criteria, respectively. At this point, we must remember that the scoring process is conducted after a comprehensive study and application to cases of all the approaches; the complete comparison and evaluation. Figures 6(a) and 6(b) show briefly the result of the valuation of the six approaches by Accessibility and Design criteria, respectively. To complete this summary, Figure 6(c) shows the average of scores for the six approaches by Criteria. We should note that for the Other Criteria, we score only the *supporting tool* topic by simply matching the options “yes” and “no” to 5 and 0, respectively.

### 3.2 Focusing on Ours

As we already said, Ours [9] allows developers to produce accessible interfaces by moving from abstract to concrete architectural views using Aspect-Orientation. This is a main advantage, since allows developers to keep in mind a clear picture of how these architectural views relate each other during the development process, while preserving their own properties: (i) the abstract view ensures clean designs --i.e. free of crosscutting symptoms, which are separated and modeled as aspects for their modularization; while (ii) the concrete view provides the implementation of these designs, but as a consequence of the weaving process that takes place at the code level. Thus, Ours uses Aspect-Orientation to propose a smooth and open transition between models (abstract and concrete views), since this transition allows the independence of the way clean designs will be implemented into accessible code. At this point, we want to state the situation about users having alternatives when browsing; for example, as shown in Figure 1, the Web page offers the student two optional links to look for help for the login process. We highlighted that browsing these pages is optional and therefore, if the student follows these help links, his/her decision will produce a different navigation path. As we said before, we focus on the UI models because, undoubtedly, is at the UI where Accessibility barrier finally show, but notice that this is one of those cases in which navigational issues can affect Accessibility. This is the reason why, to improve the user’s experience when browsing to achieve the desired functionality, we have to consider the UI designs for each alternative in the navigation path we have defined as important for the task’s functionality. This means that if we provide the user with alternatives in the navigation path, they must be explored and modeled before properly, because they can be relevant to Accessibility and therefore to the success of the user’s task. This is an advantage of Ours, because although Ours is focused on UI models, also allows to explore navigational models to avoid unexplored optional browsing that can lead to user interfaces which were not considered initially. Basically, this is possible mainly because of two reasons. In first place, the UID is the conceptual tool used by OOHDM to state transformations between Web application requirements (use case model) and the conceptual, navigational and interface models. As shown in Figure 7, this is the same principle that Ours propitiates between Web applications requirements and accessible UI models. Ours uses two conceptual tools (the UID with *integration points* and *SIG template* for Accessibility), with which the interaction between OOHDM models links and reinforces Accessibility needs.

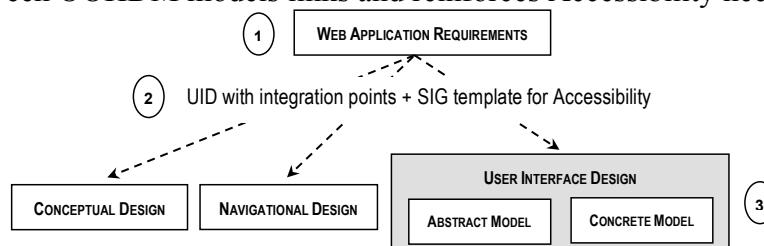


Figure 7: Ours within MDSD paradigm

In second place, since Ours is conceived within the MDSO paradigm, models are related to each other and as a consequence of an iterative and incremental development process. Thus, Ours allows: (i) going back from UI models to navigational models to look for alternatives in the navigation path, (ii) assessing the need and relevance of these alternatives to the functionality under develop, and (iii) going forward from navigational models to UI models to check the Accessibility of the UI related to these alternatives.

#### 4. CONCLUSIONS AND FUTURE WORK: A SYNTHESIS

In this work, we presented a novel WE approach to conceive, design and develop accessible Web applications using Aspect-Oriented concepts, which enabled us to address Accessibility early from requirements and through design to implementation. First, Aspect-Orientation capabilities constitute an important driver to efficiently capturing the orthogonal properties that are typical of the Accessibility's nature. Second, organizing these properties into a model-driven approach gives us better visibility of the components at different levels --i.e. from its conceptualization to its instantiation by particular Accessibility rules. In addition, we provided explicit analysis and design techniques aiming at facilitating the capture of early Accessibility concerns. However, we must take into account that the inclusion of new conceptual tools for treating Accessibility requires an extra effort for developers to get familiar with them. In this sense, we are currently incorporating our ideas into supporting tools to assist developers to design model-driven accessible Web applications. Since our proposal is strongly linked to the model-driven paradigm, we should note how this issue benefits/affects our proposal. It is a fact that applying "unified", model-driven approaches brings the benefit of having full documentation and automatic application generation at the expense of introducing some bureaucracy into the development process. Since our proposal suggests the early treatment of the Accessibility concerns through models, we may still be influenced by this reality and its disadvantages. Related to the project team and development environment, we believe it is important to highlight the following issues: (i) although our approach is completely documented and self-contained within a well-known WE approach, its application requires a prior knowledge of the WCAG guidelines [18]; (ii) although our approach helps to transfer Accessibility requirements, the engineering staff members should not be ruled by ad hoc practices or used to apply approaches without design and documentation as a standard discipline. These two issues demand changes in the development process that must be supported by the organizations. However, our proposal propitiates the reuse of design artifacts, because Accessibility concerns are quite independent from the Web application under development and Accessibility "aspects" could be developed once and be reused in different Web projects.

Finally, we should further validate our proposal and to do so, we are currently following three different but related paths: (i) generalizing the use of our approach within some of the best known WE approaches to provide accessibility support through Aspect-Orientation techniques (we are already working and have some promising results embedding AO-WAD into UWE and OOWS methods); (ii) improving the supporting tool's functionality to propitiate industry adoption; and (iii) analyzing deeply the impact of applying our proposal on quality attributes of the resulting system, such as reuse, extensibility and modularity, and the developing effort required when using the approach. We are currently carrying out some guided experiments in the area of Web-based systems for academic domains and the petroleum industry.

##### 4.1 Some of the Contributions Related to this Work

- (WWWJ 2010) World Wide Web: Internet and Web Information Systems Journal  
Engineering Accessible Web Applications. An Aspect-Oriented Approach. A. Martín, G. Rossi, A. Cechich, S. Gordillo. World Wide Web: Internet and Web Information Systems Journal ISBN: 978-1-59904-847-5 Vol 13(4); 419-440; DOI: 10.1007/s11280-010-0091-3
- (W4A 2011) World Wide Web 8th International Cross-Disciplinary Conference on Web Accessibility



Accessibility at Early Stages: Insights from the Designer Perspective. A. Martín, A. Cechich, G. Rossi. Proceedings of 8th International Cross-Disciplinary Conference on Web Accessibility, Hyderabad, India, 2011; ISBN: 978-1-4503-0476-4; ACM; DOI: 10.1145/1969289.1969302

- (ICSEA 2010) 5th International Conference on Software Engineering Advances Supporting an Aspect-Oriented Approach to Web Accessibility Design. A. Martín, R. Mazalú, A. Cechich. In: Proceedings of 5th International Conference on Software Engineering Advances, Nice, France, 2010; ISBN: 978-0-7695-4144-0; IEEE; 20-25; DOI: 10.1109/ICSEA.2010.10

## REFERENCES

- [1] Casteleyn, S., Fiala, Z., Houben, G-J., van der Sluijs, K. Considering Additional Adaptation Concerns in the Design of Web Applications. AH (2006) doi:10.1007/11768012\_28
- [2] Casteleyn, S., Van Woensel, W., Houben, G-J. A Semantics-based Aspect-Oriented Approach to Adaptation in Web Engineering. In HT (2007) doi.acm.org/10.1145/1286240.1286297
- [3] Casteleyn, S., Van Woensel, W., van der Sluijs, K., Houben, G.J.: Aspect-Oriented Adaptation Specification in Web Information Systems: a Semantics-based Approach. New Review of Hypermedia, Taylor and Francis 15(1), 39-91 (2009)
- [4] Centeno, V., Kloos, C., Gaedke, M., Nussbaumer, M. Web Composition with WCAG in Mind. W4A (2005) doi:10.1145/1061811.1061819
- [5] Chung, L., Supakkul, S. Representing FRs and NFRs: A Goal-oriented and Use Case Driven Approach. SERA (2004) doi:10.1007/11668855\_3
- [6] Fons, J., Pelechena, V., Pastor, O., Valderas, P., Torres, V. Applying the OOWS Model-Driven Approach for Developing Web Applications. The Internet Movie Database Case Study. In: Rossi, G., Pastor, O., Schwabe, D., Olsina, L. (eds.) Web Engineering: Modelling and Implementing Web Applications. pp. 65-108. Springer-Verlag, London (2008)
- [7] Koch, N., Knapp, A., Zhang, G., Baumeister, H. UML-Based Web Engineering: An Approach Based on Standards. In: Rossi, G., Pastor, O., Schwabe, D., Olsina, L. (eds.) Web Engineering: Modelling and Implementing Web Applications. pp. 157-191. Springer-Verlag, London (2008)
- [8] Larson, J.: Interactive Software: Tools for Building Interactive User Interfaces. Prentice Hall, NJ (1992)
- [9] Martín, A. Engineering Accessible Web Applications: An Aspect-Oriented Approach. PhD Thesis. <http://sedici.unlp.edu.ar/handle/10915/19685> (2012). SeDiCI repository (2012).
- [10] Moreno, L., Martínez, P., Ruiz, B. A MDD Approach for Modelling Web Accessibility. IWWOST (2008) doi:10.1.1.163.9478
- [11] Moreno, L. AWA: Methodological Framework in the Accessibility Domain for Web Application Development. PhD Thesis. [http://www.sigaccess.org/community/theses\\_repository/phd/lourdes\\_moreno.php](http://www.sigaccess.org/community/theses_repository/phd/lourdes_moreno.php) (2010). Accessed April 15<sup>th</sup> 2010.
- [12] Moreno, N., Romero, J., Vallecillo, A. An Overview of Model-Driven Web Engineering and the MDA. In: Rossi, G., Pastor, O., Schwabe, D., Olsina, L. (eds.) Web Engineering: Modelling and Implementing Web Applications. pp. 109-155. Springer-Verlag, London (2008)
- [13] Plessers P., Casteleyn S., Yesilada Y., De Troyer O., Stevens R., Harper S. & Goble C. Accessibility: A Web Engineering Approach. WWW (2005) doi:10.1145/1060745.1060799
- [14] Rossi, G., Schwabe, D. Modelling and Implementing Web Applications with OOHD. In: Rossi, G., Pastor, O., Schwabe, D., Olsina, L. (eds.) Web Engineering: Modelling and Implementing Web Applications. pp. 109-155. Springer-Verlag, London (2008)
- [15] Section 508. Electronic and Information Technology Accessibility Standards <http://www.section508.gov/index.cfm?fuseAction=stdsdoc> (2000-2010). Accessed 15 April 2010
- [16] Stanca Law. Italian Legislation on Accessibility. [http://www.pubbliaccesso.it/biblioteca/documentazione/guidelines\\_study/index.htm](http://www.pubbliaccesso.it/biblioteca/documentazione/guidelines_study/index.htm) (2004). Accessed 25 January 2010.
- [17] Vilain, P., Schwabe, D., Sieckenius de Souza, C. A Diagrammatic Tool for Representing User Interaction in UML. UML (2000) doi:10.1007/3-540-40011-7\_10
- [18] W3C: Web Content Accessibility Guidelines 1.0 and 2.0 (WCAG 1.0 and 2.0). <http://www.w3.org/WAI/intro/wcag> (1999/2008). Accessed April 15<sup>th</sup> 2010.
- [19] Woods, S. Websites for Visually Impaired Users. Thesis <http://wise.vub.ac.be/Downloads/Theses/Woods-thesis.pdf> (2006-2007). Accessed April 15<sup>th</sup> 2009.
- [20] Yesilada, Y., Harper, S., Goble, G. & Stevens, R. DANTE: Annotation and Transformation of Web Pages for Visually Impaired Users. WWW (2004) doi.acm.org/10.1145/1013367.1013540
- [21] Zimmermann, G. & Vanderheiden, G.: Accessible Design and Testing in the Application Development Process: Considerations for an Integrated Approach. Universal Access in the Information Society 7(1-2), 117-128 (2008).

# Cambio de Creencias: Operadores de Contracción utilizando Cláusulas de Horn

Néstor Jorge Valdez<sup>†</sup>

Marcelo A. Falappa<sup>‡</sup>

<sup>†</sup> Departamento de Ciencias de la Computación, Fac. de Ciencias Exactas y Naturales  
Universidad Nacional de Catamarca (UNCa)  
Av. Belgrano 300 - San Fernando del Valle de Catamarca  
Tel.: (03834)420900 / Cel: (03834) 154591186  
e-mail: njvaldez@exactas.unca.edu.ar

<sup>‡</sup> Laboratorio de Investigación y Desarrollo en Inteligencia Artificial  
Departamento de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur,  
Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)  
Av. Alem 1253, (B8000CPB) Bahía Blanca, Argentina  
Tel: (0291)4595135 / Fax: (0291)4595136  
e-mail: mfalappa@cs.uns.edu.ar

## Resumen

Esta línea de investigación tiene como principal objetivo explorar las conexiones entre las diferentes construcciones para contracciones con lógica Horn en el área de la teoría de cambio de creencias. En años recientes se han planteado diversos enfoques para el uso de lenguajes con expresividad acotada como, las cláusulas de Horn. Delgrande propuso funciones como la *entailment-based contraction* (*e-contraction*) y la *inconsistency-based contraction* (*i-contraction*). Richard Booth et al., desarrollaron a partir de la *entailment-based contraction* la operación conocida como *infra-contraction*. Zhuang y Pagnucco investigaron la correspondencia entre las relaciones de importancia epistémica *EE-relations* y las *P-relation* vinculándolas con sus respectivas construcciones de operadores de contracción. Como resultados, lograron caracterizar sus constructores *EEHornC* y *TRPMHornC*. La contribución de la presente investigación se resume en lo siguiente: establecer formalmente los problemas que surgieron en desarrollos anteriores y que soluciones proporcionan estos nuevos enfoques. Este tema de investigación

resulta de interés ya que la lógica Horn ha encontrado un amplio uso en campos como en la IA, BD Deductivas, PL, entre otros.

**Palabras Claves** Cambio de Creencias, Operadores de Contracción, Constructores de Contracción Horn.

## Contexto

Esta investigación se realizará dentro de la planificación establecida para el desarrollo (por parte del autor-miembro del presente trabajo, Lic. N. J. Valdez) de la tesis de Maestría en Ciencias de la Computación. La presente línea de investigación representará las bases para la continuidad de una investigación el cual tendrá como meta el desarrollo y defensa de un trabajo de tesis de Doctorado en Ciencias – Mención Computación – por parte del mismo autor-miembro de este trabajo. Ambas carreras de posgrado se cursaron dentro del ámbito de la Universidad Nacional de Catamarca (UNCa), la primera de ellas como parte de un convenio marco entre la UNCa y la Universidad Nacional del Sur (UNS).

## 1. Introducción

### 1.1. Cambio de Creencias

Es posible ilustrar diversos escenarios que representen un mundo y en los cuales un agente inteligente y racional tenga que verse obligado a ajustar esas creencias actuales de alguna manera apropiada cuando coteje con nueva información. El análisis de los patrones de razonamiento que implica esta tarea se denomina cambio de creencias o dinámica de conocimiento. El planteo de la problemática del cambio de creencias está basado en el uso de un formalismo lógico, ya que tanto las antiguas creencias del agente así como la nueva información se representan mediante un lenguaje lógico formal, al menos proposicional. La investigación en este área usualmente se centra en el uso de una lógica subyacente que asume una lógica proposicional clásica. No obstante, se sabe que la lógica por sí sola no resulta suficiente para conseguir respuestas aceptables a esta clase de problemas.

Se considera que el enfoque de AGM (Alchourrón, Gärdenfors y Makinson) [2, 10] constituye el marco más influyente en el área de cambio de creencias para conjuntos de creencias. Sin embargo, el uso de bases de creencias en lugar de conjuntos clausurados puede resultar más atractivo según lo estableció Sven Ove Hansson en diversos trabajos [13]. Ambos enfoques tienen mucho en común, pero tienen la necesidad de definir una estructura extra-lógica con el fin de obtener resultados satisfactorios. Para esta investigación se abordará el estudio de la teoría de cambio de creencias mediante un lenguaje más débil que la lógica tradicional en cuanto a su expresividad, pero con muchos atractivos computacionales como lo son su tratabilidad. Este lenguaje está compuesto por cláusulas de Horn. En el mismo, las creencias de un agente se representan en una base de conocimiento compuesta por cláusulas de Horn, y la entrada epistémica también es un conjunto de cláusulas de Horn.

### 1.2. Operación de Contracción de Creencias

Un enfoque muy común para abordar el cambio de creencias es proporcionando un conjunto de postulados de racionalidad para definir las funciones de cambio de creencias. El enfoque más recurrente es el AGM [2, 10] que define un marco de caracterización mediante conjuntos de postulados. Para el caso de la contracción, la misma representa una situación en la que el agente pierde información. En otras palabras, la contracción de un conjunto de creencias mediante una sentencia es otro conjunto de creencias en el que esa sentencia perdió validez. Formalmente, una contracción – es una función de la forma  $2^{\mathcal{L}} \times \mathcal{L} \Rightarrow 2^{\mathcal{L}}$  la que satisface los siguientes postulados:

$$(K - 1) \quad K - \varphi = Cn(K - \varphi)$$

$$(K - 2) \quad K - \varphi \subseteq K$$

$$(K - 3) \quad \text{Si } \varphi \notin K, \text{ entonces } K - \varphi = K$$

$$(K - 4) \quad \text{Si } \not\vdash \varphi, \text{ entonces } \varphi \notin K - \varphi$$

$$(K - 5) \quad K \subseteq (K - \varphi) + \varphi$$

$$(K - 6) \quad \text{Si } \varphi \equiv \psi, \text{ entonces } K - \varphi \equiv K - \psi$$

$$(K - 7) \quad (K - \varphi \cap K - \psi) \subseteq K - (\varphi \wedge \psi)$$

$$(K - 8) \quad \text{Si } \psi \notin K - \varphi \wedge \psi \text{ entonces } K - (\varphi \wedge \psi) \subseteq K - \psi$$

Los primeros seis postulados conforman los postulados básicos mientras que los dos restantes son los postulados extendidos. Varios tipos de construcciones han sido propuestos para caracterizar este tipo de cambio de creencias. Tenemos entre otras, la *epistemic entrenchment* propuestas de Gärdenfors [10], las operaciones de contracción que surgen del sistema de esferas propuesto por Grove [1] (quien adapta los sistemas de esferas propuestos por Lewis para contrafácticos [7]). Las operaciones de contracción de Rott [11] fallan en satisfacer el postulado (K - 5). El modelo de contracciones AGM utiliza los *remainder sets*. En base a los mismos, surgen tres clases de funciones de contracción principales definidas de la siguiente manera:

- Maxichoice contraction: selecciona un único remainder set.
- Partial meet contraction: selecciona un subconjunto no vacío de remainder sets y los intersecta.
- Full meet contraction: intersecta todos los remainder set.

Diversos trabajos han abordado la problemática de cambio de creencias, especialmente donde la operación de contracción se refiere donde la lógica subyacente está gobernada por cláusulas Horn.

## 2. Línea de Investigación: Contracción de Creencias con Cláusulas de Horn

Delgrande trabajo sobre funciones de contracción para conjuntos de creencias (conjuntos de sentencias que están clausurados bajo consecuencia lógica) restringidas a cláusulas de Horn. Sus principales resultados fueron las siguientes:

- Pasar a una lógica con cláusulas de Horn nos conduce a dos tipos diferentes de funciones de contracción, conocidas como *entailment-based contraction* (e-contraction) (i.e., cuando una base de conocimiento con cláusula Horn implica a una creencia, y queremos debilitar la base de conocimiento a fin de que la creencia no esté implicada) y la *inconsistency-based contraction* (i-contraction) (i.e., cuando al incorporar una creencia a una base de conocimiento daría lugar a una inconsistencia, por lo que debemos debilitar la base de conocimiento para que la inconsistencia no resulte si la creencia se ha añadido),
- La contracción sobre cláusulas de Horn no satisface el postulado Recovery. No obstante, presenta algunas características que se asocian generalmente con la contracción de

bases de creencias (conjuntos arbitrarios de sentencias),

- Por último, concluyó que la orderly maxichoice contraction es el método apropiado para la contracción con cláusulas de Horn.

Es cierto que las *meet constructions* de Delgrande [14, 15] son opciones apropiadas para la contracción con cláusulas de Horn. Sin embargo, Richard Booth et al. [4], demuestran que no constituyen todas las formas apropiadas de contracción sobre cláusulas de Horn. Es por ello que Richard Booth et al. [4, 5, 6] desarrollaron, a partir de la *entailment-based contraction* de Delgrande una construcción más refinada para contracción de conjuntos de creencias conocida como *infra-contraction*. Ellos demostraron que la infra-contraction se corresponde exactamente con el método de construcción para *kernel contractions* propuestas en [12] para conjuntos de creencias cuando está restringido a un lenguaje con cláusulas de Horn. Esto significó considerar el aspecto híbrido del cambio de creencias sobre cláusulas de Horn.

Otros trabajos han logrado caracterizar la semántica de la contracción Horn a través del uso de sistemas de esferas como en las investigaciones de Fotinopoulos y Papadopoulos [8], y también mediante una relación de importancia epistémica, como la propuesta por Zhuang and Pagnucco [16, 17]. A propósito de estos últimos, Zhi Qiang Zhuang y Maurice Pagnucco investigaron sobre la correspondencia entre las relaciones de importancia epistémica (EE-relation), las relaciones de preferencia (P-relation), y sus correspondientes construcciones de operadores de contracción: epistemic entrenchment based contraction (EEC) y la transitively relational partial meet contraction (TRPMC). Exploraron también las conexiones entre ambas relaciones utilizando cláusulas de Horn y, al no encontrar una correspondencia recíproca, concluyeron que *bajo cláusulas de Horn la correspondencia entre EEC (Horn) y TRPMC (Horn) resultan diferentes*. Finalmente, concluyen que sus postulados de caracterización para EEHornC son muy específicas



para TRPMHornC, y que estos últimos resultan ser más amplios que EEHornC.

Billington et al. [3] consideraron para la contracción que la lógica rebatible resulta en muchos aspectos muy diferente de la lógica de Horn. Sin embargo, establecen ciertas reglas que sutilmente tienen alguna similitud con la lógica de Horn. Además de las construcciones e-contraction e i-contraction, [4] consideraron la *package contraction* (p-contraction), similar a la contracción propuesto por Fuhrmann y Hansson [9] donde definen contracciones múltiples. Dado un conjunto de sentencias X, el objetivo es asegurar que ninguna de las sentencias en X forme parte del resultado obtenido de una p-contraction. Resulta que su aplicación en cláusulas de Horn genera ciertos matices interesantes según [15].

### 3. Resultados y Objetivos esperados

A través de esta línea de investigación la atención estará centrada en el cambio de creencias y en particular en las operaciones de contracción. Se identifican tópicos de estudio interesantes, los que todavía no han ofrecido resultados concretos. Por ejemplo, ampliar el alcance de las Horn Contraction para Full Meet Contractions, considerando para ello los postulados extendidos, analizar el alcance de los modelos teóricos de contracción Horn propuestos para los constructores TPMHC y similares, como así también sus semánticas definidas.

Mediante esta investigación se busca determinar la capacidad y alcance de estos operadores de contracción para de esta manera, poder definir nuevas propuestas de funciones que vinculen a contracciones y revisiones a través de una lógica proposicional más débil en cuanto a representatividad, pero más flexible y adaptable en cuanto a costo computacional.

Se espera alcanzar los siguientes objetivos:

- Elaborar un marco de referencia que resulte una guía de observación y consulta en la caracterización de constructores de operado-

res de cambio. Al respecto, se espera poder definir explicaciones que resulten aceptables para el investigador que recibe una recomendación.

- Estudiar distintos constructores de operadores de contracción Horn, presentando postulados característicos, y evaluar su uso y aplicación.
- Analizar el grado de utilidad en diversas situaciones de los distintos métodos de construcción, operadores y técnicas de contracciones Horn. Con esto se pretende mejorar la capacidad del usuario al momento de tomar una decisión.
- Sistematizar las caracterizaciones formales de los operadores y funciones de contracción Horn.

### 4. Formación de Recursos Humanos

El Docente-Investigador Néstor Jorge Valdez es dependiente del Programa de Desarrollo Científico y Tecnológico (adherido al Programa Nacional de Incentivos a los Docentes-Investigadores) de la Secretaría de Ciencia y Tecnología de la Universidad Nacional de Catamarca y que cursa las carreras de Maestría en Ciencias de la Computación (UNS) y el Doctorado en Ciencias Mención Computación (UNCa). Actualmente, se encuentran en desarrollo:

- Completar la tesis de Maestría en Ciencias de la Computación (en convenio con la Universidad Nacional del Sur) titulada "*Cambio de Creencias: Operaciones de Cambio en Lenguajes con Cláusulas de Horn*".
- Luego se extenderá la investigación para alcanzar el grado de Doctor en Ciencias con Mención Computación, dependiente de la Universidad Nacional de Catamarca.

## Referencias

- [1] Grove A. Two modellings for theory change. *Journal of Philosophical Logic* 17, pages 157–170, 1988.
- [2] C. Alchourrón, P. Gärdenfors, and D. Makinson. On the logic of theory change: Partial meet contraction and revision functions. *The Journal of Symbolic Logic*, (50):510–530, 1985.
- [3] Antoniou G. Governatori G. Billington, D. and M. Maher. Revising nonmonotonic theories: The case of defeasible logic. In *Proceedings of the 23rd Annual German Conference on Artificial Intelligence, No. 1701 in LNAI*, pages 101–112, 1999.
- [4] Meyer T. Booth, R. and I. Varzinczak. Next steps in propositional horn contraction. In *Boutilier, C. (Ed.), Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI)*, pages 702–707, 2009.
- [5] Meyer T. Varzinczak I. Booth, R. and R. Wassermann. A contraction core for horn belief change: Preliminary report. In *13th International Workshop on Nonmonotonic Reasoning (NMR), (2010a/b)*, 2010.
- [6] T.; Varzinczak I.; Booth, R.; Meyer and R. Wassermann. On the link between partial meet, kernel, and infra contraction and its application to horn logic. *Journal of Artificial Intelligence Research*, pages 31–53, 2011.
- [7] Lewis D. Completeness and decidability of three logics of counterfactual conditionals. *Theoria* 37, pages 74–85, 1971.
- [8] A. Fotinopoulos and V. Papadopoulos. Semantics for horn contraction. In *7th PanHellenic Logic Symposium*, pages 42–47, 2009.
- [9] A. Fuhrmann and S. Hansson. A survey of multiple contractions. *Journal of Logic, Language and Information*, pages 39–76, 1994.
- [10] P. Gärdenfors and Makinson. Revisions of knowledge systems using epistemic entrenchment. In *Proceedings of the second conference on Theoretical aspects of reasoning about knowledge*, pages 83–95, 1988.
- [11] Rott H. Preferential belief change using generalized epistemic entrenchment. *JoLLI*, pages 45–78, 1992.
- [12] O.S. Hansson. Kernel contraction. *J. of Symbolic Logic* 59(3), pages 845–859, 1994.
- [13] S. O. Hansson. A dyadic representation of belief. in belief revision. *Vol. 29 of Cambridge Tracts in Theoretical Computer Science*, Cambridge University Press, pages 89–121, 1992.
- [14] Delgrande J.P. Horn clause belief change: Contraction functions. In *Gerhard Brewka and J  rome Lang, editors, Proceedings of the Eleventh International Conference on the Principles of Knowledge Representation and Reasoning, Sydney, Australia, 2008. AAAI Press*, pages 156–165, 2008.
- [15] R. Wassermann and J. P. Delgrande. Topics in horn contraction: Supplementary postulates, package contraction, and forgetting. In *Proc. NRAC-2011*, 2011.
- [16] Zhi Qiang Zhuang and Maurice Pagnucco. Horn contraction via epistemic entrenchment. In *Tomi Janhunen and Ilkka Niemel  a, editors, Logics in Artificial Intelligence - 12th European Conference (JELIA 2010), volume 6341 of Lecture Notes in Artificial Intelligence*, pages 339–351, 2010.
- [17] Zhi Qiang Zhuang and Maurice Pagnucco. Transitively relational partial meet horn contraction. In *Proc. IJCAI-2011*, 2011.

# Aplicación de Algoritmos Genéticos a la Corrección de Matrices Inconsistentes en el Proceso Analítico Jerárquico (AHP)

Marcelo Karanik, Leonardo S. Wanderer

Grupo de Investigación Sobre Inteligencia Artificial (GISIA)  
Universidad Tecnológica Nacional – Facultad Regional Resistencia  
French 414, Resistencia (3500) – Chaco. (0362) 4 43 26 83

marcelo@frre.utn.edu.ar, leonardo.wanderer@gmail.com

## Resumen

La combinación de diversas técnicas de la inteligencia artificial da lugar a que problemas complejos puedan ser resueltos de formas más eficientes y eficaces desde el punto de vista práctico.

Puede aprovecharse así la utilización de técnicas muy estudiadas y potentes, como los algoritmos genéticos, a la solución de los problemas de consistencia de las matrices que se utilizan en el Proceso Analítico Jerárquico, para reducir su inconsistencia al mismo tiempo que se evita introducir una arbitrariedad ajena a la decisión de un experto en dichas matrices.

La inconsistencia surge producto de la complejidad inherente a la toma de decisiones en donde los factores a considerar cumplen con dos criterios: el criterio objetivo, que determina múltiples y complejas relaciones entre las variables a tener en cuenta; y el criterio subjetivo, vinculado con los conocimientos, capacidades y percepciones de la persona encargada de tomar la decisión.

**Palabras clave:** Toma de Decisiones, AHP, Juicios Inconsistentes, Comparaciones Faltantes, Algoritmos Genéticos.

## Contexto

La presente línea de investigación se desarrolla en el marco del proyecto “*Diseño de Técnicas para el Tratamiento de Situaciones de Incertidumbre en Sistemas de Soporte de Decisiones con Múltiples Expertos*” Código UTN-1315. Dicho proyecto se lleva a cabo en el ámbito del GISIA perteneciente a la Facultad Regional Resistencia de la UTN.

## Introducción

El Proceso Analítico Jerárquico (AHP) [1] es una teoría general para determinar prioridades que retorna una escala de proporciones basada en comparación de elementos de un conjunto por pares. Estas comparaciones son realizadas por mediciones reales que se pueden tomar de dichos elementos o a partir de una medida de la preferencia, o valoración subjetiva, utilizando lo que se conoce como la escala fundamental. AHP tiene una amplia aplicación en la toma de decisiones multicriterio [2], planificación y asignación de recursos [3], y en resolución de conflictos [4]. Permite realizar una conclusión o síntesis en base a múltiples factores que se desean considerar haciendo operaciones sobre las comparaciones y valoraciones realizadas.

En el contexto de la toma de decisiones (TD) multicriterio, los problemas planteados se caracterizan por la existencia de uno o más decisores o “expertos”, que ante un conjunto de alternativas y criterios de evaluación para dichas alternativas, se encargan de valorar, por un lado los criterios entre sí, y por otro lado, en base a cada uno de los criterios, a todas las alternativas entre sí. Esta comparación requiere de matrices que se diseñan para la comparación por pares caracterizadas por ser positivas y recíprocas, por lo que el decisor deberá realizar  $n \cdot (n-1)/2$  comparaciones para cada una, siendo “n” la dimensión de dicha matriz.

AHP originalmente transforma las comparaciones de un único decisor en un ranking de alternativas. Cuando se debe utilizar el conocimiento de múltiples decisores, se debe añadir en AHP un procedimiento de agregación [5] que combine dicha información.

Una situación que puede darse al comparar elementos de a pares, en una matriz, es que las valoraciones que se realizan hagan que las relaciones de transitividad entre dichos elementos no se cumpla y por lo cual “no tengan sentido” estas relaciones. Por ejemplo, si valoramos a un elemento A como más importante que uno B, y también que B sea más importante que uno C, lógicamente A debe ser bastante más importante que C; pero si el decisor indica que la relación entre C y A no es de esa forma, dada la transitividad, existiría un grado de error o “inconsistencia” en su valoración. Cuando estas relaciones involucran pocos elementos es sencillo identificar la existencia de un error, pero a medida que crecen las matrices de comparación, y debemos valorar más elementos, existen más y más complejas relaciones de transitividad y dicho error deja de hacerse notorio.

He aquí donde se debe destacar un requerimiento de AHP cuyo tratamiento involucra el desarrollo de esta línea de

investigación: la inconsistencia de las matrices del proceso.

Por otro lado, se requiere que todas las matrices estén completas para que puedan calcularse las prioridades, esto implica un problema adicional ya que se exige al decisor realice todas las comparaciones posibles.

Únicamente cuando las matrices están completas y son consistentes puede calcularse el ranking de las alternativas.

Como se mencionó anteriormente, la inconsistencia lógica-matemática puede asociarse a la complejidad de las relaciones y, además, a la “ignorancia” por parte del decisor al momento de presentarse un problema con determinadas características que hacen que no pueda responder o evaluar con seguridad un conjunto de elementos. Existen 3 formas de ignorancia [6] que influyen al momento de comparar los valores en las matrices AHP:

- **incompleto**: falta de establecimiento de un juicio de valor, a causa de una falta de comprensión o por el esfuerzo de ahorrar tiempo que implica proveer una respuesta completa.
- **imprecisión**: incapacidad para brindar un criterio o valoración numérica de forma exacta, es decir, que refleje correctamente la perspectiva del decisor en la comparación.
- **incertidumbre**: asociada a la probabilidad de que los juicios basados en las opciones que el decisor elija son realmente lo que él cree que son.

### Consistencia y Completitud

Para satisfacer el criterio de consistencia en AHP, primero debemos poder calcular el ratio de consistencia (CR) de cada matriz [7] y debemos asegurarnos que su valor se encuentre por debajo del valor 0.1, lo que significa que la matriz es suficientemente consistente para obtener un autovector, que



represente adecuadamente la importancia relativa de los elementos comparados. Sólo así puede ser utilizado en el proceso de síntesis y obtención de la mejor alternativa en base al criterio de los expertos. En el caso de tener un valor del ratio de consistencia mayor al máximo mencionado, debemos adoptar un método que lo disminuya y mejore así la consistencia de la matriz, sea automático o manual [8].

Existen varios métodos que corrigen los problemas de inconsistencias en las matrices de AHP. A continuación se exponen 2 de ellos:

El método original de Saaty [7] modifica de forma arbitraria ciertos valores de la matriz, pudiendo alterar la importancia real que el decisor le confirió sin importar si dicha valoración es correcta en base a sus preferencias o fue producto de uno de los tipos de ignorancia nombrados.

Otro método propone la utilización de redes neuronales que realizan cambios globales en la matriz y logran mejoras de la consistencia sin comprometer la importancia en la valoración que el decisor. [9]

Puede observarse que el segundo método es más conveniente por su poca arbitrariedad en el cambio de la decisión del experto, ya que permite que este último sea el responsable de la decisión y no el método en sí mismo.

A partir de ello, se puede definir como objetivo un método que asista al decisor en aquellos casos que produzcan una inconsistencia o corrija los valores de manera que esta sea aceptable, al mismo tiempo que se mantenga su decisión, es decir, en el autovector generado.

Al observar las características del problema, se puede identificar una estructura del modelo de la investigación de operaciones [10], en donde se debe optimizar un criterio objetivo sujeto a un conjunto de restricciones. En este caso, el criterio objetivo debe suponer la

introducción de los menores cambios posibles a la matriz original (minimizando las distancias entre los valores originales y los propuestos), mientras que la restricción es el valor de inconsistencia que se busca, que debe ser menor a 0.1. Alternativamente, en el criterio objetivo también se puede considerar la conservación del orden de importancia de los elementos comparados (el autovector).

### Algoritmos Genéticos

Dentro del espectro de las técnicas de la inteligencia artificial existen un conjunto de métodos que se pueden aplicar a estos problemas de optimización, entre lo que se encuentran los algoritmos evolucionarios, y dentro de este grupo, los algoritmos genéticos (AG).

Al intentar completar los valores faltantes y corregir aquellos que provocan inconsistencia, la función que deberá guiar la búsqueda dentro del espacio de soluciones (función heurística) es, en caso de que fuera posible, difícil de determinar. En este tipo de contexto los AG pueden brindar una solución práctica. Otras de las ventajas de los AG que se pueden nombrar son la simplicidad conceptual, amplio grado de aplicabilidad, facilidad para combinarse con otros métodos (por ejemplo para optimizar el rendimiento de redes neuronales, sistemas difusos, sistemas de redes y otras estructuras), procesamiento en paralelo, robusto ante cambios dinámicos del entorno (gracias a que una población solución sirve de punto de partida para posteriores mejoras) y resuelve problemas que no tienen solución o en los que no existe pericia humana suficiente para resolverlo. [11]

Los AG requieren, en primer lugar, definir un esquema de representación del "individuo", siendo esta la estructura de las posibles soluciones al problema. En segundo lugar, la función de aptitud de ese individuo, que permite

valorar y ordenar distintas soluciones, mientras se guía el proceso de búsqueda de la(s) mejor(es) alternativa(s). [11] Aplicándolo al problema en cuestión, la representación del individuo puede ser una estructura de matriz o vector, que se definirá al momento del diseño e implementación del sistema. La función de aptitud deberá definirse considerando el criterio objetivo de la optimización establecido anteriormente.

El proceso evolutivo empieza con la generación de una población inicial, en este caso, un conjunto de matrices o vectores. Esta población tiene la característica de ser aleatoria, pero puede restringirse a ciertas condiciones. Posteriormente, se aplican operadores [12] de selección, cruce y mutación de forma iterativa que produce generaciones de individuos cada vez mejores que reemplazan a los anteriores y cuando alcanza la solución, o algún criterio particular, finaliza.

## Líneas de Investigación y Desarrollo

En el proyecto “*Diseño de Técnicas para el Tratamiento de Situaciones de Incertidumbre en Sistemas de Soporte de Decisiones con Múltiples Expertos*” (UTN-1315) se diseñó e implementó un sistema de soporte de decisiones que utiliza IL (Información Lingüística) [13] y AHP. Este sistema permite que múltiples expertos puedan participar conjuntamente en proyectos que ayuden a la toma de decisiones, cuya información se procesa utilizando las técnicas anteriormente mencionadas. Cada uno de los proyectos creados tiene un conjunto de alternativas y un conjunto de criterios que el experto deberá valorar en base a su nivel de experticia. El procesamiento y obtención de la conclusión del sistema puede realizarse para un experto en particular y para todos los expertos que participen en el

proyecto. La información provista por un experto que no se encuentra completa y/o suficientemente consistente es omitida en el procesamiento.

Del proyecto de investigación nombrado surge la línea de investigación relacionada a la utilización de algoritmos genéticos aplicados a la recuperación de valores faltantes y asistencia del usuario al momento de completar las matrices de comparación entre pares.

La propuesta involucra el diseño e implementación de un módulo que se integrará al sistema ya desarrollado. Este deberá facilitar el proceso de carga de la información de alternativas y criterios por parte de los expertos, mediante sugerencias para los valores faltantes y los ingresados, mejorando así el grado de consistencia de dicha información. Las sugerencias serán realizadas mediante el uso de algoritmos genéticos y tomará como entrada del proceso la información ya ingresada por el experto.

## Resultados y Objetivos

Dados los requerimientos de AHP y los problemas que surgen en la carga de la información de las matrices, el objetivo general es mejorar la consistencia de las matrices y facilitar ayuda al experto para completar estas matrices.

Para lograr esto se definieron los siguientes objetivos específicos:

- Relevamiento de información.
- Diseño del modelo de AG.
- Diseño del módulo computacional.
- Programación.
- Pruebas del módulo.
- Rediseño de la interfaz de usuario de carga.
- Integración con el sistema de soporte de decisiones.

## Formación de Recursos Humanos

### Formación de Postgrado

A partir de las líneas de investigación desarrolladas en el proyecto se han logrado finalizar dos tesis doctorales. Los ingenieros Sergio Gramajo y Marcelo Karanik han finalizado sus tesis doctorales del programa de Doctorado en Ingeniería de Sistemas y Computación de la Universidad de Málaga (España) en temáticas de TD relacionadas al actual proyecto.

### Formación de Becarios

Los becarios Antonio Manuel Gimenez y Leonardo Simón Wanderer son alumnos avanzados de la carrera de ingeniería en sistemas de información que han sido beneficiados con una beca de finalización de carrera y colaboran en las actividades de investigación relacionadas con el proyecto.

### Equipo de Trabajo

La estructura del equipo de trabajo es la siguiente:

- Director del Proyecto: Dr. Marcelo Karanik.
- Docente-Investigador: Dr. Sergio Gramajo.
- Alumnos becarios:
  - Antonio Manuel Gimenez.
  - Leonardo Simón Wanderer.

## Referencias

1. Saaty, R. W.: The analytic hierarchy process - What it is and how it is used. *Math Modelling*, Vol. 9, No. 3-5, pp. 161-176 (1987).
2. Triantaphyllou, E., Shu, B., Nieto Sanchez, S., Ray, T.: *Multi-Criteria Decision Making: An Operations Research Approach*. *Encyclopedia of Electrical*

and Electronics Engineering, (J.G. Webster, Ed.), John Wiley & Sons, New York, NY, Vol. 15, pp. 175-186, (1998).

3. Russell, S., Norvig, P.: *Inteligencia Artificial: Un enfoque moderno*. Prentice Hall Hispanoamericana S.A., México (1996).
4. Saaty, T. L., Alexander, J.: *Conflict Resolution: The Analytic Hierarchy Process*. Praeger, New York (1989).
5. Gramajo, S., Karanik, M., Pinto, N., Cabrera, D., Alurralde, M.: *Modelo de Apoyo para la Toma de Decisiones en QoS*. CACIC (2011).
6. Ozdemir, M. S., Saaty, T. L.: The unknown in decision making - What to do about it. *European Journal of Operational Research* 174, pp. 349-359 (2006).
7. Saaty, T L.: How to make a decision: The Analytic Hierarchy Process. *European Journal of Operational Research* 48 pp. 9-26 (1990).
8. Gramajo, S., Karanik, M., Cabrera, D., Alurralde, M., Ojeda, P., Pinto, N.: *Diseño de Técnicas para el Tratamiento de Situaciones de Incertidumbre en Sistemas de Soporte de Decisiones con Múltiples Expertos*. WICC (2012).
9. Gomez-Ruiz, J.A., Karanik, M., Pelaez, J. I.: Estimation of missing judgments in AHP pairwise matrices using a neural network-based model. *Applied Mathematics and Computation*, Elsevier. Vol. 216(10), pp: 2959-2975. (2010).
10. Taha, H. A.: *Investigación de Operaciones*. Pearson Educación, México (2004).
11. Sivanandam, S. N., Deepa, S. N.: *Introduction to Genetic Algorithms*. Springer, Heidelberg (2008).
12. García Martínez, R., Servente, M., Pasquini, D.: *Sistemas Inteligentes*. Nueva Librería SRL (2003).
13. Gramajo, S., Martínez, L.: A linguistic decision support model for QoS priorities in networking. *Knowledge-Based Systems, New Trends on Intelligent Decision Support Systems* 32(0): pp. 65-75 (2012).

# Diseño de un Sistema para Alertar Riesgo en el Déficit de Micronutrientes

Marcelo Karanik, Roberto Suénaga, Fabián Favret,  
Karina Eckert, Maria Eugenia Barreyro

Departamento de Inv. y Desarrollo Universidad Gastón Dachary

Salta 1968, Posadas, Misiones - Argentina. Tel: +54 (0376) - 4438677

mkaranik@gmail.com, rsuenaga@dachary.edu.ar, fabianfavret@gmail.com,

karinaeck@gmail.com, mariaeugeniabarreyro@yahoo.com.ar

## Resumen

Hoy en día vivimos en una sociedad globalizada, donde los factores que determinan lo que se ingiere, está muy influenciado por las consecuencias sociales y culturales de los hábitos alimentarios. Esto a su vez, produce deficiencias nutricionales por exceso o déficit de micronutrientes como vitaminas y minerales. Las deficiencias más comunes y severas, son la de hierro y calcio, que pueden causar enfermedades como la anemia y osteoporosis. Por lo que es de suma importancia para la salud pública, detectar tempranamente el riesgo de déficit de los micronutrientes, analizando la ingesta de los mismos, y los factores que influyen en su absorción y utilización.

En busca de colaborar con los profesionales de la salud, este proyecto apunta al desarrollo de un sistema que alerte sobre posibles cuadros de déficit de micronutrientes, utilizando para ello, técnicas de inferencia probabilística y de redes neuronales artificiales. La idea básica es que el sistema capte la información de la historia clínica del paciente y a partir de esos datos pueda alertar, mediante el uso de redes de bayesianas el riesgo de déficit de hierro y calcio.

**Palabras clave:** Déficit de Micronutrientes, Sistemas Expertos, Inferencia Probabilística/Bayesiana, Redes Neuronales Artificiales.

## Contexto

El presente proyecto de investigación se encuentra en ejecución, adjudicado en el marco del 5° Llamado a Concurso de Proyectos de Investigación, convocado por la Secretaría de Investigación y Desarrollo de la Universidad Gastón Dachary en el año 2012.

## Introducción

### El Problema del Déficit De Micronutrientes (Desnutrición Oculta)

El Dr. Pedro Escudero definía a la nutrición como: “*el resultado o resultante de un conjunto de funciones armónicas y solidarias entre sí que tienen como finalidad mantener la composición e integridad normal de la materia y conservar la vida*” [1]. Cuando esas funciones se desarrollan de una manera incorrecta aparecen los problemas de la malnutrición. En la actualidad dichos problemas se clasifican en malnutrición por exceso como las enfermedades crónicas no transmisibles y malnutrición por déficit dentro de las cuales encontramos a la desnutrición oculta o el déficit de micronutrientes.

Entre los micronutrientes que presentan mayores riesgos de déficit en nuestra población, se encuentran el hierro, que afecta mayoritariamente a la población infantil, adolescentes, a mujeres en edad fértil y mujeres embarazadas; y como consecuencia, entre otras situaciones aparece anemia. Se define anemia como “*disminución de la masa de glóbulos rojos o de la concentración de hemoglobina*”



*por debajo del segundo desvío estándar respecto de la media para edad y sexo” [2].*

Otro mineral que presenta déficit es el calcio, su carencia afecta a las personas mayores, fundamentalmente a las mujeres en el clima-terio y esto es uno de los factores de riesgo de osteoporosis.

La anemia por deficiencia de hierro es una de las carencias nutricionales más frecuentes en el mundo. Su presencia afecta la respuesta inmune y la resistencia a las infecciones, el metabolismo del músculo, el desarrollo intelectual, la capacidad cognitiva y de trabajo, y la regulación de la temperatura corporal. Así como el retraso de talla con peso normal o elevado respecto de su talla actual es la forma prevalente de malnutrición en los niños, la deficiencia de hierro es la forma prevalente del déficit de micronutrientes. Ambas situaciones pueden coexistir en los mismos individuos, estratos sociales y regiones, por otro lado, la deficiencia de micronutrientes en individuos aparentemente sanos suele pasar inadvertida y puede traer consecuencias críticas en la salud [3]. Por otro lado, la osteoporosis es una enfermedad esquelética sistémica caracterizada por baja masa ósea y deterioro de la microestructura del tejido óseo, con el consecuente incremento en la fragilidad ósea y la susceptibilidad a las fracturas [4]. Por lo tanto es de suma importancia para la salud pública, detectar tempranamente el riesgo de déficit de los micronutrientes, analizando no solamente la ingesta de los mismos, sino también los factores que influyen en la absorción y por lo tanto en su utilización. Analizando también factores que mejoran el aprovechamiento y aquellos que inhiben; como así también el estilo de vida de los individuos que afectan positiva o negativamente en su salud.

### **Sistemas Expertos**

La detección de situaciones probables a partir de hechos concretos es una tarea que ha sido ampliamente abordada en el área de la inteligencia artificial (IA). Una gran variedad de técnicas [5][6][7][8][9][10] se han propuesto para tratar el tema del diagnóstico, generando gran cantidad de aplicaciones que dan apoyo

a las personas durante los procesos de toma de decisiones. Estas aplicaciones utilizan el conocimiento disponible para inferir posibles situaciones en base a evidencia circunstancial, de manera similar a la que lo haría un experto del área de problema. Por este motivo, se conoce a dichas aplicaciones como sistemas expertos. Este tipo de aplicaciones se integra al modelo de resolución de problemas de Simon [11] delegando al experto la etapa de selección de la alternativa que considere conveniente.

Básicamente, los sistemas expertos utilizan la representación del conocimiento del especialista para diagnosticar situaciones y, eventualmente, sugieren acciones que estiman adecuadas para resolver el problema [12].

Los sistemas de diagnóstico son muy utilizados en distintas áreas; industria, educación, medicina, etc., e implementan una gran variedad de técnicas de ia que van desde los sistemas de reglas lógicas [5][7][9][10], sistemas difusos [9][10], sistemas probabilísticos [5][7][13][14], sistemas de clasificación basados en instancias [10][15], redes neuronales artificiales [16][17][18][19], algoritmos genéticos [7][20] y sistemas de búsqueda heurísticos [5][7][9][10]. Estos sistemas realizan inferencias en función a los datos disponibles que permiten determinar las situaciones que expliquen dichos datos. Es decir, llegan al diagnóstico en base a los síntomas que se evidencian de una situación particular.

La idea subyacente en los sistemas de diagnóstico es la representación del conocimiento de un experto y la simulación del mecanismo de deducción que emplea en su tarea [10]. Claramente, la medicina es un ámbito propicio para el desarrollo de sistemas de diagnóstico y existe una elevada cantidad de este tipo de aplicaciones [21][22][23].

Específicamente, los sistemas expertos utilizados en el área de nutrición tienen, en general, el objetivo de elaborar dietas personalizadas de los pacientes en base a una serie de datos sobre hábitos alimentarios, antropométricos y bioquímicos. Estas herramientas asisten al profesional en el diseño de las dietas específicas para cada paciente, dependiendo

del objetivo que se pretenda alcanzar (aumento/disminución de peso, entrenamiento para el alto rendimiento, tratamiento de enfermedades específicas, etc.).

Se puede ver la elaboración de una dieta como el resultado del análisis de los datos recabados en una primera etapa de la atención de un paciente. Luego, una vez confeccionada la dieta, se realiza un seguimiento del paciente y de ser necesario, se ajusta dicha dieta a los cambios que puedan surgir. Los datos capturados inicialmente pasan a formar parte de la historia clínica del paciente que serán consultados con posterioridad en caso de necesidad.

Ese volumen de información puede ser útil en la detección de las alteraciones del estado nutricional del paciente y de los estados patológicos que de ellas se puedan derivar.

Por lo general los sistemas de diagnóstico bien ideados tropiezan con bases éticas y legales, aunque probablemente la velocidad de procesamiento, la integración con el esquema de flujo de diagnóstico del médico, y la base de conocimiento actualizada adecuadamente sean las verdaderas trabas a su aplicación. A pesar de esto, algunos autores apuntan a que el uso de estos sistemas mejora sustancialmente la práctica clínica.

Una solución adecuada que pueda hacer frente a las dificultades mencionadas requiere el diseño de una aplicación robusta, distribuida, de simple implementación y de fácil utilización. El desarrollo exitoso de un sistema con las mencionadas características, implica la participación y colaboración coordinadas de múltiples áreas, desde los profesionales médicos hasta los ingenieros que provean el soporte tecnológico para la implementación.

## Líneas de investigación y desarrollo

En este proyecto se apunta al desarrollo de un sistema de alerta que detecte riesgo de déficit de micronutrientes en pacientes cuando realizan una consulta médica. La idea básica es que el sistema capte la información de la historia clínica del paciente y a partir de esos datos pueda inferir sobre el estado nutricional.

Se pretende que el sistema funcione de manera complementaria y sin atención constante del profesional. Es decir, a medida se ingresan los datos a la historia clínica, el sistema de diagnóstico los captura y realiza inferencias que generan, en caso de ser necesario, las alertas correspondientes.

La funcionalidad puede describirse en las siguientes etapas:

- A) El profesional carga los datos de historia clínica del paciente (o los actualiza) teniendo en cuenta síntomas, datos antropométricos, información de análisis clínicos y mediciones que realiza en el consultorio.
- B) El sistema cliente de alerta capta los datos cargados y los envía al sistema servidor que realiza inferencias utilizando el conocimiento disponible sobre déficit de micronutrientes.
- C) En caso de que los datos recibidos denoten alguna alteración nutricional, el sistema servidor envía un informe de alerta con la situación detectada.

Los datos del paciente quedan almacenados en el sistema de historias clínicas, y serán utilizados nuevamente en alguna consulta posterior.

Cabe aclarar que esta propuesta se basa en el procesamiento remoto de los datos. Se cree conveniente utilizar esta estrategia ya que con la notable evolución de las tecnologías de telecomunicaciones y aplicaciones web, el acceso a internet es masivo y de bajo costo.

Para el procesamiento de los datos del paciente se propone utilizar un modelo probabilístico (utilizando inferencia bayesiana) que permita estimar las probabilidades de ocurrencia de cada anomalía. Con este modelo se pueden explicar los resultados obtenidos a partir del conjunto de datos inicial para que el profesional pueda entender el motivo de la alerta generada. Un aspecto importante de destacar es que la base de conocimiento en este tipo de modelos es fácilmente actualizable mediante la incorporación de los datos de diagnósticos confirmados.

Además, en el modelo de procesamiento se propone también la incorporación de un módulo neuronal de diagnóstico, paralelo e independiente al probabilístico. Con esto se busca mayor confiabilidad de los resultados obtenidos.

Es decir, utilizando el mismo conjunto de datos, se pretende compararlos con patrones previamente definidos mediante el uso de una red neuronal entrenada con la información de pacientes que hayan sufrido alguna de las anomalías.

Finalmente se propone diseñar un módulo de consenso que utilice los dos diagnósticos obtenidos y genere una salida unificada del sistema. Para ello se propone la utilización de técnicas de agregación utilizada en el área de toma de decisiones y lógica difusa. En caso que las salidas individuales sean substancialmente diferentes, el módulo remitirá ambas opciones.

## Resultados y Objetivos

### Objetivo general

Desarrollar un sistema de alerta sobre posibles cuadros de déficit de micronutrientes que utilice técnicas de inferencia probabilística y de redes neuronales.

### Objetivos específicos

- Realizar una revisión sistemática de la literatura científica.
- Adquirir mediante entrevistas con expertos locales, el conocimiento necesario para el desarrollo del sistema.
- Definir los perfiles de pacientes con problemas nutricionales.
- Representar el conocimiento del experto mediante redes de inferencia bayesiana.
- Utilizar redes neuronales de aprendizaje supervisado para la detección de situaciones de déficit de micronutrientes.
- Programar los módulos de inferencia y de consenso.
- Programar el sistema de captura de datos de hábitos alimentarios, antropométricos y bioquímicos.

- Programar el acceso web al sistema de inferencia.

## Formación de Recursos Humanos

Actualmente el equipo de trabajo de esta línea de investigación está compuesto por docentes de los Departamentos de Ingeniería y Ciencias de la Producción y de Ciencias de la Salud de la UGD.

### Director del Proyecto:

Dr. en Ing. de Sistemas y Computación: Marcelo Karanik.

### Coordinación Interdisciplinaria:

Ing. en Sistemas: Roberto Suénaga.

### Docentes-Investigadores:

Ing. en Informática: Fabián Favret, Karina Eckert.

Lic. en Nutrición: M<sup>a</sup> Eugenia Barreyro.

## Referencias

- [1] López L. Suárez M. "Fundamentos de Nutrición Normal". Editorial El Ateneo, Buenos Aires. 2003.
- [2] Donato, H. Cedola, A. Rapetti, M.C. Buys, M.C. Gutiérrez, M. Parias Nucci, R. Rossi, N. Schwartzman, G. "Anemia Ferropénica. Guía de Diagnóstico y Tratamiento". Comité Nacional de Hematología. 2009.
- [3] Kogan L. Aveyá Guillardón E. Biglieri, A. Mangialavori, G. Calvo, E. Durán, P. "Anemia: la desnutrición oculta – Resultados de la Encuesta Nacional de Nutrición y Salud". Ministerio de Salud de la Rep. Argentina. 2008.
- [4] Torresani, M. E. "Aprendamos a comer frente al cambio hormonal". Editorial Akadia, Buenos Aires. 2009.
- [5] Castillo, E. Gutierrez, J. Haidi, A. "Sistemas Expertos y Modelos de Redes Probabilísticas". Monografías de la Academia de Ingeniería, Univ. de Cantabria, España. 1998.
- [6] Dutta, A. "Integrating AI and Optimization for Decision Support: A Survey". Deci-

- sion Support Systems. Vol 18(3). Pag 217-226. 1996.
- [7] Nilsson, N. "Inteligencia Artificial, Una nueva síntesis". Mcgraw Hill. 2000.
- [8] Partridge, D. "Artificial Intelligence and Software Engineering - Understanding the Promise of the Future". Amacom. 1998.
- [9] Rich, E. Knight, K. "Inteligencia Artificial". Mc graw hill. 1996.
- [10] Russell, S. Norvig, P. "Inteligencia Artificial, un Enfoque Moderno". Pearson Prentice Hall. 2004.
- [11] Simon, H.A. "The new science of management decision". NY. Harper and brothers. 1960
- [12] Arnott, D. Pervan, G. "Eight Key Issues for the Decision Support Systems Discipline". Decision Support Systems. Vol 44(3). Pag 657-672. 2008.
- [13] Neapolitan, R. E. "Probabilistic Reasoning in Expert Systems - Theory and Algorithms". Wiley 1990.
- [14] Pearl, J. "Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference". Morgan Kaufmann Publishers Inc. San Francisco, CA, USA. 1988.
- [15] Mitchell, T. "Machine learning". Mc graw hill. 1997.
- [16] Bishop, C.M. "Neural Networks for Pattern Recognition". Clarendon Press, Oxford. 1995.
- [17] Freeman, J.A. Skapura, D.M. "Neural Networks: Algorithms, Applications and Programming Techniques". Reading. Addison wesley. 1991.
- [18] Haykin, S. "Neural Networks: a Comprehensive Foundation". NY. Prentice Hall. 1998.
- [19] Hilera, J. R. Martinez, V. J. "Redes Neuronales Artificiales. Fundamentos, Modelos y Aplicaciones". Alfaomega. 2000.
- [20] Goldberg, D.E. "Genetic Algorithms in Search, Optimization, and Machine Learning". Addison Wesley. 2007.
- [21] Clarke, K. O'moore, R. Smeets, R. Talmon, J. Brender, J. McNair, P. Nykanen, P. Grimson, J. Barber B. "A Methodology for Evaluation of knowledge-based Systems in Medicine". Artificial Intelligence in Medicine. Vol 6(2). Pag 107-121. 1994.
- [22] Leung, D. Romagnoli, J. "Dynamic Probabilistic model-based Expert System for Fault Diagnosis". Computers & Chemical Engineering. Vol 24(11). Pag 2473-2492. 2000.
- [23] Shu-hsien Liao. "Expert System Methodologies and Applications - A Decade Review from 1995 to 2004". Expert Systems with applications. Vol 28(1). Pag 93-103. 2005.



# Framework para Toma de Decisiones Multi-Experto Multi-Granular con Información Lingüística

Gramajo Sergio, Gimenez Antonio Manuel

Grupo de Investigación Sobre Inteligencia Artificial, Universidad Tecnológica Nacional Facultad Regional Resistencia, French 414, (3500) Chaco, Argentina. Tel. 362-4432683

{sergiogramajo,manuelego1}@gmail.com

## Resumen

En términos generales, el proceso de Toma de Decisión (TD) está presente en diversas actividades humanas y consiste en elegir la/s mejor/es alternativa/s entre un conjunto de alternativas posibles en pos de alcanzar un/os objetivo/s que satisfaga/n la resolución de un problema. Cuando la información que se maneja para dicho proceso de TD es incierta y cuando en él participan múltiples expertos, el proceso se vuelve aún más complejo. En este sentido, existen diversos modelos y propuestas que ayudan a gestionar este tipo de problemas. En este proyecto se propone la construcción de un framework que permita ayudar en el proceso de TD con incertidumbre de una manera flexible y genérica y que además puedan utilizarlo múltiples expertos (ME) simultáneamente obteniendo una solución grupal para cada problema. Además se evalúan diferentes modelos de información incierta con diferentes operadores de agregación para mejorar el proceso de TD ME.

**Palabras clave:** Toma de decisión multi-experto multi-granular, Información Lingüística, Framework de decisión.

## Contexto

Este proyecto está inserto en una línea de I/D presentada en la Universidad Tecnológica Nacional con código: UTN-1315. Título: “*Diseño de Técnicas para el Tratamiento de Situaciones de Incertidumbre en Sistemas de Soporte de Decisiones con Múltiples Expertos*”. Dicho proyecto se lleva a cabo en el ámbito del GISIA perteneciente a la Facultad Regional Resistencia.

## Introducción

A modo general, un proceso de TD, intenta ayudarnos a tomar decisiones desarrollando métodos y/o modelos que permitan representar fielmente cada problema. A veces el proceso de resolución de un problema de decisión es muy sencillo y puede realizarse mediante el uso de métodos algorítmicos. Sin embargo, muchos problemas de decisión no se pueden resolver de esta manera debido a la propia dificultad y complejidad del problema. Esto puede deberse a que las decisiones están relacionadas con los cambios del entorno o la existencia de vaguedad e incertidumbre en el propio dominio del problema. En este sentido, la teoría clásica de decisión ofrece modelos probabilísticos para gestionar la incertidumbre en los problemas de decisión pero, en mu-

chos de ellos, es fácil observar que algunos de sus aspectos tienen un carácter no probabilístico ya que están relacionados con la imprecisión y la vaguedad propia de la información con la que tratan.

Este proyecto se enfoca en procesos de TD en ambientes de incertidumbre o cuando la información manejada puede ser incompleta, vaga o imprecisa. Esto implica que la utilidad asignada a cada alternativa para resolver un problema tenga que ser valorada de forma aproximada. Este es el ámbito que más se aproxima a las decisiones de la vida real y del comportamiento humano, ya que la imprecisión es inherente a los individuos. Precisamente estos problemas pueden ser resueltos utilizando el Enfoque Lingüístico Difuso (ELD) [1] derivado de la Teoría de Conjuntos Difusos [2] que utiliza palabras o frases del lenguaje natural como variables (Información Lingüística, IL). Con el uso de IL se emplean valoraciones con palabras del lenguaje natural en lugar de hacerlo con números, de este modo la información disponible se representa mediante una aproximación lingüística que resulta más natural para evaluar que una numérica. Hay varias razones por las cuales los métodos lingüísticos son apropiados. En primer lugar, debido a la naturaleza del problema, puede resultar más sencillo usar palabras antes que números. Segundo, las personas pueden sentirse más cómodas y seguras evaluando con palabras a través de términos conocidos que tratando directamente con números. Tercero, en muchos casos la información para evaluaciones numéricas no está disponible o el costo de su obtención es muy elevado [1].

Por otro lado, en el proceso de resolver un problema de decisión existe la posibilidad de incluir a múltiples expertos lo que puede contribuir a su solución ya que ellos pueden tener diferentes apreciaciones y enriquecer el resultado final. Además se puede incluir a expertos con distintos grados de experiencia y conocimiento que des-

de del punto de vista de flexibilidad del framework, su participación puede ser importante y necesaria.

Para poder manejar la complejidad, la incertidumbre y necesidad de ME inherentes en este tipo de problemas, en este proyecto se propone un framework que gestione IL Multi-granular [3][4][5].

La utilización de estas herramientas en procesos de TD han sido utilizadas en diversos campos como Economía, Matemática, Inteligencia Artificial, Aplicaciones Industriales, Servicios de Internet, Gestión de Recursos Humanos, entre otros [6][7][8][9][10][11].

## Líneas de investigación y desarrollo

En el proyecto “Diseño de Técnicas para el Tratamiento de Situaciones de Incertidumbre en Sistemas de Soporte de Decisiones con Múltiples Expertos” se diseñó e implementó un sistema de soporte a las decisiones con Información Lingüística [7][12] y el Proceso Analítico Jerárquico (AHP) [7]. Este sistema permite que múltiples expertos puedan participar conjuntamente en proyectos que ayuden a la toma de decisiones, cuya información se procesa utilizando las técnicas anteriormente mencionadas. Cada uno de estos proyectos tiene un conjunto de alternativas y un conjunto de criterios que el experto deberá valorar en base a su experticia. El procesamiento y obtención de la conclusión del sistema puede realizarse para un experto en particular y para todos los expertos que participen en el proyecto. En el procesamiento se omite la información de aquellos expertos si no se encuentra completa y suficientemente consistente

## Resultados y objetivos

Para la construcción del framework inicialmente fue necesario modelar los diferentes aspectos de los problemas de TD que abordamos en este proyecto, como ser: información cualitativa del problema a través de valores lingüísticos o IL, manejo de múltiples expertos, modelado de la información lingüística multi-granular (para incluir expertos con distintos grados de conocimiento o experiencia) y debido a que se maneja IL es necesario realizar procesos de Computación con Palabras (CW) [7][8][13][14] para obtener los resultados finales.

Debido a diferentes ventajas respecto a otros métodos para representar IL, en este trabajo se emplea el Modelo de Representación Lingüística Difusa 2-tupla (MRLD2T) [8] que permite trabajar con múltiples expertos en dominios multi-granulares sin pérdida de información haciendo uso de herramientas como Jerarquías Lingüísticas Extendidas (ELH) [15].

Para implementar las características mencionadas fue necesario definir un modelo que permita satisfacer los requisitos mencionados con la mayor flexibilidad, eficiencia y simplicidad posible. Por lo que se plantea un esquema teórico genérico de decisión dividido en 6 etapas:

1. En principio se debe identificar el objetivo del problema.
2. Identificar los objetos que forman parte del proceso como el conjunto de alternativas para resolver el objetivo, los

critérios de las alternativas y los expertos que participan del proceso de decisión.

3. Se identifica el contexto de la decisión. Es decir, los términos lingüísticos que van a utilizar los expertos y la cantidad necesaria para cada experto según su conocimiento o experiencia.

Por ejemplo, un conjunto de 7 términos lingüísticos podría estar dado por el conjunto  $S=\{s_0=\text{extremadamente bajo}, s_1=\text{muy bajo}, s_2=\text{bajo}, s_3=\text{medio}, s_4=\text{alto}, s_5=\text{muy alto}, s_6=\text{extremadamente alto}\}$ . Expertos con mayor conocimiento podrían utilizar un conjunto de 9 términos y expertos con menor conocimiento podrían utilizar 5 términos, integrando así el conjunto de términos lingüísticos multi-granulares. Este concepto también está asociado con la capacidad humana de valorar diferentes elementos [16]. Cada CTL tiene asociada una sintaxis definida por la lingüística de los términos utilizados y una semántica definida por funciones de pertenencia triangulares (ver Figura 1).

4. Se realizan las valoraciones por parte de cada uno de los expertos que participa del proceso de decisión.

5. Una vez realizadas las valoraciones el sistema realiza los cálculos computacionales para obtener los resultados finales. Para ello esta etapa tiene dos pasos:

- (i) Unificación de la información lingüística multi-granular. Este se debe a que los expertos han expresado sus valoraciones en diferentes conjuntos de etiquetas lingüísticas.

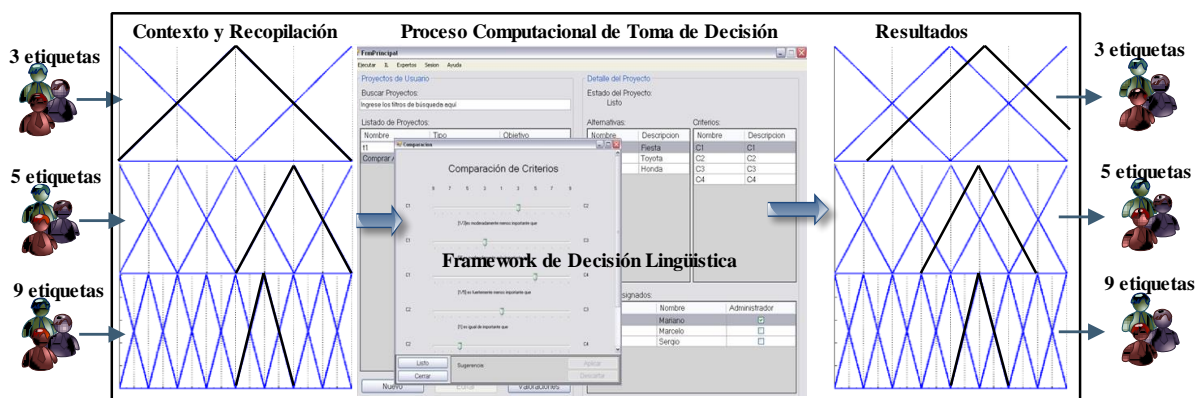


Fig. 1. Esquema del Framework de Decisión

(ii) Agregación de la información unificada en el paso previo. Es significa, obtener un valor global para cada objeto valorado utilizando diferentes operadores para ello.

6. Una vez agregada la información, los resultados finales y globales son expresados para cada experto.

El funcionamiento para la TD empleando el framework que soporte las etapas teóricas mencionadas puede verse en la Figura 1 e implica los siguientes pasos:

a) Creación del Proyecto. Aquí se crea el proyecto con el objetivo de la decisión (etapa 1).

b) Creación del contexto de la decisión. En este paso se crean las alternativas que son posibles soluciones del problema de decisión. Además los criterios que se tendrán en cuenta para las valoraciones y por otro lado, se asignan los expertos que van a participar del proyecto y se les asigna a cada uno de ellos el conjunto de etiquetas que van a utilizar en el proceso de decisión (etapas 2 y 3).

c) Recopilación de la información. En este paso cada experto valora los criterios en cada alternativa con su conjunto de términos (etapa 4).

d) Proceso computacional de TD. Este paso es transparente para los expertos y consiste en realizar los cálculos con palabras para obtener una utilidad por cada alternativa de forma global e individual (etapa 5 (i) y (ii)).

e) Resultados. Los resultados finales son mostrados al usuario, de acuerdo a su conocimiento, en un ranking de alternativas ordenadas en una escala descendente (etapa 6). Para ello se emplearon diferentes operadores de agregación basados en media y operadores ponderados. Esto permite estudiar los resultados de forma analítica para cada decisión y elegir la alternativa que solucione el problema que mejor se adapte a los objetivos del proyecto creado con el framework.

## Formación de Recursos Humanos

La formación de recursos humanos es la siguiente:

### Formación de becarios:

Los becarios Giménez Antonio Manuel y Wanderer Leonardo Simón son alumnos avanzados de la carrera de ingeniería en sistemas de información que han sido beneficiados con una beca de finalización de carrera. De esta manera es posible fomentar la actividad de investigación en alumnos que están próximos a recibirse y estimularlos a que la adopten como actividad profesional.

### Formación de postgrado:

A partir de las líneas de investigación desarrolladas en el proyecto se han logrado finalizar dos tesis doctorales. Los ingenieros Gramajo Sergio y Karanik Marcelo han finalizado su tesis doctoral del programa de Doctorado en Ingeniería de Sistemas y Computación de la Universidad de Málaga (España) en temáticas de TD relacionadas al actual proyecto.

### Equipo de trabajo:

La estructura del equipo de trabajo es la siguiente:

- Director del Proyecto: Dr. Marcelo Karanik.
- Docente-Investigador: Dr. Sergio Gramajo.
- Alumnos becados:
  - Giménez Antonio Manuel
  - Wanderer Leonardo Simón.



## Referencias

1. Zadeh, L.: The concept of a linguistic variable and its application to approximate reasoning. Part I, Part II and Part III. *Information Sciences* 199–249,301–357,143–180 (1975)
2. Zadeh, L.A.: Fuzzy sets. *Information and Control* 8, 338-353 (1965)
3. Herrera, F., Martínez, L.: A Model Based on Linguistic 2-Tuples for Dealing with Multigranular Hierarchical Linguistic Contexts in Multi-Expert Decision-Making. *IEEE Transactions on Systems, Man and Cybernetics -Part B: Cybernetics* 31, 227-234 (2001)
4. Herrera, F., Herrera-Viedma, E., Martínez, L.: A Fusion Approach for Managing Multi-Granularity Linguistic Term Sets in Decision Making", *Fuzzy Sets and Systems* 114, 43-58 (2000)
5. Herrera, F., Herrera-Viedma, E.: Linguistic decision analysis: Steps for solving decision problems under linguistic information. *Fuzzy Sets and Systems* 115, 67-82 (2000)
6. Power, Sharda, D.J., Braden, R., Shenker, D.C.a.S.: *Decision support systems*. Springer Handbook of Automation (2009)
7. Martínez, L., Ruan, D., Herrera, F.: Computing with Words in Decision support Systems: An overview on Models and Applications. *International Journal of Computational Intelligence Systems (IJCIS)* 3, 382-395 (2010)
8. Herrera, F., Martínez, L.: A 2-tuple Fuzzy Linguistic Representation Model for Computing with Words. *IEEE Transactions on Fuzzy Systems* 8, 746-752 (2000)
9. Gramajo, S., Martínez, L.: A linguistic decision support model for QoS priorities in networking. *Knowledge-Based Systems, New Trends on Intelligent Decision Support Systems* 32, 65-75 (2012)
10. Martínez, L.: Sensory Evaluation Based on Linguistic Decision Analysis. *International Journal of Approximated Reasoning* 44, 148-164 (2007)
11. Andrés, R.d., García-Lapresta, J.L.: A multi-granular linguistic model for management decision-making in performance appraisal *Soft Computing* 14, 21-34 (2010)
12. Saaty, T.L.: *How to Make a Decision - The Analytic Hierarchy Process*. *European Journal of Operational Research* 48, 9-26 (1990)
13. Delgado, M., Verdegay, J.L., Vila, M.A.: Aggregation Operations of Linguistic Labels. *International Journal of Intelligent Systems* 2, 351-370 (1993)
14. Yager, R.R.: On ordered weighted averaging aggregation operators in multi-criteria decision making. *IEEE Transaction on Systems, Man, and Cybernetics* 18, 183–190 (1998)
15. Espinilla, M., Liu, J., Martínez, a.L.: An extended hierarchical linguistic model for decision-making problems. *Computational Intelligence* 27, 489-512 (2011)
16. Miller, G.A.: The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Information Processing. *Psychological Review* 63, 81-97 (1956)

# Desarrollo e implementación de prototipos de robots móviles para la navegación autónoma utilizando técnicas de Inteligencia Artificial

Marcelo Marinelli, Monica Mounier, Enrique De Silvestre, Carlos Kornuta, Melissa Kolb, Guillermo Wurn, Lisandro Solonezen.

Departamento de Informática, Facultad de Ciencias Exactas Químicas y Naturales. Universidad Nacional de Misiones

Felix de Azara 1552, Posadas, Misiones

Te:376-4422186

marcelo@fceqyn.unam.edu.ar, monica@moniamounier@gmail.com,  
desilvestre@yahoo.com.ar, c\_kornuta@hotmail.com, meli\_kolb@yahoo.com.ar,  
guillermow77@gmail.com, lsolonezen@gmail.com

## RESUMEN

Se propone el diseño y construcción de prototipos robóticos que incluyan dispositivos de tracción, dirección, telemetría y sistemas embebidos que permitan la investigación de varias técnicas de inteligencia artificial (IA) para la navegación autónoma de robots móviles.

Los prototipos deberán poseer diversos sistemas sensoriales de entorno: sonares, infrarrojo y sistemas de visión artificial.

Las técnicas de IA utilizadas serán lógica difusa y sistemas expertos y se trabajará en los problemas clásicos de la robótica: evasión de obstáculos y seguimiento de entornos.

Palabras clave: Navegación de Robot, Inteligencia Artificial, Simulador de Entorno.

## Contexto

Este proyecto se enmarca en el "Programa de Investigación en Computación" del Centro de Investigación y Desarrollo Tecnológico de la Facultad de Ciencias Exactas

Químicas y Naturales de la Universidad Nacional de Misiones; también vinculado con el Doctorado en Ingeniería de Sistemas y Computación que funciona en la Universidad de Málaga.

Dentro del proyecto se desempeñan docentes, tesis y becarios de las carreras de Analista en Sistemas de Computación, Licenciatura en sistemas de Información y Profesorado en Física

## Introducción

La investigación en robótica permite desarrollar técnicas que pueden ser transferidas a la industria para la fabricación de plataformas móviles de transporte autónomas, robots móviles que realicen tareas automatizadas en el hogar o sistemas que se encarguen de acceder a zonas contaminadas y/o de baja visibilidad. Para cumplir con las funciones antes mencionadas se necesita que el robot posea un sistema inteligente que le permita la toma de decisiones para evaluar la trayectoria a seguir cuando se encuentre con un obstáculo o deba seguir un entorno.

Otra aplicación de los prototipos robóticos es la utilización como material

didáctico en el área de ciencias del nivel medio. La enseñanza de la robótica desarrolla la lógica y permite mejorar el proceso de enseñanza y aprendizaje de la programación. Para esto se dota a los prototipos de robots con un intérprete sencillo, que permite a los estudiantes desarrollar programas que controlen la trayectoria de los prototipos de robots móviles.

Una de las técnicas de IA que se está aplicando es la que utiliza controladores difusos combinados con técnicas de generación automática de conjunto de datos [1], [2], [3]. Para generar las reglas difusas se utiliza el algoritmo Wang and Mendel [4]. El conjunto de datos se generan mediante un algoritmo que combina el universo de discurso de todas las variables, y mediante la aplicación de las ecuaciones cinemáticas del movimiento del robot se evalúa la mejor solución mediante una función de puntuación.

Una vez obtenidas las reglas, el módulo del controlador difuso se conecta con el simulador player/stage para verificar el comportamiento.

## Líneas de investigación y desarrollo

El grupo de robótica del Programa de Investigación en Computación está trabajando en distintas líneas de investigación y desarrollo relacionadas con la robótica y la I.A., a continuación se detallan las más importantes:

- Controladores difusos aplicados al proceso de la elaboración de yerba mate.
- Sistemas de control de navegación para robots utilizando arreglos de sensores de ultra sonido.

- Aplicación de controladores difusos desarrollados con la herramienta FIS de Matlab.
- Telemetría con tecnología bluetooth para el control de navegación de robot.
- Desarrollo de aplicaciones con Java móvil y tecnología bluetooth, con el objeto de usar los dispositivos móviles para el control de robots domésticos.
- Desarrollos de prototipos de robots móviles para la enseñanza de la robótica y el aprendizaje de la programación en el nivel medio.
- Desarrollo de protocolos de control de robots mediante un intérprete para placas de hardware libre tipo Arduino [5], [6].

## Resultados y Objetivos

En el marco de este proyecto se realizó la publicación "Marinelli, Marcelo J., De Silvestre, Enrique C. "Desarrollo de un prototipo de robot móvil para la investigación y aplicación de técnicas de inteligencia artificial". Rev. de Ciencia y tecnología, F.C.E.Q. y N. Año 14, N° 17, pag.18 a 12. año 2012 [7].



Fig. 1. Prototipo de robot móvil Megabot 5S

En el área de desarrollo se está trabajando en el diseño e implementación de dispositivo de censado para los prototipos de robots desarrollados como el que se aprecia en la Fig. 1. Algunos de estos trabajos son:

a) Dispositivo de seguimiento de línea basado en el circuito de la Fig. 2, que permite que el prototipo robótico siga una línea blanca sobre un fondo oscuro.

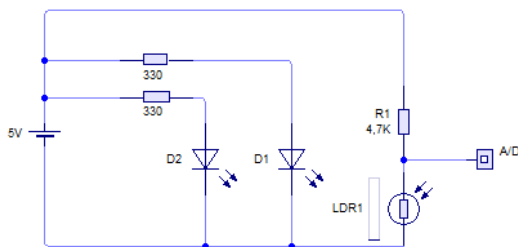


Fig. 2. Circuito de seguimiento de línea

b) Sistema de guía basado en el módulo compás, magnetómetro de tres ejes basado en un chip Honeywell HMC5883L Fig. 3.

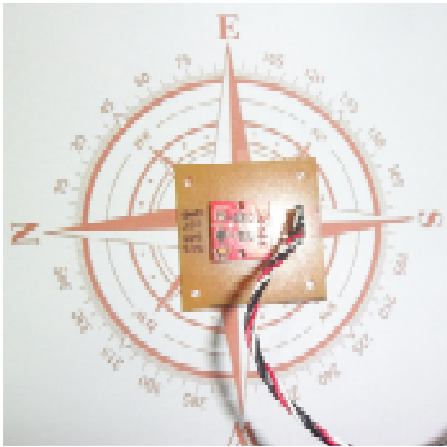


Fig. 3. Magnetómetro de tres ejes.

c) Sistema de control con módulo acelerómetro de tres ejes basado en el chip MMA7660FC. Este dispositivo se lo está ensayando para realizar un prototipo de dos ruedas de tracción y un tronco

vertical que debe mantener el equilibrio durante el movimiento.

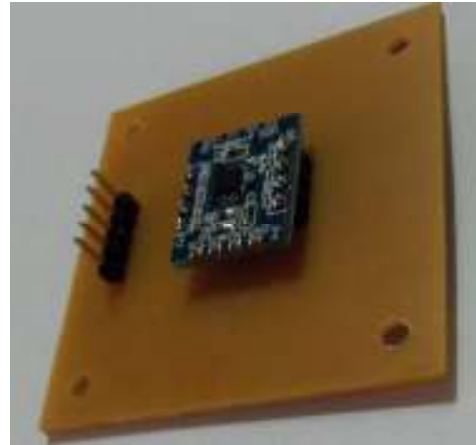


Fig. 4. Acelerómetro de tres ejes

## Formación de Recursos Humanos

En el programa hay tres doctorandos, uno de ellos con beca tipo I de CONICET y los otros dos poseen el D.E.A. aprobado del Doctorado en Ingeniería de Sistemas y Computación de la UMA, Málaga y están en etapa de desarrollo de tesis

Dos graduados investigadores y cinco tesistas de grado de la Licenciatura en Sistemas de Información de la F.C.E.Q. y N. de la Universidad Nacional de Misiones.

## Referencias

- [1] M. Mucientes, R. Iglesias, C. V. Regueiro, A. Bugarini, and S. Barro, "A fuzzy temporal rule-based velocity controller for mobile robotics". *Fuzzy Sets Syst.*, vol. 134, pp. 83–99, 2003.
- [2] B. del Brio, A. Molina "Redes Neuronales y Sistemas Borrosos". W. Pedrycz, Ed. Norwell, Alfaomega, pp. 289–306, 2007.
- [3] A. Bonarini, "Evolutionary learning of fuzzy rules: Competition and



cooperation". in Fuzzy Modelling: Paradigms and Practice, W. Pedrycz, Ed. Norwell, MA: Kluwer Academic, pp. 265–284, 1996.

[4] Wang, Li-Xin, Mendel, Jerry M. "Generating fuzzy rules by learning from examples" Systems, Man and Cybernetics, IEEE Transactions on Volume: 22, Issue: 6, Page(s): 1414 - 1427, 1992.

[5] Arduino. Plataforma de hardware libre. Disponible en <http://www.arduino.cc> (verificado en Marzo de 2013)

[6] Arduino. Software de Arduino. Disponible en <http://arduino.cc/es/Main/Software> (verificado en Marzo de 2013)

[7] Marinelli, Marcelo J., De Silvestre, Enrique C. "Desarrollo de un prototipo de robot móvil para la investigación y aplicación de técnicas de inteligencia artificial". Rev. de Ciencia y tecnología, F.C.E.Q. y N. Año 14, N° 17, pag.18 a 12. año 2012.

## DISEÑO E IMPLEMENTACIÓN DE UN SISTEMA MULTISENSORIAL EN UN ROBOT MÓVIL AUTÓNOMO CONTROLADO MEDIANTE UN SISTEMA EXPERTO BASADO EN REGLAS Y EVALUACIÓN DE LA NAVEGACIÓN EN ENTORNOS SEMIESTRUCTURADOS

**Marcelo Marinelli, Enrique De Silvestre, Carlos Kornuta, Mónica Mounier, Melissa Kolb**

Departamento de Informática, Facultad de Ciencias Exactas Químicas y Naturales. Universidad Nacional de Misiones

Felix de Azara 1552, Posadas, Misiones

Te: 376-4422186

[marcelomarinelli@fceqyn.unam.edu.ar](mailto:marcelomarinelli@fceqyn.unam.edu.ar), [desilvestre@yahoo.com.ar](mailto:desilvestre@yahoo.com.ar), [c\\_kornuta@hotmail.com](mailto:c_kornuta@hotmail.com), [monicamounier@gmail.com](mailto:monicamounier@gmail.com),  
[meli\\_kolb@yahoo.com.ar](mailto:meli_kolb@yahoo.com.ar)

### RESUMEN

El presente trabajo se basó en un prototipo de robot móvil desarrollado por este equipo de investigación, el cual estaba compuesto por un sistema sensorial formado por cinco sonares. Dicho prototipo es utilizado para la evaluación de eficiencia del sistema de control aplicando diferentes técnicas de Inteligencia Artificial. Con la intención de evaluar las mejoras en la precisión del robot en la navegación, se ha diseñado e implementado, un sistema multisensorial compuesto de una brújula electrónica y un acelerómetro además del ya mencionado sistema sonar, ampliando de ésta manera, las posibilidades de navegación del robot a entornos semiestructurados y sobre plataformas inclinadas. El desafío más interesante consistió en reprogramar el motor de inferencias del Sistema Experto basado en reglas utilizado para el control del robot, considerando las nuevas variables que aporta el sistema multisensorial.

**Palabras clave:** robot, sistemas expertos, sensores, navegación autónoma.

### CONTEXTO

Este proyecto se enmarca en el “Programa de Investigación en Computación” del Centro de Investigación y Desarrollo Tecnológico de la Facultad de Ciencias Exactas Químicas y Naturales de la Universidad Nacional de Misiones; también vinculado con el Doctorado en Ingeniería de Sistemas y Computación que funciona en la Universidad de Málaga.

Dentro del proyecto se desempeñan docentes, tesis y becarios de las carreras de Analista en Sistemas de Computación, Licenciatura en sistemas de Información y Profesorado en Física

### INTRODUCCION

La Asociación Japonesa de Robótica Industrial (JIRA) ha definido un robot inteligente como: *un robot con la capacidad de entender su medio ambiente y con la habilidad de completar con éxito una tarea a pesar de cambios en las condiciones de los alrededores bajo las cuales fue diseñado*, sin embargo las investigaciones de la robótica continúan su desarrollo para incluir procesos de aprendizaje, con el fin de que los robots además de adaptarse a los cambios de su medio ambiente, también logren aprender de sus propias experiencias.[6].

Con el agregado de sensores de distintos tipos se obtuvieron robots multisensoriales, que gracias a complejos sistemas de control permiten que el robot interactúe en mayor grado con el ambiente. Esto se conoce como la cuarta generación de robots. [2], [4], [19]. En la actualidad se encuentra en desarrollo la quinta generación de robots con la integración de componentes que permitan la interacción de módulos de conducta.

Un robot móvil es una máquina integrada por tres componentes: un sistema mecánico, un sistema electrónico y un sistema de procesamiento los cuales en conjunto le brindan capacidad de locomoción, autonomía e interacción. O sea capacidad de desplazarse de forma autónoma en un ambiente desconocido o conocido sólo parcialmente. [1], [20]. Para dotar al robot de capacidad de autonomía, debemos contar con un sistema de sensores que permitan captar el ambiente y, a partir de los datos recogidos de dichos sensores, tomar las decisiones adecuadas para un correcto desplazamiento. Para esto último, se hace necesario contar con un sistema de procesamiento de los datos obtenidos de los sensores y con un sistema de toma de decisiones, los cuales formarán parte del sistema de control que generará las señales necesarias para controlar el sistema de locomoción, dando de esta manera al robot móvil la autonomía requerida. [20], [1]. Los componentes del sistema sensorial son muy variados

y dependen del destino que se le asigne al robot, así como del entorno o ambiente en que deba desplazarse. Estos sensores pueden ser simples fines de carrera o bien sensores láser, radar, ultrasónicos o sonares y escáneres compuestos por cámaras de video comunes o infrarrojas, etc. Las señales provenientes del sistema sensorial formarán un flujo de datos a procesar por parte del sistema de procesamiento. Estos datos procesados son evaluados por el sistema de decisión el cual generará los comandos adecuados para controlar el sistema de locomoción. [2]. Toda esta secuencia de pasos: captura de datos de sensores, procesamiento, decisión y control debe hacerse en tiempo real y, la transmisión de datos entre subsistemas requerirá algún protocolo de comunicación que asegure la correcta transmisión de datos en tiempo y forma. Para este trabajo, la capa de control se ha implementado, con muy buenos resultados, los Sistemas Expertos basados en reglas.

### OBJETIVOS

El objetivo general de este trabajo fue diseñar e implementar un sistema multisensorial compuesto de sonares, brújula electrónica y acelerómetro, sobre un prototipo de robot móvil. Modificar el motor de inferencias del Sistema Experto basado en reglas que controla al robot, considerando las nuevas variables que aporta el sistema multisensorial y que permite la navegación en entornos parcialmente estructurados y sobre plataformas no necesariamente planas. Luego realizar pruebas de navegación en distintos ambientes diseñados para la evaluación en diversas condiciones y analizar estadísticamente los resultados.

### METODOLOGÍA

La metodología utilizada en el presente trabajo consistió en un análisis preliminar de la manera en que interactúan con el entorno los diversos tipos de sensores tanto de percepción interna como de percepción externa. Se recopiló información bibliográfica relevante al estudio y datos de trabajos anteriores. Se obtuvieron los lineamientos de la situación actual y se seleccionaron los sensores de percepción interna complementarios, necesarios para la navegación propuesta. Se procedió al modelado del sistema en estudio, se implementó el sistema multisensorial. Se reprogramó el motor de inferencias del sistema experto basado en reglas. Se hicieron pruebas de navegación.

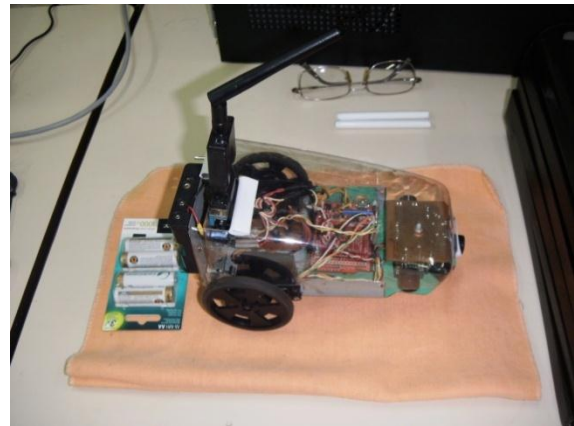


Fig.1 Prototipo inicial con tracción diferencial y un arreglo de cinco sonares y conexión bluetooth

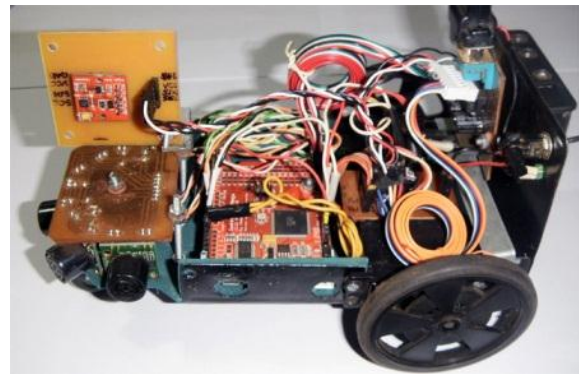


Fig.2 Prototipo Modificado un arreglo de cinco sonares y sensores de percepción interna.

### DESARROLLO

En el marco de este proyecto se desarrolló el trabajo **“Implementación de un Sistema Multisensorial en un Robot Móvil Autónomo controlado mediante un Sistema Experto basado en reglas y Evaluación de la Navegación en Entornos Semiestructurados”**. El cual consistió en las siguientes fases:

- 1.- El diseño de un sistema sensorial compuesto de un arreglo de cinco sensores de ultrasonido, y dos sensores de percepción interna, una brújula electrónica y un acelerómetro.
- 2.- Implementación del sistema sensorial en el prototipo de desarrollo.
- 3.- El desarrollo del motor de inferencias de un sistema experto basado en reglas y estratificado, con una capa de control y otra capa deliberativa que planifique la navegación a mediano plazo.
- 4.- Implementación de una conexión mediante bluetooth que permita tener la capa de control de software de más alto nivel en una computadora remota y capturar las lecturas de los sensores en tiempo real para un análisis estadístico de las mismas.

## RESULTADOS OBTENIDOS

Durante las pruebas realizadas sobre el sistema sensorial en forma aislada se obtuvieron resultados adecuados al objetivo con precisiones de cinco pulgadas para los sensores de percepción externa y tres grados para los de percepción interna. Una vez implementado el sistema sensorial al prototipo, se redujo la eficacia en la percepción interna en aproximadamente diez grados, motivados por los ruidos eléctricos del sistema de alimentación del robot. Sin embargo dicha pérdida de eficiencia no influye significativamente en el control del robot para las trayectorias propuestas inicialmente, a las que está destinado el robot. Se continúan las investigaciones en pro de mejorar la eficiencia de percepción interna aislando los ruidos eléctricos y, predisponiendo al robot, a la navegación en trayectorias más complejas.

Se concluye que: El desarrollo del sistema multisensorial implementado robot móvil autónomo y la programación de un sistema experto basado en reglas, satisface aceptablemente y a muy bajo costo, la navegación y la evasión de obstáculos en entornos parcialmente estructurados.

## LÍNEAS DE INVESTIGACION Y DESARROLLO

Dentro de las líneas de investigación que contienen el "Programa de Investigación en Computación", este proyecto se enmarca en el área de inteligencia artificial en donde se aplican técnicas de lógica difusa, algoritmos bioinspirados, sistemas expertos, etc., estas técnicas se utilizarán en los siguientes trabajos:

- Controladores difusos aplicados al proceso de la elaboración de yerba mate.
- Sistemas de control de navegación para robots utilizando arreglos de sensores de ultra sonido.
- Aplicación de controladores difusos desarrollados con la herramienta FIS de Matlab.
- Telemetría con tecnología bluetooth para el control de navegación de robot.
- Desarrollo de aplicaciones con Java móvil y tecnología bluetooth, con el objeto de usar los dispositivos móviles para el control de robots domésticos.
- Desarrollos de prototipos de robots móviles para la enseñanza de la robótica y el aprendizaje de la programación en el nivel medio.
- Desarrollo de protocolos de control de robots mediante un intérprete para placas de hardware libre tipo Arduino.

## FORMACIÓN DE RECURSOS HUMANOS

En esta línea de investigación se prevé desarrollar cinco Tesis de grado de la carrera de Licenciatura en Sistemas de Información de la F.C.E.Q y N. de la U.Na.M., dos Tesis doctorales del Doctorado en Ingeniería de Sistemas y Computación del Departamento de Lenguajes y Ciencias de la Computación de la Universidad de Málaga y uno de los integrantes obtuvo una beca tipo I de CONICET, realizando su doctorado en la UNICEM.

## BIBLIOGRAFIA

- [1] Baturone Anibal Ollero "Robótica Manipuladores y robots móviles" [Libro]. - [s.l.] : Alfaomega & Marcombo, 2001.
- [2] Brooks R. "A robust Layered Control System for a Mobile Robot" [Libro]. - [s.l.] : "IEEE Journal of Robotics and Automation", 1989.
- [3] Brox M. Gersnoviez A., Sanchez-Solano S., Baturone I. "Controlador Difuso Para Problemas De Navegación En Presencia De Obstaculos Fijos" [Informe]. - España : XIII Congreso Español de Tecnologías y Lógica Fuzzy , 2006.
- [4] Gat. Erann On three – "layer Architectures - Artificial intelligent and mobile robots" [Libro]. - [s.l.] : In R.P. Bonnaso D. Kortenkamp and R. Murphy, Editors, MIT Press, Cambridge, MA, 1998.
- [5] Kuc R. "A Spatial Sampling Criterion for Sonar Obstacle Detection" [Informe]. - [s.l.] : IEEE Transactions on Pattern Analysis and Machine Intelligence, 1990.
- [6] Ortiz Rojas Jaime "Diseño y realización de Maxcota, un robot inteligente para niños" [Informe]. - 2004.
- [7] R.Kuc & M. W. Siegel "Physically Based Simulation Model for Acoustic Sensor Robot Navigation" [Libro]. - [s.l.] : IEEE Transactions on Pattern Analysis and Machine Intelligence, 1987.
- [8] M.Mucientes, J. Casillas. "Quick Design of Fuzzy Controllers With Good Interpretability in Mobile Robotics". IEEE TRANSACTIONS ON FUZZY SYSTEMS, VOL. 15, NO. 4, pp. 636-651, 2007.
- [9] S. Russell, P. Norvig, "Inteligencia Artificial Un Enfoque Moderno". Madrid, Pearson Educacion S.A. 2004.



- [10] M. Mucientes, R. Iglesias, C. V. Regueiro, A. Bugarini, and S. Barro, "A fuzzy temporal rule-based velocity controller for mobile robotics". *Fuzzy Sets Syst.*, vol. 134, pp. 83–99, 2003.
- [11] B. del Brio, A. Molina "Redes Neuronales y Sistemas Borrosos". W. Pedrycz, Ed. Norwell, Alfaomega, pp. 289–306, 2007.
- [12] A. Bonarini, "Evolutionary learning of fuzzy rules: Competition and cooperation". in *Fuzzy Modelling: Paradigms and Practice*, W. Pedrycz, Ed. Norwell, MA: Kluwer Academic, pp. 265–284, 1996.
- [13] B. Gerkey, R. Vaughan, K. Stoy, A. Howard, G. Sukhatme, M. Mataric. "Most Valuable Player: A Robot Device Server for Distributed Control". Robotics Research Labs, University of Southern California Los Angeles, CA 90089-0721, USA.
- [14] J. Urzelai, J. P. Uribe, and M. Ezkerra, "Fuzzy controller for wall following with a non-holonomous mobile robot". *Proc. 6th IEEE Int. Conf. Fuzzy Syst. (Fuzz-IEEE'97)*, Barcelona, Spain, pp. 1361–1368, 1997.34
- [15] J. Casillas, O. Cordon, and F. Herrera, "COR: A methodology to improve ad hoc data-driven linguistic rule learning methods by inducing cooperation among rules," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 32, no. 4, pp. 526–537, 2002.
- [16] J. Casillas, O. Cordon, "COR methodology: A simple way to obtain linguistic fuzzy models with good interpretability and accuracy," in *Accuracy Improvements in Linguistic Fuzzy Modeling*, J. Casillas, O. Cordon, F. Herrera, and L. Magdalena, Eds. Heidelberg, Germany: Springer, 2003.
- [17] M. Dorigo and T. Stutzle, *Ant Colony Optimization*. Cambridge, MA: MIT Press, 2004.
- [18] O. Cordon, F. Herrera, I. F. de Viana, and L. Moreno, "A new AGO model integrating evolutionary computation concepts: The best-worst ant system," in *Proc. 2nd Int. Workshop Ant Algorithms*, Brussels, Belgium, pp. 22–29, 2000.
- [19] Balch, T. And Arkin, R.C. *Communication in Reactive Multiagent Robotic Systems*, Autonomous Robots, Vol.1 [Libro]. - 1994.
- [20] Craig John J. *Introduccion a la Robotica* 3 Ed. [Libro]. - [s.l.] : Prentice Hall, 2006.

# Estrategias de formación de grupo en videojuegos RTS con ambientes dinámicos

Daniel Trevisani

Laura Cecchi

email: {daniel.cipo, lcecchi}@gmail.com

*Grupo de Investigación en Lenguajes e Inteligencia Artificial*

Departamento de Teoría de la Computación

Facultad de Informática

UNIVERSIDAD NACIONAL DEL COMAHUE

Buenos Aires 1400 - (8300)Neuquén - Argentina

## Resumen

La relación entre los videojuegos y la Inteligencia Artificial(IA) se inició, de una manera básica, en la década de los 70. En la actualidad, continúa en forma de avanzadas simulaciones 3D presentes en las diversas sagas de videojuegos que, verdaderamente conforman una industria del entretenimiento. El género de videojuegos que mayor éxito comercial alcanzó es el de estrategia en tiempo real (RTS-Real Time Strategy). La efectividad de los videojuegos depende de la capacidad de cooperación y reacción, ante el oponente, por parte de los personajes del juego. En muchos videojuegos, esto es implementado a través de las llamadas formaciones. El propósito de este trabajo es presentar la descripción de la línea de investigación sobre el estudio de las formaciones de grupo y su implementación en videojuegos RTS. Asimismo, se presentan los resultados alcanzados hasta el momento y los trabajos en progreso y futuros.

**Palabras Clave:** REAL TIME STRATEGY, SISTEMAS MULTIAGENTES, BWAPI, STARCRAFT, SCRIPTING

## Contexto

Este trabajo está parcialmente financiado por la Universidad Nacional del Comahue, en el contexto del proyecto de investigación Agentes Inteligentes en Ambientes Dinámicos. El proyecto de investigación tiene prevista una duración de cuatro años, ha comenzado en enero del 2013 y finaliza en diciembre de 2016.

## 1. Introducción

La Inteligencia Artificial(IA) ha estado presente en los videojuegos desde la década de los 70, aunque inicialmente solo se la expresaba a través de reglas simples que determinaban el comportamiento de los personajes controlados por la computadora.

A medida que pasaron las décadas, los videojuegos evolucionaron ampliamente tanto a nivel visual como en la jugabilidad(gameplay), esto es, la forma en la que los jugadores interactúan con el juego. En los años 90, la aparición de los sistemas operativos con interfaces gráficas de usuario más atractivas, las mejoras en las tecnología de

gráficos 3D y en la velocidad de procesamiento, el auge de Internet, la continua reducción en las dimensiones del hardware presente en las computadoras y en los teléfonos celulares, produjeron un gran salto cualitativo en la tecnología de los videojuegos. Estos avances tecnológicos permitieron el desarrollo de sagas de videojuegos que perduran hasta el momento.

En la actualidad, los videojuegos son de gran importancia dentro de la industria del entretenimiento. Sin embargo, su aplicación ha ido más allá del divertimento, encontrando utilidad en otros campos como el entrenamiento militar y/o deportivo, educación y en general ludificación (gamification) [9; 10; 11; 16]. Existen características comunes en los juegos, como las severas restricciones de tiempo y la fuerte demanda de IA en tiempo real, la cual debe ser capaz de resolver tareas de decisión de manera rápida y satisfactoria. Los juegos de estrategia en tiempo real (RTS- Real Time Strategy) ofrecen una gran variedad de problemas de investigación fundamentales para la IA, entre los que podemos citar[4]:

- **Gestión de recursos.** Adquisición de recursos para construcción de instalaciones y conformar ejércitos.
- **Toma de decisiones bajo incertidumbre.** Inicialmente los jugadores no conocen, por ejemplo, la ubicación de las bases enemigas.
- **Razonamiento temporal y espacial.** Análisis estático y dinámico del terreno, así como el entendimiento de las relaciones temporales de las acciones.
- **Colaboración.** Grupos de jugadores pueden unir fuerzas e inteligencia. El desafío es cómo coordinar de manera efectiva las acciones a través de la comunicación entre las partes.
- **Modelo del oponente.** Descubrir y explotar las debilidades de los oponentes.

- **Planificación adversaria en tiempo real.** Jugar un videojuego RTS es desafiante, ya que su ambiente es altamente dinámico, hostil e inteligente. Es más desafiante aún crear un sistema autónomo de tiempo real capaz de superar la performance de un jugador humano en este campo.

Por lo anteriormente mencionado, resulta desafiante investigar sobre la implementación de sistemas multiagentes que permitan mejorar el comportamiento emergente del juego a través de técnicas de IA. Esto es de gran importancia, ya que muchos jugadores eligen qué videojuegos comprar o jugar en base a esta característica, debido, quizás, a que el nivel de calidad de la IA presente en el juego determina, en gran medida, el grado en que el jugador puede involucrarse en el juego[13].

La efectividad de los videojuegos depende de la capacidad de cooperación y reacción, ante el oponente, por parte de los personajes del juego. En muchos videojuegos, esto es implementado a través de las llamadas *formaciones* [15]. La gestión de unidades de combate tiene el propósito de maximizar el daño causado al enemigo y minimizar el propio. Las formaciones de grupo pueden proveer estas capacidades a través de movimientos cohesivos de grupo[15].

La línea de investigación presentada en este trabajo se centra en el estudio, análisis e implementación de formaciones de grupo en ambientes dinámicos en videojuegos RTS. El estudio de este aspecto permitirá gestionar, de una manera coordinada, los grupos de unidades (llamados escuadrones) presentes en el juego. Como plataforma de implementación y testeo se utilizará el videojuego RTS StarCraft: Brood War [6].

El resto de este trabajo está estructurado como sigue. En la sección 2 se describen los resultados preliminares. Luego, en la sección 3 se detallan las tareas en progreso. Finalmente, en la última sección se presentan las conclusiones y los trabajos futuros.

## 2. Resultados preliminares

El desarrollo de esta línea de investigación comenzó con el estudio de las formaciones de grupo, analizando manuales de táctica y estrategia militar [1; 2]. Entre las formaciones más comúnmente utilizadas en la milicia podemos mencionar: la formación “V”, la formación “filas” o la formación “circular”, entre otras.

Paralelamente, se analizaron los aspectos relevantes al comportamiento de los grupos de unidades [15], entre los que se encuentran la designación de un líder de grupo, la capacidad de movimiento coordinado, la selección inteligente del oponente a atacar y la cooperación en combate de las unidades.

Luego se avanzó en el modelo del juego. Se eligió la metodología Gaia [12] para el análisis y diseño de los agentes. Bajo esta metodología se definieron los esquemas de rol y los modelos de servicios asociados a estos. Los roles surgidos de la etapa de análisis son: “líder de formación”, “observador”, “scout”, “soldado” y “manager táctico”.

Estos roles fueron modelizados teniendo en cuenta que nuestro objetivo final es el comportamiento coordinado de los agentes. En este sentido, la toma de decisión sobre qué formación adoptar de acuerdo a lo observado en el ambiente, la realiza el “manager táctico”, a partir de un árbol de decisión.

## 3. Tareas en Progreso

Actualmente se ha iniciado la etapa de implementación de los roles diseñados. Esta etapa incluye la codificación de los aspectos y características relevantes de las formaciones y grupos de unidades previamente mencionados. Como resultado de esta etapa se espera generar un framework, donde se evaluarán diferentes técnicas de IA para la toma de decisión, sobre qué táctica utilizar, y en función de ella, qué formación es la adecuada para el escuadrón de acuerdo a lo observado en el

ambiente.

El videojuego elegido para testear nuestros desarrollos es StarCraft: Brood War [6]. En este sentido, se tomó como esqueleto para el juego el bot BTHAI [3; 8], desarrollado por Johan Hagelbäck para el videojuego antes mencionado. La interfaz utilizada para la comunicación e interacción con el StarCraft: Brood War [6] es BWAPI [5].

Un aspecto relevante de esta etapa es la separación de la lógica asociada al proceso de decisión sobre la formación a emplear por los grupos de unidades. Esto se logra a través de técnicas de *scripting* usando el lenguaje y la interfaz de programación GameMonkey [14]. De este modo, el proceso de decisión se establece en forma de un árbol de decisión, el cual se compone de los atributos y las decisiones para cada situación de juego descripta. El árbol de decisión está codificado en un archivo XML cuyo aspecto general puede verse en la Fig.1.

---

```
<Test attribute = duplicamos_al_enemigo
      operator="=" value="true">
  <Output decision= "circular"/>
</Test>
```

---

Fig. 1: Formato XML del árbol de decisión

El árbol es traducido a reglas simples de la forma IF-THEN, seguidamente, dichas reglas son incorporadas al código del agente. De esta manera, y de acuerdo a la dinámica del ambiente, es posible extender o modificar el comportamiento del agente sin la necesidad de reconstruir(rebuild) su código.

## 4. Conclusión y Trabajos Futuros

En este trabajo se presentó la descripción de una línea de investigación, sobre el estudio de las formaciones de grupo y su implementación en videojuegos RTS. Como caso de estudio se eligió la plataforma StarCraft.



Se describieron algunos resultados ya alcanzados en cuanto al análisis y diseño de roles y se detallaron las actividades de implementación en progreso.

Entre nuestros trabajos futuros se encuentra la aplicación de aprendizaje automático (machine learning) a fin de poder generar el árbol de decisión a partir del modelo del oponente. Realizar un modelo del oponente puede ser visto como un problema de clasificación, en el cuál se construye dicho modelo, en base a los datos recolectados durante las partidas de juego, para luego ser clasificado en uno de los modelos de oponente disponibles [15].

Asimismo, está entre nuestro trabajo futuro, utilizar al sistema DeLP [7] como mecanismo de toma de decisión entre diferentes tácticas de los escuadrones.

## Referencias

- [1] Army of United States. Field Manual 3-06.11 combined Arms Operation in Urban Terrain, 2002.
- [2] Army of United States. Field Manual 3-21.8 The Infantry Rifle Platoon and Squad, 2007.
- [3] BTHAI. StarCraft Bot using BWAPI. <http://code.google.com/p/bthai/>.
- [4] Michael Buro. Real-Time Strategy Games: A new AI Research Challenge. In *Proceedings of the International Joint Conference on AI*, pages 1534–1535, 2003.
- [5] BWAPI. A Broodwar Application Programming Interface. <http://code.google.com/p/bwapi>.
- [6] Blizzard Entertainment. StarCraft: Brood War, 1998.
- [7] A. García and G. Simari. Defeasible Logic Programming: An Argumentative Approach. *Theory and Practice of Logic Programming*, 4(1):95–138, 2004.
- [8] Johan Hagelbäck. Potential-Field Based Navigation in StarCraft. In *IEEE Conference on Computational Intelligence and Games (CIG)*, 2012.
- [9] Cork Institute of Technology. The Serious Sports Project. <http://serious-sports.org/>.
- [10] Bohemia Interactive. Virtual Battlespace 2, 2007.
- [11] Karl M. Kapp. *The Gamification of Learning and Instruction: Game -based Methods and Strategies for Training and Education*. Pfeiffer, 2012.
- [12] M.Wooldridge, N.Jennings, and D.Kinny. The gaia Methodology for Agent-Oriented Analysis and Design. *Autonomous Agents and Multi-Agent Systems*, 3(3):285–312, 2000.
- [13] Alexander Nareyek. AI in Computer Games. *ACM Queue*, 1(10):58–65, 2004.
- [14] Mathew Riek and Greg Douglas. Game-Monkey Script Reference. <http://www.gmscript.com/>.
- [15] M. van der Heijden, S. Bakkes, and P. Spronck. Dynamic Formations in Real-Time Strategy Games. In *CIG'08 2008 IEEE Symposium on Computational Intelligence in Games*, pages 47–54, 2008.
- [16] Kevin Werbach and Dan Hunter. *For the Win: How Game Thinking can Revolutionize your Business*. Wharton Digital Press, 2012.

# Sistema de Predicción de Incendios Forestales basado en el índice FWI para la Provincia de Córdoba

Marina Cardenas, Juan C. Vázquez, Julio Castillo, Sebastián Villena Ruiz

Laboratorio de Investigación de Software LIS/Dpto. Ingeniería en Sistemas de Información/  
Facultad Regional Córdoba/ Universidad Tecnológica Nacional

{ ing.marinacardenas, jcjvazquez, jotacastillo, sebastian.villena.ruiz }@gmail.com

## Resumen

El objetivo de este proyecto es elaborar un modelo que permita pronosticar los incendios forestales en la Provincia de Córdoba mediante técnicas de aprendizaje automático (machine learning), utilizando para ello modelos supervisados como redes neuronales o maquinas de soporte vectorial.

A partir del mismo, se pretende capturar la presencia de patrones de comportamiento de los factores que influyen en la producción de incendios forestales, ya sean humanos o de índole climática, tales como humedad, presión, temperatura y cantidad de lluvia caída en una zona determinada, asociados a incendios forestales.

**Palabras clave:** *Pronósticos, modelo, incendios, forestales, meteorología, predicción.*

## Contexto

El presente proyecto se encuentra consolidado dentro de la línea de investigación que se dedica a la resolución de problemas físicos, reales y sociales, a través del uso y del empleo de herramientas computacionales basadas en aprendizaje automático. Este proyecto se lleva a cabo en el Laboratorio de Investigación de Software del Dpto. de Ingeniería en Sistemas de Información de la Universidad Tecnológica Nacional Facultad Regional Córdoba, y es un proyecto acreditado por la Secretaría de Ciencia y Tecnología de la UTN.

El mismo, se enmarca dentro del Grupo GIA (Grupo de Inteligencia Artificial) de la UTN-FRC, el cual tiene como objetivo general el investigar técnicas, algoritmos de inteligencia artificial, entre los que se destacan el estudio de las redes neuronales, autómatas celulares,

análisis y procesamiento de imágenes, minería de datos, y su aplicabilidad y resolución de problemas de las ciencias naturales y de las ciencias sociales. El grupo está integrado por doctores, ingenieros, licenciados, becarios, y pasantes.

De esta manera, se puede observar que se investigan técnicas de IA tanto desde el punto de vista teórico, como desde el punto de vista práctico.

## Introducción

Los objetivos del proyecto podrían agruparse en tres tipos: relacionados a las capacidades cognitivas de sus integrantes, aquellos relacionados a la aplicabilidad concreta del sistema (por ejemplo, en el marco de una transferencia a la comunidad), y aquellos relacionados con los objetivos académicos en cuanto a la formación de futuros investigadores.

- *Cognitivos:* lograr profundo conocimiento teórico y práctico, de las técnicas, algoritmos, desarrollo y construcción de sistemas basados en conocimiento, y en particular, en sistemas basados en aprendizaje supervisado.

Asimismo se planea lograr un profundo conocimiento en la modelización de problemas complejos y no-lineales, determinando las herramientas computacionales más apropiadas, sus ventajas y desventajas, y ganar experiencia práctica en la implantación de tales sistemas.

- *Relación con el medio:* Especificar, analizar, diseñar, implementar, y probar el sistema en campo.

Obtener una realimentación y adaptar el sistema de manera tal que sea de gran utilidad a los expertos: bomberos, personal de lucha contra el

fuego, y personal de gestión de los recursos para el combate contra el fuego.

Así este sistema tendrá las funciones de ser una herramienta preventiva en la lucha contra incendios.

- *Académicos*: Generar un marco, para la enseñanza-aprendizaje y motivación de alumnos acerca de la importancia de sistemas basados en conocimiento, y de algoritmos de aprendizaje automático como manera de resolver una problemática concreta en el campo de las ciencias naturales. Como resultado, se pretende involucrar a alumnos en el estudio y la investigación de temas fundamentales de su carrera.

Este proyecto tiene como objetivo fundamental el de proveer de un modelo de pronóstico para poder detectar incendios forestales en la Provincia de Córdoba, especialmente en las sierras de Córdoba y luego tratando de abarcar la región del parque Chaqueño en la Provincia de Córdoba [1], [2].

Se pretende elaborar un modelo matemático/computacional que capture la presencia de patrones de comportamiento humanos y patrones de índole climática en factores tales como humedad, presión, temperatura y cantidad de lluvia caída en una zona determinada.

Las condiciones meteorológicas como la temperatura y el viento influyen en los incendios forestales, tal es así que desde los años 70 se conoce el Índice Canadiense de Incendios Forestales (Canadian Forest Fire Weather Index - FWI) [3], el cual se compone de 7 índices basados en 4 observaciones meteorológicas (temperatura, humedad relativa, lluvia y viento). Este índice es empleado en Argentina, y en muchos países alrededor del mundo, y es de fácil recolección por cualquier estación meteorológica.

Actualmente se cuenta con un modelo inicial que permite caracterizar el problema de los incendios forestales. Este modelo ha sido desarrollado empleando técnicas de aprendizaje automático (machine learning), y utilizando para ello modelos no supervisados como redes neuronales y maquinas vectores de soporte.

Este modelo inicial es susceptible de diversas mejoras y queda aún realizar su prueba en contraste con datos reales, ya que con nuestro grado de avance actual se ha podido realizar predicciones de incendios forestales del parque

Montesino en Portugal, ya que se contaba con datos certeros acerca de la ocurrencia de incendios en esa región y nuestras primeras pruebas experimentales auguran resultados muy buenos sobre estos datos.

Para centrarnos en el entorno del problema, se puede determinar que los incendios forestales son uno de los mayores problemas ambientales y que producen un daño ecológico, económico y humano, irreparables. Es por ello que una detección prematura de los mismos, es una herramienta clave que puede permitir una lucha más eficaz contra este flagelo.

En la actualidad existen distintas aproximaciones [4], [5] que intentan dar soluciones empleando herramientas automáticas [6] basadas en sensores locales (meteorológicos), otras se basan en satélites, y un tercer grupo de técnicas, en scanners de humo e infrarrojos [7].

En este proyecto pretendemos construir un sistema, empleando redes neuronales [8] artificiales y maquinas vectores de soporte, y alimentarlas con datos meteorológicos no costosos, como los sugeridos por índices internacionales, con el objetivo de poder predecir la existencia o no de un área quemada y en ese caso, de la extensión de la misma [9]. La información de salida que este sistema proveería, sería de vital importancia para el planeamiento estratégico de los recursos destinados a combatir los incendios forestales.

Si bien el pronóstico de incendios es una tarea difícil, se piensa que un enfoque basado en aprendizaje automático podrá inferir un modelo matemático útil y aplicable en la práctica, de manera tal de servir de un sistema de soporte de decisión a la planeación estratégica de recursos destinados a incendios forestales. Como resultados podríamos obtener un sistema que indique las áreas en las que ocurrirá un incendio y la cantidad de hectáreas que se verían afectadas.

Nuestro objetivo es aplicar este modelo en la Provincia de Córdoba que ha sido ampliamente afectada por los incendios forestales en años recientes.

Según el informe preliminar de valoración sobre el impacto del incendio en Sierras Chicas ocurrido en septiembre de 2006, los focos de fuego afectaron aproximadamente a 18.640 hectáreas distribuidas en el cordón montañoso de Sierras Chicas de la región natural de Sierras

de Sur, ubicadas en los departamentos de Colón, Punilla y Santa María. Este incendio fue uno de los de mayor magnitud registrada en Córdoba en los últimos años. La vegetación más afectada fue la arbustal-pastizal, alcanzando 11.080 hectáreas, que representa el 67% del total de la flora que fue afectada. En segundo lugar, en lo que respecta a impacto en superficie, fue el Bosque con una extensión quemada de 2.480 hectáreas.

Cabe destacar que la recuperación del estrato arbóreo podría demorar aproximadamente 60 años.

Desde el punto de vista de las pérdidas económicas derivadas del incendio forestal, se cuentan las pérdidas económicas del bosque, el costo de la extinción del incendio y costo parcial de las pérdidas de infraestructura. Así, las personas residentes permanentes del área quemada, ven perjudicada su calidad de vida, tanto en su salud como en su economía y sus actividades cotidianas, cada vez que acontece dicho flagelo.

Por esta razón, un sistema de prevención del fuego que informe de alertas tempranas podría ayudar a solventar muchas de estas pérdidas, alertando a la población y a las autoridades, para que prepare y dirija sus recursos en la posible zona de siniestro, y así evitar al mínimo pérdidas económicas, ecológicas y principalmente pérdidas de vidas humanas.

## Líneas de Investigación y Desarrollo

Como se mencionó inicialmente, el presente trabajo forma parte de la línea de investigación de aprendizaje automático, esto es, en la construcción de algoritmos que intentan inferir un modelo computacional con el fin de resolver un problema concreto, en este caso la predicción de incendios, problema que se caracteriza como complejo y altamente no lineal.

Este proyecto fue aprobado por la Secretaria de Ciencia y Técnica de la UTN, y comenzó formalmente en Marzo de 2010.

## Resultados

Durante el desarrollo del proyecto se alcanzaron los siguientes resultados:

- Predicción de la ocurrencia de un incendio forestal.
- Seguimiento y gestión de los siniestros.
- Aporte de una herramienta de gestión para la ayuda en la asignación efectiva de los recursos físicos, financieros y humanos, asociados a un siniestro.
- Formación de alumnos becarios, y pasantes en el área de IA, e introducción de los mismos en tareas de investigación.

A continuación se detallan las características esenciales de los productos construidos para la concreción de los objetivos del proyecto y de los resultados alcanzados:

### Subsistema de Recopilación de Información Meteorológica

Inicialmente en el proyecto se presuponía la existencia de registros de incendios recolectados en las estaciones de bomberos, o en estaciones meteorológicas. Sin embargo, se descubrió que no se llevaba a cabo ningún tipo de registro de incendios y que era necesario un sistema que permitiera capturar al menos los siguientes datos: fecha del siniestro, coordenadas de ocurrencia del siniestro, temperatura, humedad, presión, velocidad del viento y cantidad de hectáreas afectadas. A raíz de ello, se diseñó un sistema para realizar la carga de mediciones meteorológicas y de siniestros (incendios) en el que se puedan registrar a través de una interfaz WEB esta información.



Figura 1: Pantalla de acceso Sitio Web

En la Figura 1 se puede observar la pantalla de inicio al sistema web desarrollado que permite la administración, seguimiento de mediciones, visualización de siniestros,



estadísticas y otras operaciones propias de la gestión de incendios forestales.

A partir de este sistema, se podrán capturar las mediciones diarias que realiza cada cuartel de bomberos, las cuales servirán de entrada para los módulos de entrenamiento y predicción.

### Subsistema de Entrenamiento

Este subsistema se caracteriza por tomar como entrada un conjunto de información estadística acerca de la ocurrencia de siniestros y las condiciones en las cuales el mismo se produjo. En base a estas mediciones se infiere un modelo que es la salida de este subsistema.

Debido a que aun no se ha implementado en las estaciones de bomberos el Subsistema de Recopilación de Información Meteorológica, la información necesaria para la entrada de este subsistema es recopilada manualmente a través de la búsqueda y consulta en diarios y en sitios especializados de cuidado del medio ambiente.

### Subsistema de Predicción

Este subsistema se caracteriza por tomar como entrada el modelo producido por la etapa anterior, junto con información climática actual, y realiza la inferencia de la ocurrencia o no de un siniestro.

El sub-módulo de visualización de predicciones y siniestros emplea la API de Google Maps para la visualización de los mismos en un lapso de tiempo determinado.

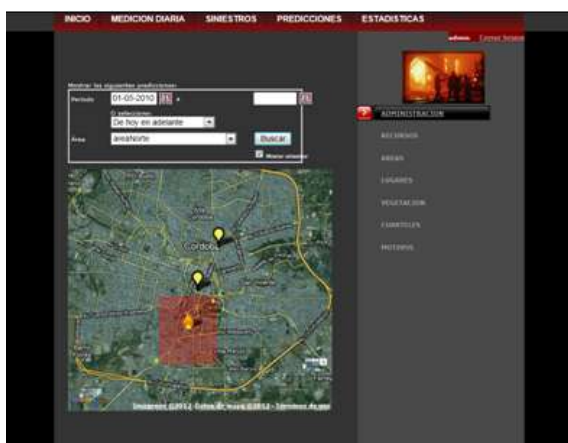


Figura 2: Ejemplo de visualización de predicciones y siniestros

En la Figura 2 pueden observarse cómo se ha utilizado la interfaz de Google Maps para la visualización de:

- siniestros: incendios ocurridos.

- predicciones: evento que el sistema pronostica que ocurrirán.

Adicionalmente, esto constituye una gran herramienta durante la etapa de testing y afinamiento del módulo de pronósticos, ya que el sistema puede probarse con datos históricos en base a siniestros ya ocurridos. De esta manera, sería posible identificar visualmente la efectividad en pronosticar incendios en las diferentes regiones de la provincia y observar gráficamente el error de la predicción (distancia del siniestro con respecto al lugar de la predicción). Esto permite analizar distintos tipos de errores: errores geográficos (si hubo una correcta predicción pero en un lugar diferente), errores temporales (si hubo una correcta predicción en un determinado lugar pero “desfasada” en el tiempo) y errores de precisión (si hubo una correcta predicción de un incendio en un lugar determinado, pero la predicción de la cantidad de hectáreas quemadas varía significativamente).

En las siguientes Figuras 3 y 4 se puede observar cómo se visualizan los datos estadísticos y la generación del reporte respectivo.

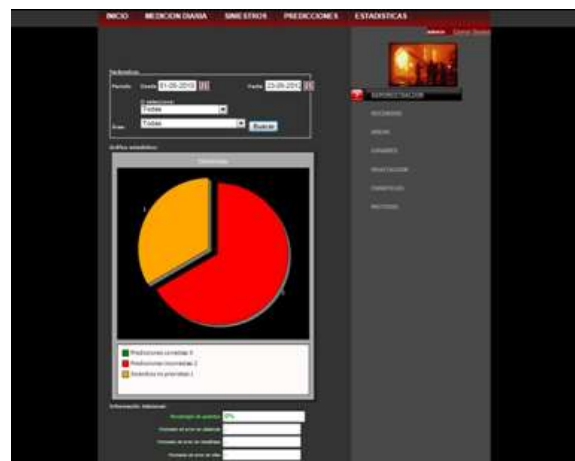


Figura 3: Ejemplo de visualización de estadísticas de siniestros y predicciones

Los reportes generados (Figura 4) por el sistema permiten conocer de manera precisa la diferencia entre las hectáreas quemadas a causa de algún siniestro y aquellas hectáreas (por quemarse) que pronosticó el sistema.

Reporte Estadístico						
Validez	Fecha Predicción	Fecha Siniestro	Hectáreas Predichas	Hectáreas Quemadas	Diferencia	Distr
Correcta	3/21/2011	3/21/2011	2.00	3.00	1.00	
Correcta	3/21/2011	3/21/2011	3.00	8.00	5.00	
Incorrecta	1/17/2010		8.00			
Incorrecta	1/3/2010		4.00			
Incendio		5/26/2011		1300.00		
Incendio		1/18/2010		20000.00		

Figura 4: Ejemplo de generación de reporte estadístico de siniestros y predicciones.

Además de la diferencia en hectáreas, el sistema provee información indicativa acerca de la diferencia en kilómetros entre un siniestro y su predicción, lo cual es particularmente útil para analizar la causa de los errores en la predicción atendiendo a la localización del error y evidenciando falencias de predicción en determinadas áreas.

## Formación de Recursos Humanos

Además de docentes, también participan de este proyecto, alumnos del último nivel la carrera de Ingeniería en Sistemas de Información de la UTN FRC, próximos a recibirse y con perspectivas de iniciarse en una carrera de posgrado o doctorado, con lo cual, uno de los objetivos del proyecto es el contribuir a la formación de dichos alumnos.

También forman parte alumnos becarios que se inician en las actividades de investigación dentro de la universidad.

El equipo de investigación y desarrollo de software, está formado por investigadores de la Universidad Tecnológica Nacional, Facultad Regional Córdoba, sin embargo también participan del proyecto, alumnos de los últimos niveles la carrera de Ingeniería en Sistemas de Información de la UTN FRC, próximos a recibirse que realizan sus prácticas supervisadas como parte de los requisitos para la obtención del grado de Ingeniero. Los alumnos intervienen, aportan su trabajo en el proyecto, y aprenden a realizar actividades de investigación, y cómo integrarse en un equipo existente. De esta manera, se está cumplimentando uno de los objetivos del proyecto que es el contribuir a la formación de alumnos en tareas de investigación, en especial fomentando el interés por los sistemas basados en IA. Asimismo, se plantea como un objetivo adicional del proyecto que sirva como marco de trabajo para que docentes-investigadores

realicen su tesis de Maestría en Sistemas de Información en el presente proyecto.

## Referencias

- [1] J. Castillo, M. Cardenas, R. Gordillo y J. Vázquez. Un Modelo de Pronósticos para Predicción de Incendios en la Provincia de Córdoba. WICC 2011, Rosario, Argentina, 2011.
- [2] J. C. Vázquez, J. Castillo, M. Cardenas, R. Gordillo, S. Villena Ruiz. Predicción de Incendios Forestales en la Provincia de Córdoba. WICC 2012, Posadas, Argentina, 2012.
- [3] C. Vega Garcia, P.M. Woodard y S.J. Titus. Dos modelos para la predicción de incendios forestales en Whitecourt Forest, Canadá. Investigación Agraria: sistemas y recursos forestales. Vol. 8, 1999.
- [4] J. Terradas J. Pinol y F. Lloret. Climate warming, wildfire hazard, and wildfire occurrence in coastal eastern Spain. Climatic Change, 38:345–357, 1998.
- [5] S. Taylor y M. Alexander. Science, technology, and human factors in fire danger rating: The Canadian experience. International Journal of Wildland Fire, 15:121–135, 2006.
- [6] S. Makridakis, S. Wheelwright y R. Hyndman. Forecasting: Methods and Applications. Wiley, 1997.
- [7] B. Arrue, A. Ottero y J. Martinez de Dios. An Intelligent System for False Alarm Reduction in Infrared Forest-Fire Detection. IEEE Intelligent Systems, 15(3):64–73, 2000.
- [8] C. Vega-Garcia, B. Lee, P. Woodard, y S. Titus. Applying neural network technology to human-caused wildfire occurrence prediction. AI Applications, 10(3):9–18, 1996.
- [9] D. Mazzoni, L. Tong, D. Diner, Q. Li, y J. Logan. Using MISR and MODIS Data For Detection and Analysis of Smoke Plume Injection Heights Over North America During Summer 2004. AGU Fall Meeting Abstracts, págs. B853+, 2005.

# Métodos, Técnicas y Herramientas para la Ingeniería de Software Orientada a Agentes

**Diana Palliotto, Gregorio N. Tkachuk, Carlos E. Vega Ugozzoli**

Departamento de Informática, Facultad de Ciencias Exactas y Tecnologías

Universidad Nacional de Santiago del Estero (UNSE)

Av. Belgrano (S) 1912, Santiago del Estero

Teléfono: +54 385 4509500 interno 1817

e-mails: {dpalliot, gregorio}@unse.edu.ar; cvegaugozzoli@hotmail.com

## Resumen

La Ingeniería de Software Orientada a Agentes es un enfoque que ha demostrado ser conveniente para construir sistemas complejos en comparación con el desarrollo estándar.

Indudablemente, el éxito de la realización de los sistemas orientados a agentes sólo puede garantizarse si se puede reducir la brecha entre el análisis y la implementación, y así desarrollar herramientas para que los conceptos y las técnicas de los sistemas multiagentes puedan implementarse fácil y directamente.

Pero, para desarrollar sistemas basados en agentes, se necesita una metodología que soporte el proceso de desarrollo como es común en otras disciplinas. En años recientes, han surgido muchos métodos y técnicas de modelado, aunque ninguna de las metodologías existentes puede considerarse completa. Se podría decir que las metodologías orientadas a agentes son el resultado de una transferencia de conocimiento desde la ingeniería del software. Seguramente, estas metodologías evolucionarán hacia el establecimiento de nuevas actividades de desarrollo lo que incrementará su alcance, provocando una mejora general de las herramientas de soporte.

Este proyecto se propone aportar al mejoramiento tanto del proceso de desarrollo como de las aplicaciones basadas en agentes, revisando, modificando y realizando propuestas metodológicas que se aplicarán y evaluarán en

sistemas basados en agentes, propiciando el avance del conocimiento científico-tecnológico.

**Palabras clave:** Ingeniería de Software, Metodología de Desarrollo de Software, Sistemas Basados en Agentes, Ingeniería de Software Orientada a Agentes.

## Contexto

Este trabajo presenta una de las líneas de investigación del programa "Sistemas de Información Web Basados en Tecnología de Agentes", aprobado por el Consejo de Ciencia y Técnica de la Universidad Nacional de Santiago del Estero (CICyT, UNSE) para el período 2012-2015. La propuesta es continuación de líneas de investigación iniciadas en el año 2005 en el proyecto "Herramientas Conceptuales, Metodológicas y Técnicas de la Informática Teórica y Aplicada" (Código 23/C062), aprobado y financiado por CICYT, UNSE.

La línea de investigación que se plantea originó el proyecto denominado "Métodos, Técnicas y Herramientas para la Ingeniería de Software Orientada a Agentes" (Código 23/C096), y está orientada a revisar, modificar y/o realizar propuestas metodológicas en el área de la Ingeniería de Software Orientada a Agentes, dirigidas particularmente al desarrollo de sistemas de información web que soporten el aprendizaje ubicuo y colaborativo.

## Introducción

Con las aplicaciones cada vez más sofisticadas exigidas por los negocios que apuntan a tomar ventajas competitivas, las tecnologías de objetos están complementándose y suplementándose con las tecnologías de agentes. Esto es especialmente cierto en áreas tales como inteligencia ambiental, e-business, servicios Web, redes y bioinformática. Estas áreas exigen software que sea robusto, que pueda operar dentro de una amplia gama de ambientes, que pueda evolucionar para cumplir con los requerimientos cambiantes, que pueda personalizarse para satisfacer las necesidades de un amplio rango de usuarios, y que sea suficientemente seguro para proteger datos personales y otros recursos. Para cumplir estos requerimientos, los desarrolladores de sistemas necesitan una metodología adecuada con el enfoque orientado a agentes. [1]

Ahora bien, existen numerosas discusiones acerca de lo que es una metodología orientada a agentes, y es sorprendente que todavía no haya ninguna definición estándar. Esto puede deberse a un abuso del término en muchos trabajos de investigación que afirman ser una metodología. Muchos investigadores han detectado este problema, aunque eso no ha provocado una autocrítica por parte de la comunidad de la ingeniería de software orientada a agentes. Como resultado, muchos artículos dicen presentar una metodología y sólo introducen un lenguaje de modelado o enumeran unas pocas actividades de desarrollo. [2]

Se podría decir que las metodologías orientadas a agentes son el resultado de una transferencia de conocimiento desde la ingeniería del software. Los investigadores de agentes son los expertos para decir si una metodología captura lo esencial del concepto de agente; sin embargo, para evaluar la capacidad de una metodología los expertos son los ingenieros de software.

Actualmente, conocer sobre ingeniería de software significa que, por lo menos, se está familiarizado con el Cuerpo de Conocimiento de la Ingeniería de Software (Software Engineering Body of Knowledge - SWEBOK) [3].

SWEBOK tiene un área denominada Métodos y Herramientas de la Ingeniería de Software y plantea que "un método puede ser heurístico, formal o de prototipado" [3]. De esta categorización de los métodos, un método orientado a agentes podría rotularse como un método heurístico donde el sistema se ve como una colección de agentes. Esto tendría sentido en la mayoría de los casos, aunque la integración con los métodos formales y/o las técnicas de prototipado puede complicar la situación. En cualquier caso, el método debe apuntar a hacer la actividad sistemática para que su éxito sea más probable, y debe proporcionar notación y vocabulario, procedimientos para realizar tareas identificables y pautas para verificar tanto el proceso como el producto.

El alcance del método puede ir desde una única fase al ciclo de vida completo. Este alcance permite hablar sobre análisis orientado a agentes el cual se concentraría en cómo determinar qué agentes son necesarios para proporcionar la funcionalidad del sistema; o bien sobre implementación orientada a agentes que dirigiría la especificación del sistema, sin tener en cuenta su orientación a agentes, para luego implementarse usando soluciones orientadas a agentes (por ejemplo, un lenguaje de programación orientado a agentes o un framework orientado a agentes). La fase o el proceso concreto se define separadamente y puede ser ad hoc o derivarse de algún modelo de ciclo de vida del software. Claramente, el proceso o la fase afectados se inclinarán por el uso de conceptos relacionados a agentes, utilizando los productos de ingeniería necesarios para referirse a esos conceptos. Finalmente, debería haber herramientas de soporte que



ayuden en la ejecución de las diferentes actividades dirigidas por el método.

Actualmente, pueden encontrarse dos grandes grupos de propuestas metodológicas orientadas a agentes [2]. En el primero de ellos, existe una metodología que primero intenta construir una notación y un vocabulario para luego ocuparse de algunas fases de desarrollo específicas. En el segundo, la notación y el vocabulario están incluidos en una herramienta que permite moverse desde el análisis y el diseño hacia la implementación.

En el primer grupo, hay trabajos que se introdujeron como metodologías en su momento, pero deberían introducirse como métodos específicos de algunas fases; GAIA [4, 5] y MESSAGE/UML [6] son ejemplos de este grupo.

En el segundo grupo, los trabajos se desarrollaron alrededor de un soporte poderoso de herramientas que cubren una parte importante del ciclo de vida. ZEUS [7] y MaSE [8] son representativas de este grupo.

Con el tiempo estos dos grupos se han unido y, como resultado, se ha incrementado el número de propuestas metodológicas que consideran más partes del ciclo de vida de software. Las metodologías como INGENIAS, PASSI, ADELFE o Prometheus consideran fases como requerimientos, implementación y prueba [1]. Cuantas más fases incorporen, están más cercanas a una metodología de calidad competitiva.

Como el área de agentes autónomos y de sistemas multiagentes se ha convertido en una tecnología prometedora, que ofrece alternativas razonables para el diseño de sistemas distribuidos inteligentes, se han realizado numerosos esfuerzos (académicos, industriales y de consorcios de estandarización) para proporcionar nuevas herramientas, con la intención de establecer las normas necesarias para lograr un uso extendido de las técnicas de los sistemas multiagentes. Para esto, es esencial que tal tecnología pueda incorporarse en las prácticas existentes en la industria del software, y no que

se vea simplemente como un nuevo paradigma prometedor. El éxito está irrevocablemente asociado con el desarrollo de lenguajes y herramientas que soporten una programación eficaz, incluso la implementación de nociones importantes en un framework unificado.

Con respecto a los lenguajes de programación, se pueden mencionar los siguientes: Jason [9, 10], 3APL [9, 11], IMPACT [9, 12], el lenguaje CLAIM y la plataforma distribuida SyMPA [9].

En relación a lenguajes y plataformas que son extensiones o están basados en Java, entre otros, se pueden indicar los siguientes: JADE [9, 13], Jadex [9, 14] y JACK [9, 15].

## Descripción de la línea de investigación

En nuestro país existen, en el ambiente académico, varios grupos de investigación cuya línea central son los sistemas basados en agentes, pero se concentran principalmente en el desarrollo de aplicaciones basadas en agentes y en la construcción de herramientas de soporte para el desarrollo de agentes.

Esta línea de investigación se propone contribuir al perfeccionamiento del proceso de desarrollo y, por ende, de las aplicaciones basadas en agentes, a través de la revisión, la modificación y/o la realización de propuestas metodológicas que luego se aplicarán y evaluarán en sistemas basados en agentes, propiciando el avance del conocimiento científico-tecnológico.

Consecuentemente, el objetivo general que guía esta investigación es:

*Aportar al mejoramiento de la calidad tanto del proceso de desarrollo como de las aplicaciones basadas en agentes.*

Considerando este objetivo general, se plantean los siguientes objetivos específicos:

- a. Alcanzar un mayor entendimiento del paradigma orientado a agentes.

- b. Evaluar las técnicas de modelado existentes, considerando tanto criterios de la ingeniería de software así como características de los sistemas basados en agentes, en especial aquellos que se utilizan como apoyo al aprendizaje ubicuo y colaborativo.
- c. Facilitar el modelado, el desarrollo, el mantenimiento y la reutilización de los sistemas basados en agentes.
- d. Realizar aplicaciones y derivaciones metodológicas y técnicas de la Ingeniería de Software Orientada a Agentes.
- e. Contribuir al mejoramiento de la calidad (portabilidad, interoperabilidad, etc.) de los sistemas basados en agentes.

## Metodología

Con el fin de dar consecución al objetivo específico **a**, se realizarán el estudio y el análisis de los siguientes temas:

- Paradigmas de la Ingeniería de Software y de la Ingeniería de Software Orientada a Agentes.
- Metodologías de desarrollo de sistemas basados en agentes, particularmente de sistemas de información web que soportan el aprendizaje ubicuo y colaborativo.
- Lenguajes y plataformas de desarrollo de sistemas basados en agentes.

Luego, para alcanzar el objetivo específico **b**, se proponen las siguientes actividades:

- Establecer criterios de evaluación provenientes de la ingeniería de software.
- Establecer las características a evaluar de los sistemas basados en agentes.
- Evaluar las técnicas en base a los criterios y las características.
- Revisar los resultados de las evaluaciones.

Con el propósito de dar cumplimiento al objetivo específico **c**, las actividades son:

- En base a las evaluaciones anteriores, determinar las ventajas y desventajas de

cada método y herramienta, indicando los problemas que se detectaron.

- Producir informes sobre la aplicación de los métodos y herramientas.
- Validar empíricamente la aplicación de los métodos y herramientas.

Para el objetivo específico **d**, las actividades son las que siguen:

- Seleccionar los métodos y modelos de agentes.
- Analizar y formular mejoras.
- Proponer modelos, métodos y/o técnicas.
- Validar e interpretar los resultados y aplicarlos.

Finalmente, para alcanzar el objetivo específico **e**, se plantean las actividades siguientes:

- Realizar un juicio técnico de los sistemas basados en agentes desarrollados.
- Detallar clara y exhaustivamente todas y cada una de las actividades completas y los productos generados.

## Resultados esperados

Los resultados que se esperan obtener de la realización de este proyecto son:

- Evaluación de métodos, técnicas y herramientas existentes para la ingeniería de software basada en agentes.
- Propuestas metodológicas (modelos, métodos y/o técnicas) para el desarrollo de sistemas basados en agentes, que deberán validarse adecuadamente aplicándolas en desarrollos concretos.

El proyecto impulsa la apertura de una nueva línea de investigación sobre ingeniería de software. Además, el desarrollo de este proyecto permitirá afianzar al grupo de investigación en la temática del proyecto.

## Formación de Recursos Humanos

El proyecto facilitará la formación de recursos humanos de la FCEyT, brindando la

posibilidad a investigadores noveles de realizar sus primeras experiencias en investigación, y permitiendo que los integrantes estudiantes realicen sus Trabajos Finales de Grado de la carrera Licenciatura en Sistemas de Información en el marco del proyecto.

## Referencias

- [1] Henderson-Sellers, Brian; Giorgini, Paolo (Editors). "Agent-oriented methodologies". Idea Group Publishing. London, UK, 2005.
- [2] Gomez-Sanz, Jorge J.; Fuentes-Fernández, Rubén; Pavón, Juan. "Understanding Agent Oriented Software Engineering Methodologies". 10 AAMAS Conference, Workshop AOSE 2011. Taipei, Taiwan, 2011.
- [3] Abran, Alain; Moore, James W. (executive editors); Bourque, Pierre; Dupuis, Robert (editors). "Guide to the software engineering body of knowledge: 2004 version (SWEBOK)". The Institute of Electrical and Electronics Engineers, 2004.
- [4] Wooldridge, M., Jennings, N. R. and Kinny, D. "The Gaia Methodology for Agent-Oriented Analysis and Design". Journal of Autonomous Agents and Multi-Agent Systems, 3 (3). pp. 285-312. 2000.
- [5] Zambonelli, F.; Jennings, N. R.; Wooldridge, M. "Developing multiagent systems: The Gaia methodology". ACM Transactions on Software Engineering and Methodology, 12(3):317-370, 2003.
- [6] Caire, Giovanni et al. "Agent Oriented Analysis using MESSAGE/UML". Proceeding AOSE '01 Revised Papers and Invited Contributions from the Second International Workshop on Agent-Oriented Software Engineering II. Springer-Verlag, London, UK, 2002.
- [7] Nwana, Hyacinth S.; Ndumu, Divine T.; Lee, Lyndon C. "ZEUS: An Advanced Tool-Kit for Engineering Distributed Multi-Agent Systems". Proceedings Agents 1999, pp. 360-361. 1999.
- [8] Deloach, Scott A.; Wood, Mark F.; Sparkman, Clint H. "Multiagent Systems Engineering". International Journal of Software Engineering and Knowledge Engineering, Vol. 11, No. 3, pp. 231-258. World Scientific Publishing Company, 2001.
- [9] Bordini, Rafael H.; Dix, Jiirgen; Dastani, Mehdi; El Fallah Seghrouchni, Amal (Editors). "Multi-Agent Programming. Languages, Platforms and Applications". Springer Science+Business Media, Inc. 2005.
- [10] Sitio Web de Jason. Disponible en: <http://jason.sourceforge.net/Jason/Jason.html>
- [11] Sitio Web de 3APL. Disponible en: <http://www.cs.uu.nl/3apl>
- [12] Sitio Web de IMPACT. Disponible en: <http://www.cs.umd.edu/projects/impact>
- [13] Sitio Web de JADE. Disponible en: <http://jade.tilab.com>
- [14] Sitio Web de Jadex. Disponible en: <http://jadex-agents.informatik.uni-hamburg.de>
- [15] Sitio Web de JACK. Disponible en: <http://www.agent-oftware.com.au/products/jack>

# Agentes inteligentes en ambientes dinámicos

Daniel Trevisani    María Laura Pino    Germán Braun    Mario Moya  
Pablo Kogan    Sandra Roger    Laura Cecchi    Gerardo Parra

email: {daniel.cipo,maria.laura.pino,germanbraun,moya.mario}@gmail.com  
{pablo.kogan,giuvago,lcecchi,gerardopar}@gmail.com

*Grupo de Investigación en Lenguajes e Inteligencia Artificial*

Departamento de Teoría de la Computación

Facultad de Informática

UNIVERSIDAD NACIONAL DEL COMAHUE

Buenos Aires 1400 - (8300)Neuquén - Argentina

## Resumen

El objetivo general de este Proyecto de Investigación es el estudio y desarrollo de técnicas de Inteligencia Artificial para dotar de inteligencia y conocimiento a agentes inmersos en mundos virtuales, interactivos y dinámicos.

El énfasis es puesto tanto en formalismos de planificación y razonamiento rebatible para la creación y control de agentes inteligentes, como en el impacto que tienen las tecnologías del lenguaje humano (TLH) en la inclusión social. En estos escenarios, el razonamiento, la toma de decisiones, la planificación de acciones y el aprendizaje ocurren bajo restricciones de tiempo críticas y en intensa interacción con el usuario.

**Palabras Clave:** AGENTES INTELIGENTES, PROGRAMACIÓN EN LÓGICA, REVISIÓN DE CREENCIAS, PLANEAMIENTO, SISTEMAS ARGUMENTATIVOS, LENGUAJE NATURAL, MINERÍA DE OPINIÓN, SIMPLIFICACIÓN DE TEXTO, INTERFACE, ACCESIBILIDAD.

## Contexto

Este trabajo está parcialmente financiado por la Universidad Nacional del Comahue, en el contexto del proyecto de investigación *Agentes inteligentes en ambientes dinámicos*. El proyecto de investigación tiene prevista una duración de cuatro años, desde enero del 2013 hasta diciembre de 2016.

## 1. Introducción

En las últimas décadas, el concepto de agente ha revolucionado el área de Inteligencia Artificial (de ahora en más IA) y se ha convertido en un pilar dominante en las Ciencias de la Computación. Actualmente, la investigación en IA concentra sus esfuerzos principalmente en el análisis, diseño y construcción de agentes autónomos que actúan en forma racional.

En la actualidad, son cada vez más las aplicaciones cotidianas que son soportadas por la IA, a través de agentes inteligentes: filtros de email, búsqueda en la web, juegos, e-commerce (comercio electrónico), control de tráfico, entre otras. Es casi imposible pensar



la computación actual sin los agentes, pero esto nos enfrenta a nuevos desafíos.

En todo ambiente relevante, el dinamismo del entorno provoca, usualmente, que los agentes deban adaptarse a nuevas situaciones y resolver problemas impredecibles para poder alcanzar los objetivos finales. Por otra parte, el agente tiene restricciones de tiempo para adecuarse a los cambios constantes en el entorno de acción.

Por esta razón, se requiere el estudio de diferentes técnicas para modelar agentes inteligentes, a fin de determinar cuáles se adaptan de manera más adecuada a ambientes dinámicos, donde el tiempo de respuesta es crítico.

Estos agentes interactúan con otros, posiblemente humanos, formando un sistema multiagentes. Es fundamental en estos casos, el uso de las Tecnologías del Lenguaje Humano evitando que los sistemas desarrollados aumenten el riesgo de exclusión social.

Los párrafos anteriores brindan el contexto a partir del cual, resulta clara la importancia estratégica de desarrollar conocimiento y recursos humanos en esta área de las Ciencias de la Computación. Los sistemas inteligentes, como tecnología, han representado y representan una de las aplicaciones más importantes desde el punto de vista económico representando inversiones de millones de dólares para un mercado de un orden de magnitud mayor.

A lo largo de este proyecto se estudiarán cuestiones fundacionales de la teoría y la construcción de aplicaciones en Agentes Inteligentes inmersos en ambientes dinámicos. Se buscará como meta el desarrollo de conocimiento especializado en el área de Inteligencia Artificial, estudiando técnicas de representación del conocimiento y razonamiento, junto con métodos de planificación y tecnologías del lenguaje natural aplicadas a dicho agentes.

Existen diversos dominios sobre los que podemos aplicar y evaluar los resultados teóricos alcanzados a partir de nuestra investigación. Se prevé la generación de prototipos de

sistemas inteligentes que puedan ser testeados particularmente en un dominio dinámico e interactivo y cuya implementación involucre a agentes inteligentes como el fútbol con robots y los videojuegos, como así también en el análisis de TLH en el ámbito de la inclusión social, como facilitadora del acceso a la información y a la educación.

## 2. Líneas de investigación y resultados esperados

El proyecto de investigación *Agentes inteligentes en ambientes dinámicos* tiene varios objetivos generales. Por un lado, el de desarrollar conocimiento especializado en el área de Inteligencia Artificial. Además, se estudian técnicas de representación de conocimiento y razonamiento, junto con métodos de planificación[15, 20] y tecnologías del lenguaje natural aplicadas al desarrollo de sistemas multiagentes.

### 2.1. Tecnologías del lenguajes humano

Dentro de la línea de investigación relacionada a las tecnologías del lenguaje humano la temática de desarrollo está basada, primordialmente, en la inclusión social. Uno de los objetivos de esta investigación es la creación de herramientas inteligentes que puedan acortar la brecha entre la tecnología y la sociedad. Para ello se han determinado tres sublíneas.

**Simplificación de texto:** El objetivo de esta línea es el desarrollo de tecnologías para facilitar el acceso a la información para todas aquellas en el que la inteligibilidad del texto es de gran importancia.

La forma en la que están estructuradas las sentencias, sean éstas por su longitud y complejidad, ocasionan un gran problema tanto a nivel de comprensión como a nivel de entendimientos del texto en sí mismo. Como se menciona en [8]: “Si la complejidad del texto

puede hacerse más simple, las oraciones resultarán más fáciles de procesar tanto para las personas como para los programas”.

Como trabajo futuro, se analizarán los estudios y herramientas disponibles según las diferentes personas destinatarias con el fin de lograr el desarrollo de las tecnologías necesarias para el idioma español.

### **Interfases accesibles en Lenguaje Natural para consultas a bases de datos:**

Una Interface en Lenguaje Natural para consultar Bases de Datos (A Natural Language Interface to Query Databases (NLIDB)) es un sistema que permite a los usuarios acceder a la información almacenada en las bases de datos por medio de requerimientos expresados en algún lenguaje natural [1, 18] tal como español, inglés, etc. Existen muchas teorías interesantes, enfoques sobre cómo deberían construirse para mejorar su precisión, cómo hacerlos más flexibles en términos de las expresiones que acepta, etc.

**Minería de opinión:** Millones de mensajes aparecen diariamente en los sitios más populares de *microblogging*, comentarios de noticias en diarios web, blogs, etc. Los autores de estos mensajes escriben acerca de sus vidas, comparten sus opiniones sobre una variedad de temas y discuten sobre éstos. Toda esta información que los usuarios generan en las publicaciones acerca de los productos que utilizan, visión política o religiosa, se vuelve un recurso de gran valor para el análisis de opiniones y sentimientos de la opinión pública.

El estudio de estos temas mediante un seguimiento continuo junto con la determinación sobre los acontecimientos o hechos causales de variaciones en la opinión pública son cruciales a la hora de tomar una decisión. Tanto a nivel de consumidor como de proveedor esta información tiene un gran valor estratégico, que les brinda una tendencia y/o comparativa del valor mundial a través del tiempo.

El objetivo de esta línea es lograr una

herramienta web accesible. De esta manera, el usuario puede proponer una temática para analizar el comportamiento de la opinión pública sobre dicho tema. Actualmente, se está trabajando en producir resultados que sirvan de base de comparación a futuros análisis y mejoras. En este sentido, se pretende analizar diferentes fuentes de búsqueda, algoritmos de clasificación, herramientas lingüísticas, etc. para un mejor desempeño del sistema.

## **2.2. Complejidad Descriptiva del sistema DeLP**

La Complejidad Descriptiva[16, 11] caracteriza a las clases de complejidad por la categoría de la lógica que se requiere para expresar los lenguajes que pertenecen a la clase. Así, podemos clasificar los problemas, en términos de la complejidad de los lenguajes formales que permiten expresarlos.

La Programación en Lógica Rebatible (Defeasible Logic Programming), de ahora en más DeLP, es una herramienta basada en argumentación para representar conocimiento tentativo y razonar con él [13].

Esta línea de investigación estudia a la DeLP como una estructura lógica finita, con el objeto de determinar su poder expresivo y compararlo con otros sistemas, continuando el trabajo realizado en [6, 3, 4, 5, 7] donde se definió una semántica basada en juegos para DeLP y se comenzó con el análisis de la complejidad temporal y descriptiva. Algunos de los resultados obtenidos en esta línea son:

- Definición de las complejidades de dato, de programa y combinada para DeLP.
- El problema de decisión: “dado un árbol dialéctico  $\mathcal{T}$ , determinar si la raíz está etiquetada con U” es PSPACE-completo y AP-completo.
- Introducción de DeLP como un lenguaje de consulta rebatible para bases de datos ordenadas finitas, basado en sistemas argumentativos.

- Se analizaron aquellas consultas que requieren de un operador de clausura transitiva y se mostró que el conjunto de las consultas  $FO(TC)$ , que coinciden con las consultas de complejidad computacional  $NL$ , son expresables en el lenguaje de consulta rebatible.

A través de la Complejidad Descriptiva relacionamos las clases de complejidad con el poder expresivo necesario para caracterizar el problema. Si bien la complejidad descriptiva del lenguaje, no coincide necesariamente con la complejidad del cómputo de una consulta, esta relación ayuda a determinar y circunscribir el poder expresivo del lenguaje. Por esta razón, es que entre nuestras tareas en progreso se encuentran analizar si las fórmulas de primer orden extendidas con el menor punto fijo y las fórmulas de segundo orden extendidas con el operador de clausura transitiva son expresables en el lenguaje de consulta rebatible. Asimismo, se pretende relacionar estos resultados con las complejidades de dato y combinada, concepto introducido en [4].

### 2.3. Planificación

Uno de los objetivos de esta línea de investigación es el intento de dotar a un agente inteligente tanto de capacidades reactivas como deliberativas. Las capacidades deliberativas se logran a partir de la implementación de un planificador novedoso, denominado *planificador continuo* [17], una de las alternativas para planificación en ambientes reales planteadas en [19]. En esta aproximación, se presenta un agente que persiste indefinidamente en un entorno, posiblemente cambiante y dinámico. Tal agente no se detiene al alcanzar un meta determinada. Por el contrario, continúa actuando en una serie de fases que se repiten e incluyen la formulación de metas, planificar y actuar.

Se ha concretado la implementación de un traductor de un subconjunto relevante del lenguaje PDDL para la descripción de los dominios y de las acciones, de manera tal que puedan ser manipuladas por el framework de

*planificación continua*[2]. El uso de técnicas de planificación de IA para mejorar los procesos de *E-learning*[14] ofrecen una posibilidad concreta de aplicación del módulo desarrollado.

Otro punto de interés dentro de esta línea es el estudio, análisis e implementación de estrategias de tratamiento de metas[9, 10, 12]. El objetivo puntual es el diseño de un módulo destinado al manejo de metas que pueda ser anexado al planificador continuo existente. La incorporación de este módulo enriquecerá al planificador, pues contribuirá a enfrentar de mejor manera las características dinámicas del ambiente en el que actúa.

## 3. Formación de Recursos Humanos

Durante la ejecución del proyecto se espera lograr la culminación de dos tesis doctorales, la iniciación de, al menos, dos tesis de postgrado y la finalización de seis tesis de grado dirigidas y/o co-dirigidas por los integrantes del proyecto. Ya se han iniciado, en el contexto del grupo y dirigidas por algunos de sus miembros, dos tesis de Licenciatura en Ciencias de la Computación. Finalmente, se espera la consolidación como investigadores de los miembros más recientes del grupo de investigación.

## Referencias

- [1] S. Abiteboul, V. Hull, and R. Viannu. *Foundations of Database Systems*. Addison Wesley, 1995.
- [2] G. Braun, M. Moya, and G. Parra. Sistemas multiagentes en ambientes dinámicos: Planificación continua mediante pddl. In *Actas del XIII Workshop de Investigadores en Ciencias de la Computación*, pages 163–167, Rosario, Santa Fe, Argentina, 2011. Universidad Nacional de Rosario.
- [3] L. A. Cecchi, P. R. Fillottrani, and G. R. Simari. An Analysis of the

- Computational Complexity of DeLP through Game Semantics. In *XI Congreso Argentino de Ciencias de la Computación*, pages 1170–1181, Argentina, Octubre 2005. Universidad Nacional de Entre Ríos.
- [4] L. A. Cecchi, P. R. Fillottrani, and G. R. Simari. On the complexity of DeLP through game semantics. In J. Dix and A. Hunter, editors, *XI International Workshops on Nonmonotonic Reasoning*, pages 386–394, Clausthal University, 2006.
- [5] L. A. Cecchi and G. Simari. *DeLP marking procedure for dialectical trees is PSPACE-complete*. EDULP - Editorial de la Universidad de La Plata, 2012.
- [6] L. A. Cecchi and G. R. Simari. Sobre la relación entre la Semántica GS y el Razonamiento Rebatible. In *X CACiC - Universidad Nacional de La Matanza*, pages 1883–1894, San Justo - Pcia. de Buenos Aires, 2004.
- [7] L. A. Cecchi and G. R. Simari. DeLP como lenguaje de consulta: un enfoque preliminar al poder expresivo. In *XVIII Congreso Argentino en Ciencias de la Computación*, Universidad Nacional del Sur, Argentina, Octubre 2012.
- [8] R. Chandrasekar, C. Doran, and B. Srinivas. Motivations and methods for text simplification. In *In Proceedings of the Sixteenth International Conference on Computational Linguistics (COLING '96)*, pages 1041–1044, 1996.
- [9] M. T. Cox and C. Zhang. Mixed Initiate Goal Manipulation. *AI Magazine*, 28(2), 2007.
- [10] C. da Costa Pereira and A. G. B. Tetamanzi. Goal generation with relevant and trusted beliefs. In *Proceedings of the 7th Intl. Joint Conf. on Autonomous Agents and Multiagent Systems*, 2008.
- [11] H.-D. Ebbinghaus and J. Flum. *Finite Model Theory*. Springer-Verlag, 1995.
- [12] R. Feldmann, G. Brewka, and S. Wenzel. Planning with Prioritized Goals. In *Proceedings of the 10th International Conference on Principles of Knowledge Representation and Reasoning*, 2006.
- [13] A. J. García and G. R. Simari. Defeasible Logic Programming: An Argumentative Approach. *Theory and Practice of Logic Programming*, 4(1):95–138, 2004.
- [14] A. Garrido, L. Morales, and I. Serina. Using AI Planning to Enhance E-Learning Processes. In *Proceedings of the XXII ICAPS*, 2012.
- [15] M. Ghallab, D. Nau, and P. Traverso. *Automated Planning. Theory and Practice*. Morgan Kaufmann, 2004.
- [16] N. Immerman, editor. *Descriptive Complexity*. Springer-Verlag, New York, 1999.
- [17] M. Moya and C. Vaucheret. Agentes deliberativos basados en planificación continua. In *X Workshop Agentes y Sistemas Inteligentes (WASI)*, Martiarena esquina Italia - S.S. de Jujuy, Octubre 2009. Universidad Nacional de Jujuy - Facultad de Ingeniería.
- [18] N. Nihalani, S. Silakari, and M. Motwani. Natural language interface for database: A brief review. In *International Journal of Computer Science Issues*, vol. 8, no. 2, pages 600–608, 2011.
- [19] S. Russell and P. Norvig. *Artificial Intelligence: A modern approach*. Prentice Hall, New Jersey, third edition, 2009.
- [20] J. Zhang, X. Nguyen, and R. Kowalczyk. Graph-based Multi-agent Replanning Algorithm. In *Proceedings of the Sixth Intl. Joint Conf. on Autonomous Agents and Multiagent Systems*, 2007.



# Agentes inteligentes en ambientes dinámicos: Análisis de opinión

Sandra Roger

Pablo Kogan

email: {giuvago,pablo.kogan}@gmail.com

*Grupo de Investigación en Lenguajes e Inteligencia Artificial*

Departamento de Teoría de la Computación

Facultad de Informática

UNIVERSIDAD NACIONAL DEL COMAHUE

Buenos Aires 1400 - (8300)Neuquén - Argentina

## Resumen

La meta fundamental de este proyecto está puesta tanto en los formalismos de planificación y razonamiento rebatible para la creación y control de agentes inteligentes, como en el impacto que tienen las tecnologías del lenguaje humano (TLH) en la inclusión social. En estos escenarios, el razonamiento, la toma de decisiones, la planificación de acciones y el aprendizaje ocurren bajo restricciones de tiempo críticas y en intensa interacción con el usuario. Esta línea de investigación se centra en el desarrollo de una aplicación destinada al estudio y seguimiento de la opinión pública sobre un tema determinado.

El crecimiento de internet junto con el desarrollo de la Web 2.0 (Web Social) posibilita que personas de todo el mundo compartan información global. acerca de un tema (producto, servicio, persona política, etc.) logrando así tomar una decisión más apropiada a la hora de elegir. Con el Objetivo de compartir Millones de mensajes aparecen diariamente en los sitios más populares de *microblogging*, comentarios de noticias en diarios web, blogs, etc. Los autores de estos mensajes escriben acerca de sus vidas, comparten sus opiniones sobre una variedad de temas y discuten sobre estos. Toda

esta información que los usuarios generan en las publicaciones acerca de los productos que utilizan o visión política y religiosa, se vuelve un recurso de gran valor para el análisis de opiniones y sentimientos de la opinión pública.

El estudio de estos temas mediante un seguimiento continuo junto con la determinación sobre los acontecimientos o hechos causales de variaciones en la opinión pública son cruciales a la hora de tomar una decisión. Tanto a nivel de consumidor como de proveedor esta información tiene un gran valor estratégico, que les brinda una tendencia y/o comparativa del valor mundial a través del tiempo.

**Palabras Clave:** AGENTES INTELIGENTES, LENGUAJE NATURAL, MINERÍA DE OPINIÓN, ANÁLISIS DE SENTIMIENTOS, ACCESIBILIDAD.

## Contexto

Este trabajo está parcialmente financiado por la Universidad Nacional del Comahue, en el contexto del proyecto de investigación *Agentes inteligentes en ambientes dinámicos*. El proyecto de investigación tiene prevista una duración de cuatro años, desde enero del 2013 hasta diciembre de 2016.

## 1. Introducción

La información textual disponible en la web podría ser categorizada en expresiones de hecho o de opinión. Las expresiones de hechos están relacionadas a entidades, eventos y sus propiedades. Por otro lado, las de opinión son usualmente expresiones subjetivas que describen algún sentimiento o valoración sobre las personas, entidades, eventos y sus propiedades [5].

Junto con el desarrollo de la tecnología y el creciente acceso a la información, hemos sido testigos del nacimiento de un nuevo tipo de sociedad: la sociedad de la interactividad y comunicación [8]. En este nuevo contexto, el papel de los sentimientos expresados en la web se han vuelto crucial. Las personas expresan las emociones que determinan ciertos hechos. Por otra parte, otras personas que tienen acceso a estas expresiones, las transforman bajo sus propias influencias en otras expresiones.

Uno de los *microblogging* más populares es Twitter. Los usuarios de Twitter pueden expresar las siguientes intenciones en sus tweets: charla diaria, conversaciones, intercambio de información, noticias y presentación de informes. Además, los usuarios de Twitter tienden a publicar las opiniones personales con respecto a ciertos temas y acontecimientos noticiosos. Una gran ventaja de estas opiniones es que se facilitan de manera libre y voluntariamente por los usuarios. Por lo tanto, los datos textuales de opiniones publicadas podría ser agregada y usada para medir la opinión pública implícita. Sin embargo, la elevada cantidad de opiniones generadas diariamente en aplicaciones de medios sociales, hace una evaluación humana de este contenido imposible de lograr. Por esta razón, estas opiniones textuales se suele evaluar mediante métodos computacionales

La minería de opinión o análisis de sentimientos se refiere a la aplicación de técnicas del campo del procesamiento del lenguaje natural, recuperación de información y clasificación de texto, para identificar y extraer información a partir de datos textuales subjetivos. Algunas de las tareas más importantes del campo son:

distinguir entre la información objetiva y opiniones en fuentes de datos textuales, y para detectar los sentimientos en los textos dogmáticos al identificar si una tiene una opinión positiva o negativa relacionadas al tema tratado.

Teniendo en cuenta la riqueza del lenguaje humano y su gran poder expresivo y ambigüedad inherente al mismo, el problema de la clasificación de sentimientos no es trivial.

Las aplicaciones en donde puede ser crucial el análisis de los sentimientos pueden ser: Resúmenes de opiniones [2], opiniones de libros o películas [3, 4], análisis de opiniones políticas: Análisis de candidatos políticos [1, 6], e-government [7] para analizar impacto de decisiones, análisis del impacto de productos y marcas [1].

Los recursos lingüísticos para OM definen algunas propiedades relacionadas a los sentimientos. Los avances sobre este tópico tratan con tres tareas principales: *Determinación de la orientación del término*: positivo, negativo, neutro; *Determinación de la subjetividad de un término*, si un término tiene una naturaleza subjetiva u objetiva: verde, alto, líquido, etc; y, *Determinación de la fuerza de la determinación del término (orientación o subjetividad)*, como el grado de positividad o negatividad del término.

Las investigaciones sobre OM han tomado tres líneas de investigación interrelacionadas [5]

- Desarrollo de recursos lingüísticos para el análisis de sentimientos tal como corpus léxico anotado manualmente;
- Implementación de diferentes algoritmos para el análisis del texto y clasificación de acuerdo a su orientación semántica y subjetiva;
- Extracción de opiniones del texto, incluyendo diferentes tipos de relaciones con contenido asociado.

En este trabajo se pone principal énfasis en el seguimiento continuo de una temática en la web, junto con la determinación de los acontecimientos o hechos causales de variaciones en

la opinión pública, siendo esto crucial a la hora de la toma de decisión. Brindando una información de gran valor estratégico que nos muestra una tendencia y/o comparativa de su valor mundial a través del tiempo.

## 2. Líneas de investigación y desarrollo

El proyecto de investigación *Agentes inteligentes en ambientes dinámicos* tiene varios objetivos generales. Por un lado, el de desarrollar conocimiento especializado en el área de Inteligencia Artificial. Además, se estudian técnicas de representación de conocimiento y razonamiento, junto con métodos de planificación y tecnologías del lenguaje natural aplicadas al desarrollo de sistemas multiagentes.

Específicamente, esta línea se centra en el estudio de un sistema multiagente en ambientes dinámicos para el seguimiento continuo de la opinión pública sobre un determinado tema de interés.

Las encuestas fueron tradicionalmente, la forma de obtener información acerca de la opinión pública, siendo estas estáticas en un tiempo discreto. A diferencia de este tipo de encuestas, este trabajo está enfocado en realizar un seguimiento continuo de la opinión pública. Esta opinión está expresada públicamente en diferentes sitios de la web. Teniendo este corpus a disposición el proceso continúa realizando una clasificación de la información obtenida acerca de la temática a analizar. Por ejemplo si la temática a analizar es el “asignación universal por hijo” se pueden buscar los comentarios de las noticias relacionadas con este tema, y clasificarlos si están a favor o en contra.

El objetivo de esta investigación es desarrollar una herramienta para hacer este proceso de forma automática.

La Figura 1 muestra la arquitectura básica del sistema multiagentes destinada al análisis de opinión. La misma está dividida en cuatro agentes principales: Agente Buscador, Agente Filtrador, Agente Analizador y Agente Compositor.

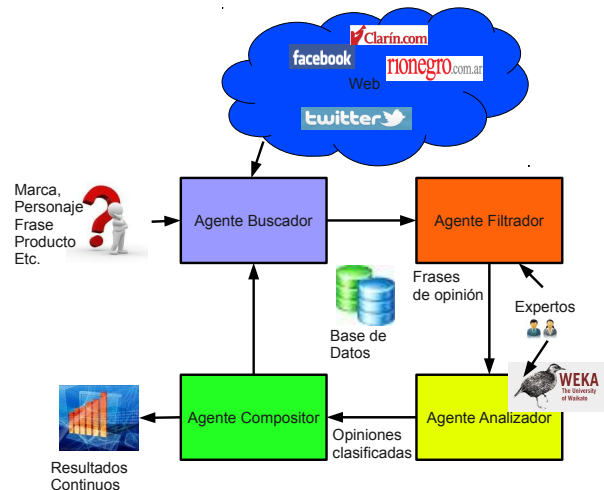


Figura 1: Arquitectura de agentes para análisis de opinión

El primer agente que entra en juego es el **Agente Buscador**, el cual tiene tres tareas principales: análisis de la entrada o consulta, búsqueda sobre la web, y finalmente almacenar lo buscado en una base de datos.

Se producen además dos procesos. En el primero, se realiza una desambiguación de la entrada en el caso de ser necesario. Por ejemplo si se está buscando a Riquelme, se producirá una desambiguación entre el jugador de fútbol argentino “Juan Román Riquelme”; la modelo paraguaya “Larissa Riquelme”; el niño holandés cuyo nombre es “Riquelme Van Gool”, en homenaje al jugador de fútbol; entre otros. El segundo proceso, trata también con la consulta pero en el sentido de producir la relajación de la entrada, esto quiere decir que si estamos buscando a “Cristina Fernandez de Kirchner” también se considere, por ejemplo, al término “presidenta de la Argentina” como equivalente. En este sentido debemos identificar un algoritmo óptimo para la construcción de la entrada y sus limitaciones. Usando las consultas correctas se podrá encontrar las sentencias adecuadas en el proceso de recuperación.

El proceso de búsqueda es realizado en diferentes ámbitos:

- *Búsqueda en microblogs conocidos:* Los *microblogs* se han convertido en una he-

ramienta muy popular entre los usuarios de internet. Millones de mensajes aparecen diariamente en los sitios más populares de *microblogging* como *Twitter*, *Facebook*, *Tumblr*, etc. Los autores de estos mensajes escriben acerca de su vida, comparten sus opiniones sobre una variedad de temas y discuten sobre estos. Como el formato de los mensajes es libre y de fácil acceso, los usuarios tienden a modificar su forma de comunicación de blogs y listas de correos tradicionales a servicios de *microblogging*.

- *Comentarios de noticias de diarios (La Nación, Clarín, etc.):* La proliferación de los diarios en su versión *on-line*, posibilitan a los lectores la opción de comentar las noticias, con el objetivo de hacer al diario más interactivo. Esta fuente de información es muy rica en contenido y en opinión. La búsqueda sobre los comentarios no es tan trivial como la anterior. Esta se realiza a través de un robot web que va navegando las noticias relacionadas con el tema y almacenando los comentarios.
- *Búsqueda en la web a través de buscadores:* aprovechando el resultado que arrojan los buscadores se realiza un robot web que navega los *links* y devuelve resultados de blogs, listas de correos públicos y noticias de sitios poco conocidos.

Toda la información obtenida se almacena en una base de datos con toda la información que se puede obtener de la persona que publica su opinión.

El **Agente Filtrador** se encarga de descartar todos los datos del corpus que no sirven, como por ejemplo entradas duplicadas, entradas que no demuestran sentimientos, etc..

El **Agente Analizador** se encarga de clasificar las entradas del corpus en sentimientos. Inicialmente comenzaremos a trabajar con una ontología de dos sentimientos: “amor” y “odio”. Este agente es el encargado de realizar un proceso de entrenamiento sobre análisis de sentimientos. Esta tarea es realizada con la he-

rramienta Weka<sup>1</sup> e inicialmente utilizado el clasificador *Support Vector Machine* (SVM) dado su relativo éxito en el tratamiento del lenguaje natural. Posteriormente se realizará un estudio comparativo más profundo sobre otros clasificadores.

Finalmente, el **Agente Compositor** es el encargado de componer los resultados obtenidos por el agente analizador en un lapso de tiempo determinado. El factor tiempo en conjunto con los resultados son los puntos más importantes a analizar. Los resultados obtenidos podrían modificar el comportamiento del agente buscador antes de comenzar un nuevo ciclo.

## 2.1. Primera versión del sistema

En una primera versión del sistema, se pretende crear un *baseline* con el cual comparar posteriormente diferentes mejoras en las diferentes etapas del sistema. Dicho *baseline* se definirá de la siguiente manera.

El *agente buscador* toma Twitter y se encarga de hacer una recuperación o filtrado de tweets sobre un tema específico. Almacenando los mismos en una base de datos con información del usuario emisor del tweet. No se produce, además, ninguna consideración relacionada al tweet. Es decir, no se tiene en cuenta si el tweet es de opinión o informativo. No se produce ningún proceso de desambiguación ni de relajación de la entrada. En esta instancia el agente buscador sólo conoce la entidad sobre la cual se desea realizar un análisis de opinión y el rango de fechas que se desea utilizar para medir la opinión de los usuarios acerca del tema planteado.

El *Agente Filtrador* sólo se encarga de descartar los datos redundantes del corpus, como por ejemplo entradas duplicadas. En este sentido, se ha decidido realizar esta eliminación dado que no se utiliza ninguna información relacionada al usuario emisor de la opinión.

El **Agente Analizador** encargado de clasificar las entradas del corpus en sentimientos, comenzará a trabajar con una ontología de dos sentimientos: “amor” / “odio” o “positivo” /

<sup>1</sup>[www.cs.waikato.ac.nz/ml/weka/](http://www.cs.waikato.ac.nz/ml/weka/)



“negativo”. Este agente es el encargado de realizar un proceso de entrenamiento sobre análisis de sentimientos. En este caso el algoritmo será considerado positivo si tienen un número mayor de palabras positivas que negativas.

Finalmente, el **Agente Compositor** pondrá los resultados obtenidos por el agente analizador en el lapso de tiempo introducido como entrada en el agente buscador.

### 3. Resultados esperados

El objetivo de este sistema es lograr una herramienta web accesible. De esta manera, el usuario puede proponer una temática para analizar el comportamiento de la opinión pública sobre dicho tema. Actualmente, se está trabajando en producir resultados que sirvan de base de comparación a futuros análisis y mejoras. En este sentido, se pretende analizar diferentes fuentes de búsqueda, algoritmos de clasificación, herramientas lingüísticas, etc. para un mejor desempeño del sistema.

### 4. Formación de Recursos Humanos

Durante la realización de este sistema se espera lograr, como mínimo, la culminación de 3 tesis de grado dirigidas y/o co-dirigidas por los integrantes del proyecto.

Finalmente, es constante la búsqueda hacia la consolidación como investigadores de los miembros más recientes del grupo

## Referencias

- [1] C. G. Akcora, M. A. Bayir, M. Demirbas, and H. Ferhatosmanoglu. Identifying breakpoints in public opinion. *1st Workshop on Social Media Analytics (SOMA 10)*, 2010.
- [2] A. Bossard, M. Génereux, and T. Poibeau. Cbseas, a summarization system integration of opinion mining techniques to summarize blogs. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics: Demonstrations Session, EACL '09*, pages 5–8, Stroudsburg, PA, USA, 2009. Association for Computational Linguistics.
- [3] S. Morinaga, K. Yamanishi, K. Tateishi, and T. Fukushima. Mining product reputations on the web. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '02*, pages 341–349, New York, NY, USA, 2002. ACM.
- [4] B. Pang and L. Lee. Thumbs up? sentiment classification using machine learning techniques. In *In Proceedings of EMNLP*, pages 79–86, 2002.
- [5] B. Pang and L. Lee. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2):1–135, 2008.
- [6] M. J. Silva, P. Carvalho, L. Sarmento, E. Oliveira, and P. Magalhães. The design of OPTIMISM, an opinion mining system for portuguese politics. In *New Trends in Artificial Intelligence: Proceedings of EPIA 2009 - Fourteenth Portuguese Conference on Artificial Intelligence*. Universidade de Aveiro, Oct. 2009.
- [7] G. Stylios, D. Christodoulakis, J. Besharat, M.-A. Vonitsanou, I. Kotrotsos, A. Koumpouri, and S. Stamou. Public opinion mining for governmental decisions. In F. Bannister, editor, *ECEG Conference Issue*, volume 8, pages 202 – 213, 2010.
- [8] M. Wiberg. *The Interaction Society: Theories, Practice and Supportive Technologies*. Information Science Publishing, 2004.

# Prototipo de Sistema de Control Fuzzy Aplicado al Proceso de Curado de Tabaco Virginia

Miguel A. Azar, Sergio L. Martínez, Jesús M. Manero

Facultad de Ingeniería / Universidad Nacional de Jujuy

Av. Italia y Av. Martearena / S. S. de Jujuy / Provincia de Jujuy/Tel. 0388-4221591

auazar@live.com, smartinez@fi.unju.edu.ar, jmanero@fca.unju.edu.ar

## Resumen

El aprovechamiento de las ventajas de la inteligencia artificial en el ámbito de la producción de cultivos industriales es uno de los elementos en los que muchas soluciones se apoyan para lograr un producto acorde a las exigencias del mercado. Dentro del dominio de la IA los sistemas Fuzzy Logic constituyen una de las técnicas más clásicas para el diseño de sistemas de control inteligentes. El presente proyecto trata acerca del desarrollo de un prototipo fuzzy para el control de estufas de curado de tabaco.

**Palabras clave:** Cultivos industriales, Fuzzy Logic, Inteligencia Artificial, Secadero Bulk Curing, Sistema de control, Tabaco Virginia.

## Contexto

El presente proyecto está basado en un trabajo interdisciplinario entre las Facultades de Ingeniería y Ciencias Agrarias de la UNJu. Por otro lado, la documentación fue extraída de dossiers elaborados por la Cámara del Tabaco de Jujuy y Massalin Particulares.

## Introducción

La lógica fuzzy se concibe como un paradigma diferente del procesamiento de la información y es muy útil para la representación de datos inexactos, incompletos y el manejo de la incertidumbre [Kasavob, 1998].

El conjunto de características que reúne la técnica fuzzy la convierten en la herramienta ideal tanto para la representación de datos inciertos como para el control de sistemas que habitualmente son manejados por expertos. Las estufas empleadas para el secado de tabaco actualmente son manipuladas por expertos (o estuferos) que utilizan una sucesión de reglas ambiguas que no pueden ser modeladas mediante la lógica convencional.

## El proceso del curado

La producción de tabaco es un proceso que se desarrolla en cuatro fases [Guzmán Pérez *et al.*, 1983], que son: 1) cultivo, 2) Recolección, 3) Curado (o secado de la hoja) y 4) Procesado. El *curado* es la fase de mayor importancia y se define como el proceso mediante el cual se dirigen las reacciones biológicas (químicas) de la hoja de tabaco para lograr la calidad prefijada. Para el tabaco tipo Virginia se emplea el curado “a fuego indirecto” [1] mediante estufas a gas que permiten, a través del calor que generan, “amarillar” y secar las hojas de tabaco. La estufa consiste en una cámara o recinto cerrado que utiliza calor artificial proveniente de la combustión de gas. Actualmente se ha extendido el uso de estufas Bulk-Curing o de Circulación Forzada [Agrovisión NOA, 1997]. Por ello el curado se realiza en forma semiautomática, debiendo ser necesaria la intervención de una persona (“estufero”) dedicada completamente al control de

dicho proceso. Este control difiere de acuerdo a la hoja con la que se esté tratando. Las hojas se clasifican por la posición que ocupan en el tallo (pisos foliares) [Llanos Company, 1981]. Los pisos foliares son: Bajera, Media o Superior.

### El proceso de curado

El curado es un proceso de secado o pérdida de agua en condiciones controladas para que las hojas de tabaco, mantengan el mayor tiempo posible su actividad biológica a fin de que los cambios químicos y bioquímicos se produzcan del modo más apropiado para conseguir un producto de alta calidad [Llanos Company, 1981]. El curado se divide en 4 fases: 1) Amarillamiento (o *amarilleo*), 2) Fijación de color, 3) Secado de lámina y 4) Secado de nervadura (o *secado de vena*). En ellas el operador o estufero manipula las válvulas de gas a fin de lograr un control de temperatura que depende de la madurez, la humedad inicial y el piso foliar del cual provienen las hojas.

No existe un programa o secuencia absoluta de pasos a seguir para lograr un estufado exitoso. Cada bibliografía proporciona diferentes formas de llevar a cabo el control del estufado no obstante ello, la curva de temperatura y humedad relativa que debe emplear el estufero, depende de varios factores.

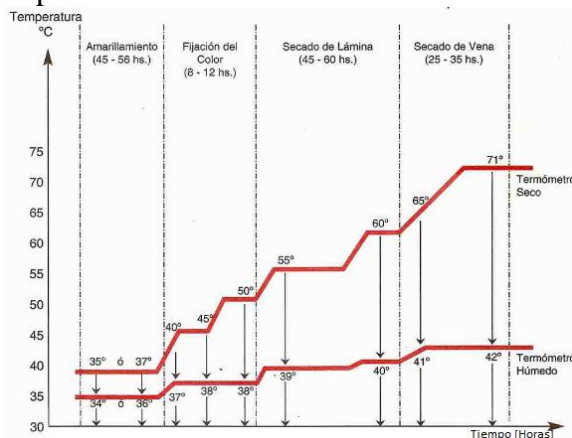


Figura 1

### Análisis y modelado

Para el caso en estudio, se puede observar una tendencia muy notable de requisitos definidos en forma ambigua. Así por ejemplo, durante la educación el experto emplea la expresión siguiente: “*Para comenzar el amarilleo hay que medir la temperatura de bulbo seco en ese momento. Si es alta se abre la válvula de gas al mínimo para que aumente la temperatura lentamente*”. Esto denota claramente que desde el punto de vista lingüístico existen definiciones imprecisas tales como: “alta”, “mínimo” y “lentamente”. El hecho de que las expresiones vertidas por parte del experto tengan características difusas no significa que sean inconsistentes. El principal control efectuado por parte del estufero es el de la apertura de la válvula de gas para aumentar o disminuir la temperatura en el interior de la estufa. El estufero a partir de una temperatura de referencia, ajusta la apertura de la válvula de gas de manera que el error entre lo medido y lo deseado sea nulo.

### Análisis y modelado de las entradas de referencia

Ahora bien, la mayoría de los sistemas de control posee una entrada de referencia constante, sin embargo para el caso en estudio el estufero mantiene una referencia variable. La temperatura de referencia para el experto está regida por una gama de curvas predeterminadas. Esta situación se ve reflejada en la Figura 1, en donde se puede observar que a medida que avanza el tiempo la temperatura se va incrementando gradualmente hasta llegar a un máximo de 71 °C.

La variación de temperatura de referencia va a estar en función de las siguientes condiciones iniciales a saber: 1) *Madurez de la hoja*. La hoja que ingresa al secadero puede estar verde, madura o sobremadura (estos son valores nítidos). 2) *Piso Foliar*.

Es la procedencia o ubicación de la hoja respecto al tallo y puede ser bajera, media, superior o corona. Las hojas de la parte media y superior poseen el mismo cuerpo o grosor, por lo tanto son indistintas para el experto. 3) *Temperatura inicial*. Es la temperatura que posee la estufa antes de iniciar el curado. Esta puede ser alta o baja. 4) *Fase*.

La temperatura inicial es un parámetro fuzzy ya que el estufero la define como alta o baja. En palabras del experto “una temperatura baja es de unos 20 °C. A partir de los 25 °C a 30 °C ya es alta”.

La Fase tiene una definición precisa (amarilleo, fijación de color, secado de lámina y secado de vena), sin embargo el momento de cambio de una fase hacia la otra no está nítidamente definido. La bibliografía especializada y los expertos coinciden en que no se inicia una fase una vez que finaliza la anterior sino que el cambio de fase es gradual. Esto es lo que permite definir a esta variable como fuzzy (Figura 2).

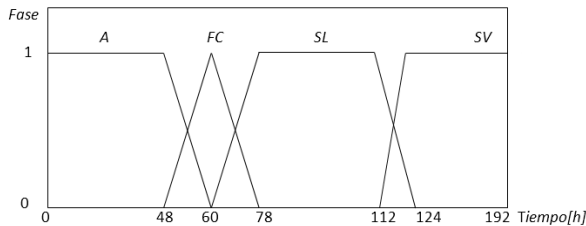


Figura 2

Este conjunto de variables condiciona las decisiones a tomar por parte del experto sobre como realizará el estufado. Dicho de otro modo, el experto posee un programa diferente de curvas de temperatura para cada combinación de variables iniciales. Este conjunto de combinaciones están modeladas en la Tabla 1.

		AG				
		N	B	ME	A	MA
d(AG)	N	N	LEP	LIP	VP	VP
	B	LEN	N	LEP	LIP	VP
	ME	LIN	LEN	N	LEP	LIP
	A	VN	LIN	LEN	N	LEP
	MA	VN	VN	LIN	LEN	N

Tabla 1

### Análisis y modelado del sistema controlado

Para el caso en estudio el *sistema controlado* representa a la estufa o secadero. Muchos sistemas dinámicos ya sean mecánicos, eléctricos, térmicos, etc., pueden ser caracterizados por ecuaciones matemáticas [Ogata, 1980].

La apertura de la válvula de gas también es una variable fuzzy dado que el experto define el control de su apertura o cierre mediante expresiones lingüísticas difusas, tales como: “mínimo”, “máximo”, “menos del máximo” y “un poco más del mínimo”. La salida del FIS (Fuzzy Inference System) de control es la apertura de gas (AG). El universo en discurso está comprendido entre 0 y 1 (0 apertura nula, 1 apertura máxima).

El experto hace referencia a “*aumentar lentamente la temperatura*” o “*mantener la temperatura sin dejar enfriar el secadero*”, es por ello que el resultado de las reglas también se expresa en esos términos

La base de reglas pertenecientes al control de la apertura de gas está diseñada a partir de la plantilla de reglas de MacVicar-Whelan [Cheong, 2008]

### Implementación

En la Figura 3 se presenta el modelo del prototipo de sistema desarrollado en Simulink®. Se trata de un sistema continuo de simulación normal cuyo tiempo límite se estableció en 11.520 unidades de tiempo. Este valor se determinó debido a que cada intervalo de tiempo representa un minuto (el tiempo máximo de estufado son 8 días, es decir 11.520 minutos).

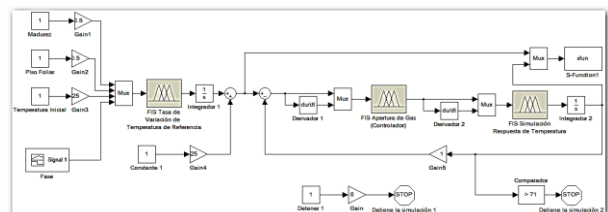


Figura 3



Este modelo hace uso S-functions que proporcionan un mecanismo eficaz para ampliar las capacidades del entorno Simulink® [2].

Un S-function es la descripción del lenguaje máquina de un bloque de Simulink® [MathWorks, 2001].

### Líneas de Investigación y desarrollo

Como resultado del trabajo realizado se pueden inferir las siguientes líneas de investigación futuras que permitirían complementar y adicionar nuevas características al sistema bajo estudio: 1) Desarrollo de un controlador híbrido neuro-fuzzy en donde la incorporación de una red neuronal artificial pudiera capturar e incorporar a la base de conocimientos – en tiempo real–, las actividades del operario. 2) Desarrollo de un control central multiplexado de n estufas con un solo computador o microcontrolador, con posibilidades de control a distancia.

### Resultados y objetivos

#### Objetivo Principal

Desarrollar un prototipo de sistema de control fuzzy aplicable al curado del tabaco Virginia sobre estufas Bulk Curing.

#### Objetivos Secundarios

- Conocer y utilizar un entorno matemático de simulación computacional que permita el desarrollo y la aplicación de técnicas basadas en lógica fuzzy.
- Investigar y analizar los parámetros que emplean los expertos en el curado de tabaco, para modelar un conjunto de variables de control apropiadas.
- Diseñar un Modelo de Inferencia Difuso mediante el uso de herramientas de diseño fuzzy para la configuración de las variables manipulables.

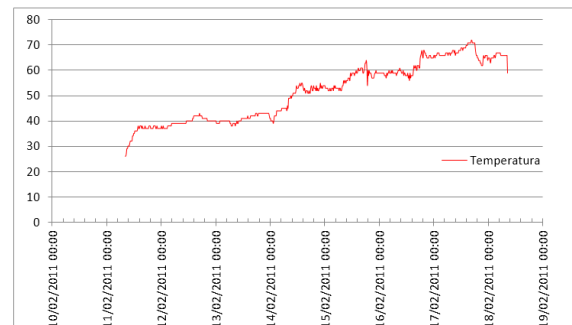
La lógica fuzzy, además de su utilización muy difundida y relativamente reciente en sistemas de control, es una buena alternativa para el modelado de sistemas

controlables en los cuales su función transferencia es desconocida o bien de difícil configuración matemática debido a su alta complejidad.

El empleo del entorno MATLAB® constituyó un excelente medio para la determinación de los resultados del modelo y la constatación con los datos empíricos en forma rápida y confiable.

Los resultados obtenidos a través de las diferentes gráficas generadas por el sistema fueron contrastados con mediciones efectuadas *in situ* en estufas Bulk Curing, para las mismas condiciones de entrada, mediante el empleo de un registrador de temperatura electrónico.

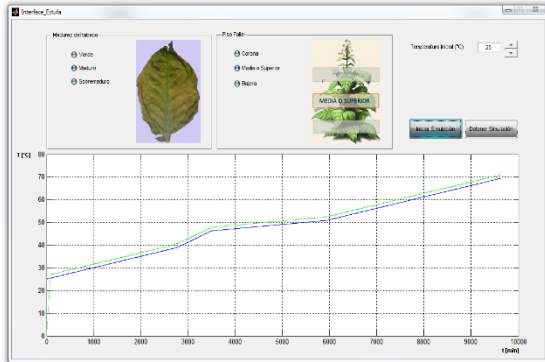
Para la evaluación del prototipo, se tomó como parámetros de comparación las pendientes obtenidas y el tiempo total de cada fase en el sistema contra los mismos parámetros medidos en casos exitosos de mediciones sobre estufas Bulk Curing. Las pendientes comparadas fueron tomadas por tramos, es decir para cada fase del proceso se consideró su pendiente por separado. Para obtener las pendientes de las curvas medidas se las linealizó utilizando el método de mínimos cuadrados. En la Figura 4 se muestra una gráfica de una medición realizada con los siguientes parámetros: Estufa: Bulk Curing, Madurez del Tabaco: Maduro, Piso Foliar: Superior, Temperatura Inicial: 26 °C.



**Figura 4**

El eje de ordenadas graduado entre 0 y 100 representa la temperatura en grados Celsius. La curva posee al inicio un valor de 26 °C y de 71 °C al finalizar el proceso

de curado. Los picos y valles observados en las curvas denotan la complejidad de controlar en forma manual la temperatura. La Figura 5 ilustra la aplicación una vez ejecutada la simulación. Se puede apreciar la linealización de la curva para los mismos parámetros de entrada.



**Figura 5**

La linealización como puede observarse se realizó por tramos, en donde cada uno de estos representa una fase del proceso de curado,

### **Formación de recursos humanos**

El equipo de trabajo estuvo formado por: Ing. Sergio Martínez (Facultad de Ingeniería - UNJu), Ing. Jesús Manero (Facultad de Ciencias Agrarias - UNJu), Ing. Miguel Azar (Facultad de Ingeniería - UNJu), Cr. Roberto Strizic y Sr. Candelario Casasola.

### **Referencias**

[Agrovisión NOA, 1997] Revista Agrovisión NOA - Compendio tabacalero, **Circulación Forzada y Optimización de Estufas**. Editor: Horacio David & Asociados. (1997)

[Cheong, 2008] Cheong F., **A hierarchical fuzzy system with high input dimensions for forecasting foreign exchange rates**. Int. J. Artificial Intelligence and Soft Computing, Vol. 1, No. 1. (2008).

[Guzmán Pérez *et al.*, 1983] Guzmán Pérez, J. L. Muñoz Muñoz B. **Recolección**

**y Curado de los Tabacos Amarillos Tipo Virginia**. Hojas Divulgadoras Núm. 11/83. Publicaciones de Extensión Agraria. Neografis. Madrid, España. (1983).

[Kasabov, 1998] Kasabov, N. K. **Foundations of Neural Networks, Fuzzy Systems, and Knowledge Engineering** The MIT Press. London, England. (1998).

[Llanos Company, 1981] Llanos Company, M. **El Tabaco - Manual Técnico para el Cultivo y Curado** Ediciones Muni-Prensa. Madrid, España. (1981).

[Ogata, 1980] Ogata, K. **Ingeniería de Control Moderna** Prentice Hall Hispanoamericana S.A. Minnesota, USA. (1980).

### **Referencias Web**

[1] Philip Morris International **Growing Tobacco**[online].<[http://www.pmi.com/en/g/our\\_products/growing\\_tobacco/pages/growing\\_tobacco.aspx](http://www.pmi.com/en/g/our_products/growing_tobacco/pages/growing_tobacco.aspx)>, consultada el 19 de Febrero de 2011.

[2] MathWorks® R2011b **Documentation Simulink® - What is an S-function?** [online].<<http://www.mathworks.com/help/toolbox/simulink/sfg/f6-21998.html>>, consultada el 09 de Junio de 2011.

# Mejoramiento de los Servicios de Razonamiento basados en Programación Lógica Rebatible

Sebastián Gottifredi      Luciano H. Tamargo      Diego R. García  
Alejandro J. García      Marcelo A. Falappa      Guillermo R. Simari

Laboratorio de Investigación y Desarrollo en Inteligencia Artificial  
Departamento de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur,  
Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)  
Av. Alem 1253, (B8000CPB) Bahía Blanca, Argentina  
Tel: (0291) 459-5135 / Fax: (0291) 459-5136  
e-mail: {lt, sg, drg, ajg, maf, grs}@cs.uns.edu.ar

## 1. Resumen

Esta línea de investigación se enfoca en mejorar las capacidades de razonamiento de agentes que participan en Sistemas Multi-Agente (SMA). Su objetivo general es mejorar y desarrollar nuevas técnicas para los servicios de razonamiento basados en Programación Lógica Rebatible (Defeasible Logic Programming o DeLP) combinando revisión de creencias, argumentación y mecanismos de confianza y reputación para su aplicación en SMA. En particular, se espera avanzar en la implementación de nuevos servicios y facilidades de los Servicios de Razonamiento basados en DeLP. Dicha integración permitirá realizar un avance en las tres áreas de investigación, y además, proveerá de técnicas avanzadas aplicables a los modelos de razonamiento de agentes inteligentes y sistemas multi-agente para entornos dinámicos.

**Palabras Clave:** Razonamiento, Argumentación, Sistemas Multi-Agente.

## 2. Contexto

Esta línea de investigación se realizará dentro del ámbito del Laboratorio de Investigación y Desarrollo en Inteligencia Artificial (LIDIA) del Departamento de Ciencias e Ingeniería de la Computación de la Universidad

Nacional del Sur (U.N.S.). Está asociada con los proyectos de investigación: “Representación de conocimiento y Razonamiento Argumentativo: Herramientas Inteligentes para la Web y las Bases de Datos Federadas” PGI 24/N030, y “Formalismos Argumentativos aplicados a Sistemas Inteligentes para Toma de Decisiones”. Código: PGI 24/ZN18 financiado por la U.N.S.

## 3. Introducción

Los sistemas de argumentación resultan útiles para toma de decisiones, tales deliberaciones pueden modelarse de una forma dialéctica usando argumentos a favor y en contra de cada opción [11, 13, 5, 4]. La implementación de este tipo de sistemas ha cobrado impulso en los últimos años, en particular para brindar soporte a las capacidades de razonamiento de los agentes, por ejemplo para mejorar los mecanismos de toma de decisión automática. Una aproximación a la implementación de este tipo de sistemas está dada por los servicios de razonamiento basados en Programación Lógica Rebatible o DeLP-servers.

Los DeLP-servers introducidos en [14] y desarrollados luego [26] constan de tres componentes: conocimiento público, un mecanismo de inferencia, y un conjunto de operadores que definen cómo integrar nuevo conocimiento en-

viado por los clientes que consultan a un DeLP-server. Para realizar dichas consultas se define el concepto de Consulta Contextual, la cual consta de dos elementos: la consulta particular que el cliente pretende hacer, y un contexto dentro del cual debe procesarse dicha consulta particular. Este contexto podrá contener conocimiento privado incluido por el cliente para ser utilizado en el cálculo de la respuesta, y operadores (conocidos por el DeLP-server) a través de los cuales se especificará qué tratamiento darle a dicho conocimiento privado enviado en la consulta. Por ejemplo, un DeLP-server puede disponer de un operador para dar prioridad al conocimiento privado sobre el público, otro operador para excluir conocimiento público, o un operador para dar prioridad al conocimiento público. De esta manera, utilizando los operadores que son indicados en el contexto, se puede integrar adecuadamente el conocimiento privado enviado por el cliente al conocimiento público; y calcular la respuesta, utilizando tanto el conocimiento público almacenado, como el contexto en el cual el cliente quiere que se trate su consulta. El cómputo de una respuesta para una consulta contextual no afecta a futuras consultas del mismo agente, ni tampoco a otras consultas simultáneas de otros agentes. Es importante destacar que la misma consulta contextual realizada en varias oportunidades al mismo servidor, siempre tendrá la misma respuesta. Sin embargo, la misma consulta realizada al mismo DeLP-server, pero con diferente contexto, podrá tener una respuesta diferente.

Por ejemplo, un DeLP-server podría implementar un Servicio de Recomendación de Noticias, y este servicio podría ser consultado por diferentes clientes para obtener sugerencias sobre qué publicaciones leer en la Web. Está claro que una consulta como “¿Qué noticia me recomienda?” puede ser demasiado general para que el servicio genere una respuesta satisfactoria para el cliente. En cambio, “Estoy interesado en noticias sobre tenis, preferentemente publicaciones deportivas ¿qué noticia me recomienda?” provee, además de la consulta, un contexto, el cual puede ser utilizado por el servicio para realizar una sugerencia que sea satis-

factoria. En otro dominio de aplicación como es el diagnóstico, un cliente podría enviar en el contexto de la consulta los síntomas a tener en cuenta para procesar su consulta.

Los operadores para el procesamiento del conocimiento privado permiten un manejo adecuado de la información enviada por el cliente. Por ejemplo: la consulta “¿noticias sobre tenis?” puede incluir un contexto en el cual se indique un operador para excluir noticias sobre accidentes de tenistas, y dar prioridad a noticias de tenistas argentinos.

En esta línea de investigación se pretende mejorar las capacidades de los DeLP-server combinando los resultados que se han obtenido en el tratamiento de información proveniente de diferentes informantes, formalización de los operadores con resultados en revisión de creencias y argumentación, y el uso de técnicas de optimización en el cálculo de garantías para la obtención de respuestas.

## 4. Líneas de Investigación y Desarrollo

Esta línea de investigación toma como punto de partida el desarrollo de los DeLP-servers elaborado en [26] y resultados obtenidos por los autores que han sido publicados recientemente [23, 22, 15, 20, 17].

Su objetivo general es desarrollar técnicas para mejorar las capacidades de razonamiento de agentes que participan en Sistemas Multi-Agente (SMA). Para esto se propone desarrollar nuevas técnicas para los servicios de razonamiento basados en DeLP combinando revisión de creencias, argumentación y mecanismos de confianza y reputación para su aplicación en SMA. En particular, se espera avanzar en la implementación de nuevos servicios y facilidades de los Servicios de Razonamiento basados en DeLP. Dicha integración permitirá realizar un avance en las tres áreas de investigación, y además, proveerá de técnicas avanzadas aplicables a los modelos de razonamiento de agentes inteligentes y sistemas multi-agente para entornos dinámicos.



El uso de credibilidad de informantes en dinámica de creencias es algo novedoso [25]. Este formalismo puede ser usado como punto de partida para combinar revisión de creencias y mecanismos de confianza y reputación en SMA. En [21], Sabater y Sierra sostienen que la importancia de la confianza y reputación en sociedades humanas está fuera de discusión, por lo cual no es sorprendente que varias disciplinas, cada una desde una perspectiva diferente, haya estudiado y utilizado ambos conceptos. La investigación científica en el área de mecanismos computacionales de confianza y reputación en sociedades virtuales, es una disciplina reciente orientada a incrementar la fiabilidad y performance de comunidades electrónicas. En artículos recientes como [10, 21, 9, 3] podemos notar que en ciencias de la computación hay dos elementos que han contribuido sustancialmente a incrementar el interés en confianza y reputación: el paradigma de SMA y la evolución creciente del e-commerce.

El estudio de confianza y reputación tiene muchas aplicaciones en tecnologías de Comunicación e Información. Estos sistemas han sido reconocidos como factores claves para la adopción del comercio electrónico. Los mismos son usados por agentes de software inteligentes como un mecanismo para buscar compañeros confiables y como un incentivo en la toma de decisiones acerca de si se tiene en cuenta un contrato. La reputación es usada en el mercado electrónico como un mecanismo para evitar fraudes y estafas [9]. Los e-markets no son el único campo de aplicación; por ejemplo, en [3] usan la confianza para mejorar la performance de mecanismos de revisión de creencias.

En el contexto de SMA, un agente puede a menudo recibir información a través de otro que, por lo general, llamamos informante. Estos informantes son agentes independientes quienes tiene sus propios intereses y, por lo tanto, no son completamente fiables. Es natural para un agente creerle más a un informante sobre otro. Es por esto que en algunos trabajos se ha propuesto la organización de los informantes en un orden parcial que compara la plausibilidad de los mismos. En esta inves-

tigación se pretende combinar formalismos de revisión de creencias, argumentación y actualización de conocimiento con técnicas de mantenimiento de confianza y reputación de agentes en un ambiente distribuido.

Desde el punto de vista teórico, elaborar tales conexiones entre argumentación, revisión de creencias y mecanismos de confianza y reputación, y poder aplicarlas a los SMA, permitirá un progreso científico substancial en el campo de la representación de conocimiento [19, 2, 16].

El estudio de los vínculos y relaciones entre revisión de creencias y argumentación ha comenzado hace muy poco. Aunque ya existen en la literatura comentarios [7, 12] de que ambas áreas podrían combinarse con buenos resultados, el tema carece de resultados importantes. Por lo tanto, con la línea de investigación propuesta se esperan obtener resultados relevantes en el área de razonamiento y representación de conocimiento. Por ejemplo, las técnicas de revisión de creencias proveen a los agentes la capacidad de corregir sus creencias cuando existe una discrepancia con la información nueva que pueden obtener de su entorno [1, 6, 8, 18]. Uno de los resultados esperados es mejorar los sistemas de revisión de creencias utilizando argumentación y mecanismos de confianza y reputación para evaluar si es importante que la nueva información obtenida sea incorporada o no a las creencias del agente.

## 5. Resultados Esperados

Los DeLP-servers fueron introducidos en [14] y posteriormente formalizados en [26]. En esta línea de investigación se espera poder mejorar las capacidades de los DeLP-servers, utilizando resultados obtenidos y publicados recientemente. A continuación se incluye un resumen de estas publicaciones.

En el artículo *Deliberative DeLP agents with multiple informants* [23], se define un formalismo de razonamiento argumentativo basado en confianza (trust), donde la información que posee una agente puede ser recibida de diferentes informantes. En este formalismo se utiliza

la confianza asignada a la fuente de la información para decidir ante la presencia de información contradictoria y poder obtener respuestas garantizadas por un proceso de argumentación rebatible. En este formalismo, las conclusiones tentativas del agentes están sustentadas por argumentos, y esos argumentos pueden ser atacados por otros argumentos. El mecanismo de inferencia compara los argumentos y contra-argumentos utilizando la información de la confianza asignada a la fuente de información. El formalismo permite además que la confianza asignada a las fuentes pueda ser cambiada dinámicamente utilizando los resultados reportados en [24].

En el artículo *On Influence and Contractions in Defeasible Logic Programming* [15], se desarrollan diferentes nociones de contracción basadas en la Programación Lógica Rebatible. En particular, estas nociones de contracción se basan tanto en las diferentes formas de inferencia implícitamente existentes en un sistema argumentativo, así como también en la influencia que exhiben los literales en el proceso de razonamiento dialéctico en este tipo de sistemas. Todos los operadores buscan lograr que un literal no sea garantizado a partir de un Programa Lógico Rebatible. Sin embargo, cada operador puede ser visto como un refinamiento del anterior en términos del cambio que producen respecto a las inferencias del programa. Estos resultados pueden utilizarse para definir nuevos operadores para los servicios de razonamiento que permitan la integración de conocimiento privado a través de las consultas contextuales de forma más sofisticada.

En el artículo *Selective revision with multiple informants and argumentative support* [22], consideramos el problema de revisión de creencias en un SMA con información proveniente de diferentes agentes con distintos grados de credibilidad. En este contexto, un agente tiene que elegir cuidadosamente que información acepta en la revisión para evitar creer en información poco confiable. En este artículo hemos propuesto un proceso de revisión en la cual se combina revisión selectiva, argumentación deductiva, y la credibilidad de la información

para su manejo adecuado. La nueva información es evaluada basada en la credibilidad de la fuente y todos los argumentos a favor y en contra de la misma. El proceso de evaluación determina que parte de la nueva información va a ser aceptada para luego ser incorporada en la base de creencias por medio de un operador de revisión apropiado.

En los artículos *A Heuristics-Based Pruning Technique for Argumentation Trees* [20] y *Using Argument Strength for Building Dialectical Bonsai* [17], se propone una técnica de poda basada en heurística para los árboles de dialéctica de los sistemas argumentativos. En los trabajos se muestra que el uso de la técnica propuesta reduce considerablemente el tamaño de los árboles de dialéctica, lo cual acelera significativamente la velocidad de respuesta del sistema. La heurística utilizada se basa en la noción de fuerza de un argumento, la cual indica la probabilidad de que un argumento sea derrotado y es calculada de manera *offline*. La heurística es utilizada para ordenar los argumentos, de manera tal que al construir el árbol de dialéctica se exploran primero los argumentos con menos chances a estar derrotados. Estos resultados pueden utilizarse para acelerar el computo de los servicios de razonamiento estudiando como calcular la fuerza de los argumentos potenciales que utilizará el servidor. Además, la noción de fuerza podría utilizarse como una medida para determinar el grado de garantía de una respuesta dada por el servidor.

## Referencias

- [1] Carlos E. Alchourrón, Peter Gärdenfors, and David Makinson. On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic*, 50(2):510–530, 1985.
- [2] L. Amgoud, C. Devred, and M. Lagasquie. A constrained argumentation system for practical reasoning. In *Seventh International Conference on Autonomous Agents and Multiagent Systems (AAMAS'08)*, pages 429–436, 2008.
- [3] Kathleen Suzanne Barber and Joonoo Kim. Belief revision process based on trust: Simulation experiments. In *In Proceedings of Autonomous Agents '01 Workshop on Deception, Fraud, and Trust in Agent Societies*, pages 1–12, 2001.

- [4] Christoph Beierle, Bernhard Freund, Gabriele Kern-Isberner, and Matthias Thimm. Using defeasible logic programming for argumentation-based decision support in private law. In *Third International Conference on Computational Models of Argument (COMMA'10)*, Desenzano del Garda, Italy, September 2010.
- [5] Philippe Besnard and Anthony Hunter. A logic-based theory of deductive arguments. *Artificial Intelligence*, 128(1-2):203–235, 2001.
- [6] R. Booth, T. Meyer, I. Varzinczak, and R. Wassermann. A contraction core for horn belief change: Preliminary report. In *NMR*, 2010.
- [7] Claudette Cayrol, Florence Dupin de Saint Cyr-Bannay, and Marie-Christine Lagasquie-Schiex. Change in abstract argumentation frameworks: Adding an argument. *Journal of Artificial Intelligence Research*, 38:49–84, may 2010.
- [8] J. Delgrande and R. Wassermann. Horn clause contraction functions: Belief set and belief base approaches. In *KR*, 2010.
- [9] Chrysanthos Dellarocas. The digitalization of word-of-mouth: Promise and challenges of online reputation mechanisms. In *Management Science*, 2003.
- [10] Changyu Dong and Naranker Dulay. Shinren: Non-monotonic trust management for distributed systems. In *IFIPTM*, pages 125–140, 2010.
- [11] Phan Minh Dung. On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and n-Person Games. *Artificial Intelligence*, 77(2):321–358, 1995.
- [12] Marcelo Alejandro Falappa, Gabriele Kern-Isberner, and Guillermo Ricardo Simari. Belief revision and argumentation theory. In *Argumentation in Artificial Intelligence*, pages 341–360. Springer, 2009.
- [13] A. Garcia and G. Simari. Defeasible logic programming: An argumentative approach. *Theory and Practice of Logic Programming (TPLP)*, 4:95–138, 2004.
- [14] Alejandro Javier García, Nicolás D. Rotstein, Mariano Tucac, and Guillermo Ricardo Simari. An argumentative reasoning service for deliberative agents. In *KSEM*, pages 128–139, 2007.
- [15] Diego R. García, Sebastian Gottifredi, Patrick Krümpelmann, Matthias Thimm, Gabriele Kern-Isberner, Marcelo A. Falappa, and Alejandro Javier García. On influence and contractions in defeasible logic programming. In *LPNMR*, pages 199–204, 2011.
- [16] Sebastian Gottifredi, Alejandro Javier García, and Guillermo Ricardo Simari. Query-based argumentation in agent programming. In *IBERAMIA*, pages 284–295, 2010.
- [17] Sebastian Gottifredi, Nicolás D. Rotstein, Alejandro J. García, and Guillermo R. Simari. Using argument strength for building dialectical bonsai. *To Appear in AMAI*, 2013.
- [18] Sven Ove Hansson. Kernel Contraction. *The Journal of Symbolic Logic*, 59:845–859, 1994.
- [19] N. D. Rotstein, A. J. Garcia, and G. R. Simari. Reasoning from desires to intentions: A dialectical framework. In *AAAI Conference on Artificial Intelligence*, pages 136–141, 2007.
- [20] Nicolás D. Rotstein, Sebastian Gottifredi, Alejandro J. García, and Guillermo R. Simari. A heuristics-based pruning technique for argumentation trees. In *SUM*, pages 177–190, 2011.
- [21] Jordi Sabater and Carles Sierra. Review on computational trust and reputation models. *Artificial Intelligence Review*, 24(1):33–60, 2005.
- [22] Luciano H. Tamargo, Alejandro J. García, Matthias Thimm, and Patrick Krümpelmann. Selective revision with multiple informants and argumentative support. *AEPIA vol. 15 nro. 50*, pages 4–17, 2012.
- [23] Luciano H. Tamargo, Sebastian Gottifredi, Alejandro J. García, and Guillermo R. Simari. Deliberative delp agents with multiple informants. *AEPIA vol. 15 nro. 13*, pages 13–30, 2012.
- [24] Luciano Héctor Tamargo, Marcelo Alejandro Falappa, Alejandro Javier García, and Guillermo Ricardo Simari. A change model for credibility partial order. In *Proceeding of the 5th International Conference on Scalable Uncertainty Management (SUM)*, pages 317–330, 2011.
- [25] Luciano Héctor Tamargo, Alejandro Javier García, Marcelo Alejandro Falappa, and Guillermo Ricardo Simari. Modeling knowledge dynamics in multi-agent systems based on informants. *The Knowledge Engineering Review (KER)*, 27(1):87–114, 2012.
- [26] Mariano Tucac. Grupos de servicios de razonamiento para el procesamiento de consultas contextuales en paralelo. In *Tesis de Doctor en Ciencias de la Computación, Universidad Nacional del Sur*, 2012.

## Metaheurísticas avanzadas y de población descentralizada para problemas de ruteo de vehículos.

Villagra S., Pandolfi D., Villagra A., Lasso M., Rasjido J., Mercado V., Orozco S., Serón N., Vidal P., Valdez J., Montenegro C.

Laboratorio de Tecnologías Emergentes (LabTEm) - Unidad Académica Caleta Olivia  
 Universidad Nacional de la Patagonia Austral  
 {svillagra, dpandolfi, avillagra, mlasso, jrasjido, vmercado, sorozco, nseron, pjvidal, jcvaldez, cmontenegro}@uaco.unpa.edu.ar

Leguizamón G.

Laboratório de Investigación y Desarrollo en Inteligencia Computacional (LIDIC)  
 Departamento de Informática - Universidad Nacional de San Luis  
 legui@unsl.edu.ar

### RESUMEN

En el presente trabajo se describe la línea de investigación en el campo de las metaheurísticas, en especial de las metaheurísticas híbridas y paralelas para resolver el problema de ruteo de vehículos y posiblemente otros dominios afines. Se hará hincapié en la investigación y desarrollo de algoritmos híbridos basados en cGA enriquecido con componentes de otras metaheurísticas y en la propuesta de una metodología simple para esta hibridación y que pueda extenderse a otros algoritmos.

Los algoritmos estudiados serán validados a través de estudios experimentales intensivos y analizados rigurosamente bajo la teoría estadística apropiada.

**Palabras clave:** Metaheurísticas, Paralelismo, Hibridación, Problemas de ruteo de vehículos

### CONTEXTO

La línea de investigación descrita en esta presentación se enmarca en el Proyecto de Investigación "Metaheurísticas avanzadas y de población descentralizada para problemas de ruteo de vehículos" llevado a cabo en el Laboratorio de Tecnologías Emergentes (LabTEm) en el marco del programa de Investigación en Ciencia y Tecnología de la Universidad Nacional de la Patagonia Austral.

### 1. INTRODUCCION

El diseño de algoritmos cada vez más eficientes para resolver problemas complejos, tanto de optimización como de búsqueda, ha sido tradicionalmente uno de los aspectos más importantes de la investigación en Informática. El objetivo perseguido en este campo es, fundamentalmente, el desarrollo de nuevos métodos capaces de resolver los problemas complejos con el menor esfuerzo computacional posible, mejorando así los resultados obtenidos por los algoritmos existentes. En consecuencia, esto no sólo permite

afrontar problemas actuales de forma más eficiente, sino también tareas vedadas en el pasado debido a su alto costo computacional. En este contexto, la actividad investigadora en algoritmos exactos, heurísticos ad hoc y metaheurísticos, para resolver problemas complejos de optimización, está creciendo de forma evidente en estos días. La razón es que continuamente se están afrontando nuevos problemas de ingeniería, mientras que, al mismo tiempo, cada vez se dispone de mejores recursos computacionales, como nuevos tipos de computadoras, redes y entornos como Internet. La principal ventaja de la utilización de algoritmos exactos es que garantizan encontrar el óptimo global de cualquier problema, pero tienen el grave inconveniente de que en problemas reales (que suelen ser NP-hard en la mayoría de los casos) su tiempo de ejecución y/o los requisitos de memoria crecen de forma exponencial con el tamaño del problema. En cambio, los algoritmos heurísticos ad hoc son normalmente bastante rápidos, pero la calidad de las soluciones encontradas está habitualmente lejos de ser óptima. Otro inconveniente de estos heurísticos ad hoc es que no son fáciles de definir en determinados problemas. Las metaheurísticas ofrecen un equilibrio adecuado entre ambos extremos: son métodos genéricos que ofrecen soluciones de buena calidad (el óptimo global en muchos casos) en un tiempo razonable.

En la actualidad, uno de los mecanismos que ha probado ser exitoso en el diseño de metaheurísticas eficientes para la resolución de problemas de optimización es el paralelismo. Utilizar algoritmos paralelos es una forma de aliviar los problemas vinculados a tiempos intensivos de ejecución e importantes requerimientos de memoria para resolver instancias complejas de interés actual.

El problema de ruteo de vehículos (VRP) [Dantzing y Ramster 1959] consiste en minimizar el costo de repartir la mercancía a un conjunto de clientes desde un depósito. Reducir el costo implica minimizar el número de vehículos utilizados y la longitud de sus rutas. El interés de este problema viene dado por dos causas principales. Por un lado,



el VRP es un problema NP-hard [Lenstra y Rinnooy Kan 1981] y de alto interés académico debido a su dificultad en las restricciones que incluye, y en la multitud de variantes existentes. Por otro lado, muchos problemas del mundo real pueden ser visualizados (o concebidos) como variantes de VRP, siendo aplicable tanto a grandes empresas logísticas como a empresas locales de reparto (prensa, alimentación, etc.). De hecho, el uso de métodos computarizados a menudo, en empresas de transporte puede llegar a proporcionar un ahorro significativo, llegando en muchos casos a suponer hasta el 20 % del costo total del producto [Toth y Vigo 2001]. En la historia reciente de VRP existe una evolución constante en la calidad de las metodologías empleadas para resolverlo. Estas técnicas comprenden tanto algoritmos exactos como métodos heurísticos (secuenciales y paralelos). De cualquier forma, debido a la dificultad que presenta el problema, no existe ningún método exacto capaz de resolver instancias de más de 50 clientes [Toth y Vigo 2001], [Golden et al. 1998].

## 2. LINEAS DE INVESTIGACION y DESARROLLO

En esta sección se describe la línea de investigación que se lleva a cabo en el proyecto:

- Metaheurísticas paralelas e híbridas aplicadas al problema de ruteo de vehículos

La optimización es una disciplina fundamental en campos de las ciencias tales como Informática, Inteligencia Artificial, Logística, Biología, Tecnología de la Producción, Física, etc. El concepto de optimización puede verse como el proceso de encontrar y mejorar el rendimiento de una publicación o dispositivo a partir de determinados cambios lógicos o físicos. A este tipo de problemas se los puede dividir en dos categorías [Blum y Roli 2003], [Papadimitriou y Steiglitz 1998]: aquellos en los que la solución está codificada mediante valores reales y aquellos cuya solución está codificada con valores enteros. En los últimos encontramos los problemas denominados de optimización combinatoria. Algunos ejemplos muy conocidos son el problema del viajante de comercio (TSP – Travelling Salesman Problem), el problema de asignación cuadrática (QAP – Quadratic Assignment Problems) o los problemas de planificación (Scheduling Problems), entre otros, [Cook et al. 1998], [Papadimitriou y Steiglitz 1998]. En las dos últimas décadas ha emergido un nuevo tipo de técnicas aproximadas que consiste básicamente en la combinación de métodos heurísticos (técnicas aproximadas con componentes aleatorios guiados) básicos en plataformas de más alto nivel con el fin de explorar el espacio de

búsqueda de una forma eficiente y efectiva. Estos métodos son comúnmente conocidos con el término metaheurísticas. En [Blum y Roli 2003] se pueden encontrar recopiladas varias definiciones de metaheurísticas dadas por diferentes autores, pero en general podemos decir que las metaheurísticas son estrategias de alto nivel que planifican de manera estructurada la aplicación de varias operaciones para explorar espacios de búsqueda de elevada dimensión y complejidad intrínseca. Pueden clasificarse en multitud de diferentes formas, en [Blum y Roli 2003] se da una clasificación que depende de un conjunto de características seleccionadas que las diferencian. Así, se puede tener metaheurísticas inspiradas en la naturaleza, basadas en poblaciones o en trayectorias, con función objetivo estática o dinámica, que utilizan una o varias estructuras de vecindario, o que alternativamente memorizan estados anteriores de la búsqueda. De entre las metaheurísticas más conocidas se pueden destacar, entre otras: Simulated Annealing (SA) [Cerny 1985], [Kirkpatrick et al. 1983], Tabu Search (TS) [Glover 1986], [Glover y Laguna 1993], procedimiento de búsqueda adaptativo aleatoriamente voraz (Greedy Randomized Adaptive Search Procedure - GRASP) [Feo y Resende 1995], [Resende y Ribeiro 2003], búsqueda por vecindario variable (Variable Neighborhood Search - VNS) [Hansen y Mladenovic 1997], y otros basados en una población tales como algoritmos evolutivos (EAs) [Bäck et al. 1997], optimización por colonia de hormigas (ACO - Ant Colony Optimization) [Dorigo y Caro 1999], optimización por cúmulo de partículas (PSO - Particle Swarm Optimization) [Eberhart y Kennedy 1995], etc.

En los últimos años, ha aumentado considerablemente el interés en las metaheurísticas híbridas en el campo de la optimización. Se han obtenido buenos resultados en muchos problemas de optimización clásicos y de la vida real utilizando metaheurísticas híbridas. Talbi en [Talbi 2002], [Talbi 2009] propone una taxonomía para algoritmos híbridos y presenta dos clasificaciones para este tipo de algoritmos: jerarquizada y plana. Esta clasificación establece esquemas híbridos específicos en donde, en general, diferentes algoritmos son combinados de acuerdo a ciertos criterios. Sin embargo, en este trabajo se pretende brindar una perspectiva diferente al esquema híbrido planteado previamente. Más precisamente, se pretende crear algoritmos cuya construcción siga criterios similares a los establecidos para diseñar algoritmos híbridos, pero incorporando componentes de los algoritmos antes que al algoritmo como un todo. En este sentido, se puede realizar una analogía entre la hibridación con componentes de otras metaheurísticas y la simbiosis en la naturaleza. En biología, la simbiosis [Fabian

2001] funde especies genéticamente distintas en un único organismo. Simbiosis es la reunión de dos estructuras que han evolucionado por caminos separados. El resultado es una estructura compuesta con un comportamiento diferente del que tenían las componentes por separado. La simbiosis permite que la evolución avance a pasos agigantados. Una criatura simbiótica puede saltar a un nivel superior de complejidad mucho más de prisa que una criatura que evoluciona mediante los procesos habituales de la mutación y la especiación.

En lo que sigue se describe brevemente el uso de metaheurísticas paralelas, tema que conforma el marco global de la investigación a ser abordada en este proyecto. Aunque el uso de metaheurísticas (simples o híbridas) permite reducir significativamente la complejidad temporal del proceso de búsqueda, este tiempo puede seguir siendo muy elevado en algunos problemas de interés real. Con el aumento de plataformas paralelas eficientes, la implementación paralela de estas metaheurísticas surge de forma natural como una alternativa para acelerar la obtención de soluciones precisas a estos problemas. La literatura es muy extensa en cuanto a la paralelización de metaheurísticas se refiere [Alba 2005], [Crainic y Toulouse 2003], [Cung et al. 2003], [Luna et al. 2006], [Luque y Alba 2011] ya que se trata de una aproximación que puede ayudar no sólo a reducir el tiempo de cómputo, sino a producir también una mejora en la calidad de las soluciones encontradas. Esta mejora está basada en un nuevo modelo de búsqueda que alcanza un mejor balance entre intensificación y diversificación. De hecho, muchos investigadores no utilizan plataformas paralelas de cómputo para ejecutar estos modelos paralelos y, aún así, siguen obteniendo mejores resultados que con los algoritmos secuenciales tradicionales. Tanto para las metaheurísticas basadas en trayectoria como para las basadas en población se han propuesto modelos paralelos acorde a sus características con la utilización no sólo de modelos paralelos, sino también de plataformas paralelas, ya que una simple evaluación del problema puede llevar minutos o incluso horas si se consideran problemas en los que intervienen, por ejemplo, complejas simulaciones.

En cuanto a los modelos paralelos de metaheurísticas basadas en trayectoria encontrados en la literatura se pueden clasificar, generalmente, dentro de tres posibles esquemas: ejecución en paralelo de varios métodos (modelo de múltiples ejecuciones [Alba et al. 2004], [Luna et al. 2005]), exploración en paralelo del vecindario (modelo de movimientos paralelos), y cálculo en paralelo de la función de fitness (modelo de aceleración del movimiento).

Teniendo en cuenta a los modelos paralelos de metaheurísticas basadas en población el paralelismo surge de manera natural cuando se trabaja con

poblaciones, ya que cada individuo puede manejarse de forma independiente. Debido a esto, el rendimiento de los algoritmos basados en población suele mejorar bastante cuando se ejecutan en paralelo. A alto nivel podemos dividir las estrategias de paralelización de este tipo de métodos en dos categorías: (1) paralelización del cómputo, donde las operaciones que se llevan a cabo sobre los individuos son ejecutadas en paralelo, y (2) paralelización de la población, donde se procede a la estructuración de la población. Uno de los modelos más utilizados que sigue la primera de las estrategias es el denominado maestro-esclavo (también conocido como paralelización global). En este esquema, un proceso central realiza las operaciones que afectan a toda la población mientras que los procesos esclavos se encargan de las operaciones que afectan a los individuos independientemente. Al margen del modelo maestro-esclavo, la mayoría de los algoritmos paralelos basados en población encontrados en la literatura utilizan alguna clase de estructuración de los individuos de la población. Este esquema es ampliamente utilizado especialmente en el campo de los algoritmos evolutivos y es el que mejor ilustra esta categorización. Entre los esquemas más populares para estructurar la población encontramos el modelo distribuido (o de grano grueso) [Alba 1999] y el modelo celular (o de grano fino) [Alba y Tomassini 2002]. En el caso de los algoritmos distribuidos la población se divide entre un conjunto de islas que ejecutan una metaheurística secuencial. Las islas cooperan entre sí mediante el intercambio de información (generalmente individuos, aunque nada impide intercambiar otro tipo de información). Esta cooperación permite introducir diversidad en las subpoblaciones, evitando caer así en los óptimos locales. Por otro lado, las metaheurísticas celulares [Alba y Dorronsoro 2008] se basan en el concepto de vecindario. Cada individuo tiene a su alrededor un conjunto de individuos vecinos donde se lleva a cabo la explotación de las soluciones. La exploración y la difusión de las soluciones, al resto de la población, se produce debido a que los vecindarios están solapados, lo que produce que las buenas soluciones se extiendan lentamente por toda la población. A parte de estos modelos básicos, en la literatura también se han propuesto modelos híbridos donde se implementan esquemas de dos niveles. Por ejemplo, una estrategia bastante común en la literatura es aquella donde en el nivel más alto tenemos un esquema de grano grueso, mientras que cada subpoblación se organiza siguiendo un esquema celular. A continuación se describe brevemente a los algoritmos evolutivos celulares (cEAs) ya que es una de las líneas principales de investigación y serán los primeros a utilizar como base para el diseño de nuevas propuestas algorítmicas. En los EAs celulares la población está

estructurada utilizando el concepto de vecindario [Alba y Tomassini 2002], de forma que los individuos sólo pueden interactuar con sus vecinos más próximos en la población.

En un cEA [Tomassini 2005], la población está usualmente estructurada en una rejilla (o malla) bidimensional de individuos. Los individuos sólo pueden interactuar con sus vecinos en el bucle reproductor que aplica los operadores de variación. El solapamiento de los vecindarios provee al cEA de un mecanismo implícito de migración, ya que las mejores soluciones se extenderán suavemente por toda la población, por lo que de esta forma se logra mantener la diversidad genética en la población por más tiempo con respecto a otros EAs no estructurados.

En lo relacionado a metaheurísticas paralelas aplicadas al problema de ruteo de vehículos existen muchas contribuciones y tendencias. Pueden encontrarse diferentes métodos paralelos para variantes del problema de ruteo de vehículos en [Crainic 2008] [Golden et al. 2008]. No obstante, el campo de las metaheurísticas paralelas es muy rico en cuanto a la variedad de metodologías utilizadas, y no todas las variantes del problema de ruteo de vehículos se han abordado. Por lo tanto es un desafío la exploración de algunas variantes de problema, la búsqueda de metodologías y estrategias de paralelización.

### 3. RESULTADOS OBTENIDOS/ESPERADOS

A continuación se mencionan los resultados esperados del proyecto:

- Investigar y estudiar la aplicación de técnicas paralelas en el ámbito de problemas de ruteo de vehículos.
- Realizar una descripción formal de una metodología para la incorporación de componentes activas de diferentes metaheurísticas en una metaheurística anfitriona.
- Desarrollar algoritmos avanzados y paralelos basados en inteligencia computacional para el problema de análisis.

### 4. FORMACION DE RECURSOS HUMANOS

Un integrante de este proyecto de investigación está desarrollando su Tesis de Doctorado en temáticas afines.

Tres integrantes han comenzado su Maestría orientando sus cursos y trabajos a esta línea de investigación.

Se cuenta con dos becarios de investigación de grado.

### 5. BIBLIOGRAFIA

- [Alba 1999] E. Alba. Análisis y Diseño de Algoritmos Genéticos Paralelos Distribuidos. PhD thesis, University of Málaga, 1999.
- [Alba y Tomassini 2002] E. Alba y M. Tomassini. Parallelism y evolutionary algorithms. *IEEE Transactions on Evolutionary Computation*, 6(5):443 – 462, 2002.
- [Alba et al. 2004] Alba, F. Luna, y A. J. Nebro. Advances in parallel heterogeneous genetic algorithms for continuous optimization. *International Journal of Applied Mathematics y Computer Science*, 14(3):101 – 117, 2004.
- [Alba 2005] E. Alba, editor. *Parallel Metaheuristics: A New Class of Algorithms*. Wiley, 2005.
- [Alba y Dorronsoro 2008] E. Alba y B. Dorronsoro. *Cellular Genetic Algorithms*. Springer 2008
- [Bäck et al. 1997] T. Bäck, D.B. Fogel, y Z. Michalewicz, editors. *Handbook of Evolutionary Computation*. Oxford University Press, 1997.
- [Blum y Roli 2003] C. Blum y A. Roli. Metaheuristics in combinatorial optimization: Overview y conceptual comparison. *ACM Computing Surveys*, 35(3):268–308, 2003.
- [Cerney 1985] V. Cerney. A thermodynamical approach to the travelling salesman problem: an efficient simulation algorithm. *Optimization Theory an Applications*, 45, 1985.
- [Cook et al. 1998] W.J. Cook, W.H. Cunningham, W.R. Pulleyblank, y A. Schrijver. *Combinatorial Optimization*. Wiley Interscience, 1998.
- [Crainic y Toulouse 2003] T. G. Crainic y M. Toulouse. Parallel strategies for metaheuristics. In F. W. Glover y G. A. Kochenberger, editors, *Handbook of Metaheuristics*, Norwell, MA, USA, 2003. Kluwer Academic Publishers.
- [Crainic 2008] Crainic, T.G. Parallel solution methods for vehicle routing problems. *The Vehicle Routing Problem: Latest Advances y New Challenges*. Pages 171-198, 2008, Springer.
- [Cung et al. 2003] V.-D. Cung, S. L. Martins, C. C. Ribeiro, y C. Roucairol. Strategies for the Parallel Implementation of Metaheuristics. In C.C. Ribeiro y P. Hansen, editors, *Essays y Surveys in Metaheuristics*, pages 263–308, Norwell, MA, USA, 2003. Kluwer Academic Publishers.
- [Dantzing y Ramster 1959] G.B. Dantzing y R.H. Ramster, “The truck dispatching problem,” *Management Science*, vol. 6, pp. 80–91, 1959.
- [Eberhart y Kennedy 1995] R. Eberhart y J. Kennedy. A new optimizer using particles swarm theory. In *Sixth International Symposium on Micro Machine y Human Science (Nagoya, Japan)*, IEEE Service Center, Piscataway, pages 39–43, 1995.
- [Feo y Resende 1995] T.A. Feo y M.G.C. Resende. Greedy randomized adaptive search procedures. *Journal of Global Optimization*, 6:109–133, 1995

- [Golden et al. 2008] Golden, B.L. y Raghavan, S. y Wasil, E.A. The vehicle routing problem: latest advances y new challenges. Vol. 43. 2008. Springer Verlag.
- [Glover 1986] F. Glover. Future paths for integer programming y links to artificial intelligence. *Computer y Operations Research*, 13:533–549, 1986.
- [Glover y Laguna 1993] F. Glover y M. Laguna. Tabu search in Modern Heuristic Techniques for Combinatorial Problems. John Wiley y Sons, 1993.
- [Golden et al. 1998] B.L. Golden, E.A. Wasil, J.P. Kelly, y I-M. Chao, Fleet Management y Logistics, chapter The Impact of Metaheuristics on Solving the Vehicle Routing Problem: algorithms, problem sets, y computational results, pp. 33–56, Kluwer, Boston, 1998.
- [Hansen y Mladenovic 1997] P. Hansen y N. Mladenovic. Variable neighborhood search for the p-median. *Location Science*, 5:207–226, 1997.
- [Kirkpatrick et al. 1983] S. Kirkpatrick, J. Gelatt, y M. Vecchi. Optimization by simulated annealing. *Science*, 220:671–680, 1983.
- [Lenstra y Rinnooy Kan 1981] J.K. Lenstra y A.H.G. Rinnooy Kan, “Complexity of vehicle routing y scheduling problems,” *Networks*, vol. 11, pp. 221–227, 1981.
- [Luque y Alba 2011] G. Luque y E. Alba, “Parallel Genetic Algorithms”, Springer, 2011.
- [Luna et al. 2005] F. Luna, E. Alba, y A. J. Nebro. Parallel heterogeneous metaheuristics. In E. Alba, editor, *Parallel Metaheuristics*, pages 395 – 422. Wiley, 2005.
- [Luna et al. 2006] F. Luna, A. J. Nebro, y E. Alba. Parallel evolutionary multiobjective optimization. In N. Nedjah, E. Alba, y L. de Macedo, editors, *Parallel Evolutionary Computations*, volume 22 of *Studies in Computational Intelligence*, chapter 2, pages 33 – 56. Springer, 2006.
- [Papadimitriou y Steiglitz 1998] C. Papadimitriou y K. Steiglitz. *Combinatorial Optimization*. Dover Publications, 1998.
- [Resende y Ribeiro 2003] M.G.C. Resende y C.C. Ribeiro. *Handbook of Metaheuristics*, chapter Greedy randomized adaptive search procedures, pages 219–249. Kluwer Academic Publishers, 2003.
- [Talbi 2002] E.-G. Talbi. A taxonomy of hybrid metaheuristics. *Heuristics*, (8):541–564, 2002.
- [Talbi 2009] E.-G. Talbi. *Metaheuristics: From design to Implementation*. Wiley, 2009.
- [Tomassini 2005] M. Tomassini. *Spatially Structured Evolutionary Algorithms: Artificial Evolution In Space y Time*. Natural Computer. Springer-Verlag, Heidelberg, 2005.
- [Toth y Vigo 2001] P. Toth y D. Vigo, *The Vehicle Routing Problem*, Monographs on Discrete Mathematics y Applications. SIAM, Philadelphia, 2001.



# Redefinición de un Framework que formaliza la Argumentación basada en Expertos

Paola D. Budán<sup>1,2,3</sup>

Guillermo R. Simari<sup>1,2</sup>

<sup>1</sup>Laboratorio de Investigación y Desarrollo en Inteligencia Artificial

<sup>2</sup>Departamento de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur,  
Av. Alem 1253, (B8000CPB) Bahía Blanca, Argentina

<sup>3</sup>Departamento de Informática, Universidad Nacional de Santiago del Estero,  
Av. Belgrano 1912, (4200CPB) Santiago del Estero, Argentina  
e-mail: pbudan@unse.edu.ar - grs@cs.uns.edu.ar

## Resumen

Desde la Argumentación, la información es evidencia que prueba un punto de vista, para tomar una decisión, o para convencer a las personas de pensar o actuar de cierta manera. El término argumento se usa en un sentido especial, para aludir a las razones para apoyar o rebatir cierta afirmación, que es cuestionable o dudosa [6]. Una de las formas de representar los argumentos son los denominados *Esquemas de Argumentación*, en las cuales los argumentos representan estructuras inferenciales utilizadas en el discurso cotidiano. Entre estos esquemas de argumentación se encuentran los *Esquemas que apelan a la Opinión del Experto o Esquemas Basados en la Posición del Saber*[9, 10]. En un trabajo previo [5] se propuso un *framework* denominado *Framework de Argumentación Experto (EAF)*. El mismo está basado en una 4-tupla  $\theta = \langle AR, EXP, ATTACK, ASSERT \rangle$ . EAF permite representar a un experto, las propiedades de dicho experto, y en particular la asociación de un experto con un argumento. En esta línea de investigación se pretende refinar la definición de la 4-tupla en el conjunto que contiene las características del experto, de manera que

dichas características puedan ser ponderadas.

**Palabras clave:** Argumentación - Argumentos - Esquemas basados en la Opinión del Experto - Framework de Argumentación

## Contexto

Esta línea de investigación está inserta en el marco del desarrollo de una tesis doctoral para optar por el título de Doctor en Ciencias de la Computación del Departamento de Ciencias de la Computación de la Universidad Nacional del Sur, trabajo éste que se denomina “Formalización de las estructuras de los Esquemas de Argumentación”. Se llevará a cabo dentro del ámbito del Laboratorio de Investigación y Desarrollo en Inteligencia Artificial (LIDIA), y está asociada a los siguientes proyectos de investigación:

- “Sistemas De Apoyo a la Decisión Basados en Argumentación: formalización y aplicaciones”. PIP-CONICET (PIP-112-200801-02798). Director: Carlos Iván Chesñear. Período 01/2009 - 12/2011. CONICET (aún en vigencia).
- “Formalismos Argumentativos aplicados a Sistemas Inteligentes para Toma de Deci-

siones”. Código: PGI 24/ZN18. Director: A. J. García. Co-director: M. A. Falappa. Acreditado con evaluación externa para el período 01/2009 - 12/2011. Universidad Nacional del Sur (aún en vigencia).

- “Representación de Conocimiento y Razonamiento Argumentativo: Herramientas Inteligentes para la Web y las Bases de Datos Federadas”. 24/N030, 01/01/11 - 31/12/2014.

## Introduccion

Las investigaciones en argumentación contribuyen a la formalización del razonamiento de sentido común. En un sentido general, la argumentación se puede asociar con el proceso en el cual se ofrecen argumentos a favor y en contra de las conclusiones, con el fin de determinar qué conclusiones son aceptables. Existen varios formalismos basados en argumentos con aplicaciones en muchas áreas, tales como el razonamiento jurídico, los sistemas de recomendación y los sistemas multi-agente. [4, 1] Los esquemas de argumentación [9, 10] ofrecen la posibilidad de representar los argumentos en una manera semi-formal. Este sencillo dispositivo permite representar argumentos en una forma cercana al lenguaje natural, y es útil generalmente para ser utilizada en textos libres. Un tipo particular de esquemas de argumentación corresponde al Esquema de Argumentación basado en la Opinión del Experto, o en la Posición del Saber. La mayoría de los frameworks de argumentación están basados en el trabajo de Dung [3], quien propuso un *Framework Abstracto de Argumentación (AF)*. Tomando como base el AF de Dung es que se propuso un framework denominado *Expert Argumentation Framework (EAF)*. El EAF permite representar a un experto, las propiedades de dicho experto, y la asociación en particular de un experto con un argumen-

to. El mismo está representado en una 4-tupla  $\theta = \langle AR, EXP, ATTACK, ASSERT \rangle$ . En esta 4-tupla, EXP representa el conjunto de expertos o proponentes, ASSERT representa una función que relaciona los expertos con los argumentos. En la definición del conjunto EXP se consideraron las seis cuestiones críticas sostenidas por Walton al momento de definir este esquema. En el EAF, estas características pueden tomar el valor 1 (uno) si están presentes en el Experto, o 0 (cero) en caso contrario. Con el desarrollo de esta investigación se pretende refinar la definición del conjunto de características del experto, de manera que las mismas puedan ser ponderadas, en vez de limitarse a los valores binarios.

## Líneas de investigación y desarrollo

La presente línea de investigación estudiará la formalización del esquema de argumentación propuesto y evaluará la necesidad de ponderar las características del experto. Para ello:

- Se analizará si todas las características propuestas por Walton para el experto están realmente contempladas en el framework.
- Se aplicará el EAF a casos de prueba.
- Se evaluarán las situaciones en las cuales la propuesta de formalización presentada no sea adecuada.
- Se definirá una nueva formalización para las características del experto.

## Argumentos

Un argumento es un par  $\langle \Phi, \alpha \rangle$ , donde el primer elemento del par es un conjunto mínimo y consistente de fórmulas que confirman

(prueban) el segundo elemento. Es decir, se cuenta con el respaldo y la afirmación de un argumento, aunque no se indique el método de inferencia ya que dicho método no difiere de un argumento a otro [6]. Para determinar si un argumento es sólido, tiene que ser sometido a las siguientes preguntas: otra persona que disponga de la misma información podría llegar a una conclusión diferente? Si la respuesta a esta pregunta es afirmativa, el argumento dado es débil. ¿Por qué puede alguien llegar a una conclusión diferente si está basado en la misma información? ¿Qué hipótesis diferentes manejan las dos personas que partiendo de la misma información llegan a conclusiones distintas?. El último punto a evaluar es que tan sólidas son las afirmaciones en las cuales se basa el argumento. Estas afirmaciones, ¿proviene de conocimiento común?, ¿de la opinión de un experto?, ¿de una estadística?, ¿del conocimiento o la experiencia personal?, ¿por qué alguien debería aceptarlas? [7]. En definitiva, un argumento es un conjunto de fórmulas apropiadas que pueden ser utilizadas para probar una afirmación. Un argumento es un subconjunto de un conjunto  $\Delta$ , y el conjunto  $\Delta$  es gran repositorio de información, desde el cual se pueden construir argumentos por y en contra de las afirmaciones arbitrarias. No se espera que  $\Delta$  sea consistente. Ni siquiera sus fórmulas individuales deben ser consistentes. Las fórmulas de  $\Delta$  pueden representar información cierta o incierta, y pueden representar declaraciones subjetivas, objetivas o hipotéticas. Así  $\Delta$  puede representar hechos, creencias, vistas, etc. de diferentes agentes que no tienen la misma opinión [2]. En síntesis, un argumento es una pieza de información perteneciente a un repositorio de información.

### Esquemas de Argumentación

Los esquemas de argumentación son las formas abstractas de argumentos más

comúnmente usadas en la conversación cotidiana, y en otros contextos como el científico y el legal. Constituyen una manera de argumentar que capturan los patrones estereotipados del razonamiento humano [10]. La mayoría de estos esquemas son formas plausibles de razonamiento que no se ajustan a las formas tradicionales de argumentos deductivos o inductivos. Dentro de los esquemas de argumentación se encuentran: los que recurren a la opinión del experto, los que se basan en la opinión del experto, los que argumentan desde la analogía, los que tienen en cuenta la correlación entre causas-efectos, basados en signos, entre otros. Cada esquema contiene un conjunto de preguntas claves que hacen coincidir el esquema que representan con una manera estándar de sondear críticamente un argumento para encontrar sus potenciales puntos débiles. La motivación original de los esquemas fue enseñar a los estudiantes universitarios habilidades para el desarrollo de un pensamiento crítico. Por ello, deberían ser formalizados de otra manera, para que se vuelvan más útiles en el campo de la Inteligencia Artificial [8].

### Argumentación Abstracta

Phan Minh Dung desarrolló una abstracción de la argumentación rebatible mediante los *frameworks de argumentación* (Argumentation Framework - AF). En un AF la estructura de los argumentos permanece sin especificar y el análisis se centra en la interacción existente entre los mismos. Un framework de argumentación se define como un par compuesto por un conjunto de argumentos y una relación binaria que representa los ataques entre argumentos. Un argumento es una entidad abstracta que cumple un determinado rol al estar relacionado con otros argumentos. Formalmente, un framework de argumentación está definido como un par  $AF = (AR, ATTACKS)$  en

donde  $AR$  es un conjunto de argumentos y  $ATTACKS$  es una relación binaria en  $AR$ , es decir,  $ATTACKS \subseteq AR \times AR$ , y representa la relación de ataques entre argumentos. En este marco, un argumento  $A$  se dice *acceptable* si se puede defender de todos los ataques. Sea un  $AF = (AR, ATTACKS)$  un framework de argumentación:

- Un conjunto  $S \subseteq AR$  se denomina *libre de conflicto* si no existen dos argumentos  $A, B \in S$  tal que  $(A, B) \in ATTACKS$ .
- Un argumento  $A \in AR$  es acceptable con respecto a  $S \subseteq AR$  si y solo si para cada  $B \in AR$  se cumple que si  $B$  ataca a  $A$  entonces  $B$  es atacado por  $S$ .
- Un conjunto libre de conflictos  $S \subseteq AR$  es admisible si y solo si cada argumento en  $S$  es acceptable con respecto a  $S$ .
- Un conjunto admisible  $E \subseteq AR$  es una extensión completa de  $AF$  sí y sólo sí  $E$  contiene cada argumento que es acceptable con respecto a  $E$ .

### Framework de Argumentación Experto (EAF)

El EAF sobre el cual se basa esta investigación, toma en cuenta las cualidades o características del experto que esgrime los argumentos y se define como una 4-tupla  $\theta = \langle AR, EXP, ATTACK, ASSERT \rangle$  donde:

- $AR$  es un conjunto de argumentos.
- $ATTACK$  es una relación binaria de ataque entre argumentos que pertenecen al conjunto  $AR$ ,  $ATTACK \subseteq AR \times AR$ .
- $EXP$  es el conjunto de expertos o propo- nentes que esgrimen un argumento.

- $ASSERT$  es una función definida de la siguiente manera:  $ASSERT : EXP \rightarrow AR$  y determina qué experto esgrime cuál argumento.

El componente  $EXP$  de la tupla se define de la siguiente manera, siguiendo los criterios estipulados por Walton: Sea  $EXP$  un conjunto de expertos, se define un experto  $X_i \in EXP$  como una 4-tupla  $X_i = (Q, C, B, S)$  donde:

- $Q$  denota que el argumento proviene de un experto cualificado (qualified) en el asunto.
- $C$  denota que el argumento proviene de una fuente confiable.
- $B$  significa que el argumento del experto es mejor que cualquier otro argumento de cualquier otro experto (better).
- $S$  el argumento es seguro, capaz de sobreponerse a cualquier ataque.

Cada argumento de  $AR$  tiene asociado unas propiedades  $Q, C, B, S$ . La presencia de una propiedad se representa con un 1 y la ausencia de la propiedad se representa con un 0. A partir de estas consideraciones, se expande la definición de Dung, ampliando las definiciones de conjunto libre de conflictos y elementos aceptables:

- Un conjunto  $S \subseteq AR$  se denomina *libre de conflicto* si no existen dos argumentos  $A, B \in S$  tal que  $(A, B) \in ATTACKS$ , ya sea que sean esgrimidos por el mismo o por diferentes expertos.
- Un argumento  $A \in AR$  es acceptable con respecto a  $S \subseteq AR$  si y solo si para cada  $B \in AR$  se cumple que si  $B$  ataca a  $A$  entonces los expertos que esgrimen los argumentos que se atacan son diferentes.



## Resultados y Objetivos

El objetivo general de esta línea de investigación es estudiar en el contexto presentado, la suficiencia de la extensión al framework de Dung propuesta para formalizar los Esquemas Basados en la Opinión del Experto. El objetivo particular del plan es desarrollar una serie de pruebas que permitan verificar si todas las características que debe reunir el experto están contempladas en el EAF, y si las mismas necesitan ser expresadas de alguna manera que permita ponderarlas.

## Formación de Recursos Humanos

Actualmente el equipo de trabajo de esta línea de investigación se encuentra compuesto una estudiante de posgrado de la Universidad Nacional de Bahía Blanca y su director. Por otra parte se vincula con un grupo de trabajo sobre argumentación compuesto por doctorandos e investigadores formados.

## Referencias

- [1] García A. and Simari G. *Defeasible Logic Programming An Argumentative Approach*. 2003.
- [2] A. Maguitman C. Chesñevar and P. Ronald. *Logical models of argument*. ACM Computing Surveys, 32(4),337-383, 2000.
- [3] Phan Minh Dung. *On the acceptability of arguments and its fundamental rol in monotonic reasoning, logic programming and n-person games*.
- [4] A. J. García and G. R. Simari. *Defeasible logic programming: An argumentative approach*. Journal Theory and Practice of Logic Programming archive Volume 4 Issue 2, NY, USA, 2004.
- [5] G. Simari P. Budán, M. Budán. *A Preliminary Study of Argumentation Frameworks and Argumentation Schemes that appeal to Expert Opinion*.
- [6] I. Rahwan and G. Simari (eds.). *Argumentation in Artificial Intelligence*. Springer-Verlag, 2009.
- [7] Yanna Rider. *The critical edge*. <http://www.theage.com.au/news/education-news/the-critical-edge/2007/03/23/1174597895195.html>, 2007.
- [8] Walton and Gordon. *Visualization Tools, Argumentation Schemes and Expert Opinion Evidence in Law*. Oxford University Press, 2007.
- [9] Douglas Walton. *Justification of Argumentation Schemes*. Received by Greg Restall, <http://www.philosophy.unimelb.edu.au/ajl/2005>, 2005.
- [10] Douglas Walton. *Fundamentals of Critical Argumentation*. Cambridge University Press, Cambridge. UK, 2006.

## Un enfoque para el debilitamiento y fortalecimiento entre argumentos

Maximiliano C. D. Budán      Mauro Gomez Lucero      Guillermo R. Simari

Laboratorio de Investigación y Desarrollo en Inteligencia Artificial  
 Departamento de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur,  
 Av. Alem 1253, (B8000CPB) Bahía Blanca, Argentina  
 Departamento de Informática, Universidad Nacional de Santiago del Estero,  
 Av. Belgrano 1912, (4200CPB) Santiago del Estero, Argentina  
 e-mail: {mcd,mg, grs}@cs.uns.edu.ar

### Resumen

El objetivo general de esta línea de investigación es estudiar el fortalecimiento y debilitamiento de los argumentos en el contexto de los sistemas argumentativos (AS), añadiendo un meta-nivel de información a los argumentos en la forma de etiquetas extendiendo así las capacidades de representación de la estructura. Las etiquetas serán utilizadas para introducir la representación de incertidumbre, fiabilidad, disponibilidad de tiempo, o cualquier otra característica de importancia para los argumentos. Uno de los objetivos particulares es la formalización y definición de un marco argumentativo, para representar las características especiales de los argumentos, el fortalecimiento y debilitamiento entre los mismos, y utilizar la información que contienen las etiquetas en el proceso para obtener el conjunto de argumentos aceptados.

**Palabras clave:** Sistemas Argumentativos, Etiquetas, Fortalecimiento, Debilitamiento.

### Contexto

El presente plan de trabajo está estrechamente relacionado con los siguientes

proyectos de investigación que se desarrollan dentro del ámbito del Laboratorio de Investigación y Desarrollo en Inteligencia Artificial (LIDIA):

- “Sistemas De Apoyo a la Decisión Basados en Argumentación: formalización y aplicaciones”. PIP-CONICET (PIP-112-200801-02798). Director: Carlos Iván Chesñevar. Período 01/2009 - 12/2011. CONICET (aún en vigencia).
- “Formalismos Argumentativos aplicados a Sistemas Inteligentes para Toma de Decisiones”. Código: PGI 24/ZN18. Director: A. J.García. Co-director: M.A. Falappa. Acreditado con evaluación externa para el período 01/2009 - 12/2011. Universidad Nacional del Sur (aún en vigencia).
- “Representación de Conocimiento y Razonamiento Argumentativo: Herramientas Inteligentes para la Web y las Bases de Datos Federadas”. 24/N030, 01/01/11 – 31/12/2014.

## Introducción

Dentro de las Ciencias de la Computación, el problema de la representación del conocimiento ha sido abordado mayormente desde el campo de investigación de la Inteligencia Artificial (IA) estudiando diferentes modelos de representación del conocimiento y del tratamiento de dicho conocimiento con un determinado fin, por ejemplo, tomar decisiones de acuerdo a cambios en el entorno.

La argumentación ha contribuido a la formalización de un modelo de razonamiento, similar al razonamiento humano. En un sentido general, la argumentación puede estar asociada a la interacción de los argumentos a favor y en contra de ciertas conclusiones, con el fin de determinar qué conclusiones son aceptables. Varios formalismos basados en la teoría de la argumentación han surgido, con aplicaciones en muchas áreas, tales como el razonamiento legal, sistemas de recomendación, agentes autónomos y sistemas multi-agente [1, 7, 10, 3, 4].

En ciertas aplicaciones de la argumentación, sería interesante añadir meta-nivel de información a los argumentos en la forma de etiquetas para extender así las capacidades de representación de la estructura. La razón de esta extensión es que la solidez de un argumento no es el único factor a tener en cuenta [2]; argumentos puede tener asociada una característica particular que determina la fuerza de un argumento. De esta forma, las etiquetas pueden ser utilizadas para introducir la representación de incertidumbre, fiabilidad, o cualquier otra característica que sea de importancia para un modelo de representación en particular y ayudar en el proceso de obtener el conjunto de argumentos aceptados.

Recientemente, Bench-Capon y Pollock en [2, 8], han distinguido tendencias que están influenciando actualmente la investigación sobre este tema en Inteligencia Artificial. Bench-

Capon en sus investigaciones sostiene que en muchos casos de desacuerdo, sobre todo en situaciones de razonamiento práctico, es imposible demostrar de manera concluyente que ninguna de las partes está mal. El papel de la argumentación en estos casos es el de persuadir en lugar de probar, demostrar o refutar. En estos casos, la fuerza de un argumento depende de los valores sociales que tenga atribuido los argumentos, por lo que, el éxito del ataque de un argumento a otro depende de la comparación de esas fuerzas. Por otro lado, Pollock analiza que en la mayoría de semántica argumentativas ignoran el hecho de que algunos argumentos son mejores que otros, apoyando sus conclusiones con más fuerza. Pero una vez que reconocemos que los argumentos pueden variar en fuerza, es decir, pueden diferir en su grado de justificación, es más difícil la definición de que argumentos son aceptados y con qué grado de justificación.

Tomando como base estas investigaciones, establecemos como objetivo la construcción de un marco argumentativo, en el cual las etiquetas son la forma de representar las características particulares de los argumentos. La interacción entre los argumentos pueden afectar las etiquetas que tienen asociados, por lo que, estos cambios puede producir fortalecimientos y debilitamientos entre argumentos. Es importante tener en cuenta que las características o propiedades asociadas a un argumento podrían variar en el tiempo y ser influenciadas o afectadas por distintas fuerzas que gobiernan el mundo real, por ejemplo, la confiabilidad de una determinada fuente de información.

## Líneas de Investigación y Desarrollo

La presente línea de investigación estudiará en el contexto de la argumentación, la problemática asociada con la representación

del conocimiento y el tratamiento de características especiales asociadas al mismo. En particular se formalizara e implementara un marco argumentativo, que permita la representación de características asociadas a los argumentos, y el modelado de fortalecimiento y debilitamiento entre argumentos.

## Argumentación

Aunque en la teoría de la argumentación se incluye el debate y la negociación, las cuales están dirigidas a alcanzar unas conclusiones de mutuo acuerdo aceptables, su principal motivación es quizás el debate social en el que la victoria sobre un oponente es el principal objetivo. Este arte y ciencia es con frecuencia el medio por el cual algunas personas protegen sus creencias o propios intereses en un diálogo racional, en simples coloquios o durante el proceso de argumentación o defensa de ideas. La argumentación es usada en los juicios para probar y/o refutar la validez de ciertos tipos de evidencias. Los estudiantes de argumentación estudian las racionalizaciones post hoc mediante las cuales un individuo puede justificar decisiones que originalmente pudieron haber sido realizadas de forma irracional [12].

Desde la antigüedad, la argumentación ha sido objeto de interés en todas las áreas donde se practica el arte de hablar y de escribir de manera persuasiva. En la actualidad, el estudio de la argumentación ha recobrado vigencia debido a la gran influencia que los medios de comunicación tienen sobre la sociedad. Esta influencia se manifiesta en el planteamiento de estrategias argumentativas para convencer al público acerca de ciertos valores e ideas. Ejemplo de esto son los discursos argumentativos relacionados con la publicidad o el pensamiento político. Así pues, la principal motivación del estudio de la argumentación (por parte de los argumentadores), consiste en establecer si

el razonamiento planteado es verosímil, es decir, si quien es objeto de la argumentación estará dispuesto a aceptarla.

## Sistemas Argumentativos

En argumentación es una proposición es aceptada o no de acuerdo a un análisis de las razones de las que se dispone para creer o no en la misma, donde estas razones o justificaciones toman la forma de argumentos [5]. Además, la manera en que estos argumentos son considerados permite la automatización de este tipo de razonamiento. En los sistemas argumentativos basados en reglas (SABR), existe un conjunto de reglas de inferencia con las cuales, a partir de cierta información (antecedente o conjunto de premisas) se puede inferir de manera tentativa nueva información (consecuente). En este tipo de sistemas, las reglas son almacenadas en una base de conocimiento, junto a otra información en forma de hechos o presuposiciones, que representan la evidencia que el agente obtiene de su entorno. A partir de esta evidencia, el agente puede usar las reglas de inferencia para construir argumentos a favor o en contra de una afirmación. Una vez hecho esto se evalúan todos los argumentos construidos, y se determina cuáles de ellos son aceptados, buscando concluir si, a partir de la base de conocimiento del agente, está afirmación puede aceptarse o no. Estos formalismos son no-monótonos dado que la introducción de nueva información al sistema puede generar nuevos argumentos que resultan contradictorios con algunos de los ya existentes. Como resultados de esta investigación se incorporo a los SABR el tratamiento de reglas que están activas en determinados intervalos de tiempo que brinda la posibilidad de modelar conocimientos que esta disponible solo en determinado intervalos de tiempo aumentando así la capacidad de representación del mundo real y mo-



delos mas simples de analizar. En general, en la mayoría de estos formalismos, argumentos y contra-argumentos son comparados utilizando un criterio de preferencia pre-determinado permitiendo decidir si un ataque tiene éxito. En el últimos años, el campo de aplicación de la argumentación se ha expandido velozmente, en gran parte debido a los avances teóricos, pero también gracias a la demostración exitosa de su uso práctico en un gran número de dominios de aplicación, tales como el razonamiento legal [9], los sistemas multi-agentes [11], entre muchos otros.

## Argumentos y Etiquetas

En ciertas aplicaciones de la argumentación, será beneficioso añadir meta-nivel de información a los argumentos. Por esta razón, las etiquetas extenderían las capacidades de representación de la estructura. La razón de esta extensión es que la solidez de un argumento no es la única consideración; argumentos puede tener asociada una característica particular que determina la fuerza de un argumento. Entonces, las etiquetas pueden ser utilizadas para introducir la representación de incertidumbre, fiabilidad, o cualquier otra característica que puede ayudar en el establecimiento de una característica importante de los argumentos y ayudar en el proceso de obtener el conjunto de argumentos aceptados.

## Algebra de Etiquetas

En el álgebra de etiquetas, definiremos las operaciones entre las etiquetas. Las etiquetas puede ser usada para diferentes fines dentro del contexto de los sistemas de argumentación, tales como: establecer el grado de fuerza (define una preferencia entre los créditos), definen debilitamiento y fortalecimiento de las operaciones entre las etiquetas asociadas a las con-

clusiones, y establecer la condición de aceptabilidad para los argumentos.

## Objetivos y Resultados Esperados

En el LIDIA a través de los años se han llevado a cabo diferentes proyectos sobre Sistemas de Argumentación, en particular investigaciones dedicadas a desarrollar sistemas de argumentación masiva. Varias publicaciones proponiendo la creación de mecanismos que pudieran mejorar la complejidad computacional de los sistemas de argumentación basados en Defeasible Logic Programing (DeLP) [6] fueron propuestas y publicadas en conferencias y revistas internacionales. El objetivo general de esta línea de investigación, es mostrar que las etiquetas se pueden utilizar para introducir la representación de incertidumbre, fiabilidad, o cualquier otra característica que puede ayudar en el establecimiento de una característica importante de los argumentos y ayudar en el proceso de obtener el conjunto de argumentos aceptado. Los objetivos específicos de esta línea de investigación consiste en:

- La formalización de un marco de argumentativo con la capacidad de representar y tratar características asociadas a los argumentos;
- Posibilitando el debilitamiento y el fortalecimiento entre argumentos;

Para representar estas características particulares asociadas a los argumentos se utilizara una estructura de etiquetas. Para la manipulación de las etiquetas, se definirá un algebra de etiquetas, la cual establecerá como serán las operaciones de fortalecimiento, debilitamiento y propagación entre etiquetas.

## Formación de Recursos Humanos

Actualmente el equipo de trabajo de esta línea de investigación se encuentra compuesto por un becario de posgrado que accedió a una beca interna del CONICET y su director. Por otra parte se vincula con un grupo de trabajo sobre argumentación y agentes inteligentes compuesto por doctorandos e investigadores formados.

## Referencias

- [1] Leila Amgoud and Henri Prade. *Using arguments for making and explaining decisions*, volume 173. Elsevier Science Publishers Ltd., Essex, UK, 2009.
- [2] Trevor J. M. Bench-Capon. *Persuasion in Practical Argument Using Value-based Argumentation Frameworks*. 2003.
- [3] Cristian E. Briguez, Maximiliano Celmo Budán, Cristhian A. D. Deagustini, Ana Gabriela Maguitman, Marcela Capobianco, and Guillermo Ricardo Simari. *Towards an Argument-based Music Recommender System*. 2012.
- [4] Maximiliano C. D. Budán, Mauro Gómez Lucero, Carlos I. Chesñevar, and Guillermo R. Simari. *Modeling Time and Reliability in Structured Argumentation Frameworks*. 2012.
- [5] A. Maguitman C. Chesñevar and P. Ronald. *Logical models of argument*. ACM Computing Surveys, 32(4),337-383, 2000.
- [6] A. J. García and G. R. Simari. *Defeasible logic programming: An argumentative approach*, *Journal of Theory and Practice of Logic Programming*. 4 (1), pp. 95-138., 2004.
- [7] Philippe Pasquier, Ramon Hollands, Iyad Rahwan, Frank Dignum, and Liz Sonenberg. *An empirical study of interest-based negotiation*, volume 22. 2011.
- [8] John L. Pollock ? *Defeasible Reasoning and Degrees of Justification*, volume 1. 2010.
- [9] H. Prakken and G. Sartor. *The role of logic in computational models of legal argument - a critical survey*. In: A. Kakas, F. Sadri (eds.), *Computational Logic: Logic Programming and Beyond*, Springer, pp. 342-380, 2002.
- [10] Iyad Rahwan, Sarvapali D. Ramchurn, Nicholas R. Jennings, Peter Mchurney, Simon Parsons, and Liz Sonenberg. *Argumentation-based negotiation*, volume 18. Cambridge University Press, New York, NY, USA, December 2003.
- [11] C. Sierra S. Parsons and N. Jennings. *Agents that Reason and Negotiate by Arguing*. *Journal of Logic and Computation* 8,pp. 261-292., 1998.
- [12] Charles Arthur Willard. *A theory of argumentation*. University of Alabama Press, c1989, 1989.

# Integración de Ontologías Datalog+/-

Cristhian A. D. Deagustini    María Vanina Martínez    Marcelo A. Falappa  
Guillermo R. Simari

Laboratorio de Investigación y Desarrollo en Inteligencia Artificial  
Departamento de Ciencias e Ingenierías de la Computación  
Universidad Nacional del Sur  
Alem 1253 - Bahía Blanca - Buenos Aires - Argentina  
(0291) 459-5135

cadd@cs.uns.edu.ar, vanina.martinez@cs.ox.ac.uk, mfalappa@cs.uns.edu.ar, grs@cs.uns.edu.ar

## Resumen

En los últimos tiempos, la colaboración y el intercambio de información se han vuelto aspectos cruciales de muchos sistemas. En estos entornos es de vital importancia definir métodos automáticos para resolver conflictos entre el conocimiento compartido por distintos sistemas. Este conocimiento es frecuentemente expresado a través de ontologías que pueden ser compartidas por los sistemas que utilizan el mismo.

En la presente investigación se busca la definición de métodos automáticos de integración de ontologías Datalog+/- . En base a lo logrado en este aspecto se buscará la adaptación del framework desarrollado para su aplicación tanto en la creación de federaciones de Bases de Datos (Data Federation) como en el intercambio de datos (Data Exchange). En estos campos de aplicación estos métodos podrán contribuir brindando la posibilidad de obtener de forma automática un esquema universal que respete tanto como sea posible a los originales manteniendo la coherencia del mismo con respecto a las restricciones de integridad impuestas a los datos, y definiendo que datos pueden ser mantenidos en la federación resolviendo incoherencias en el proceso.

Adicionalmente, se analizarán posibles extensiones a Datalog+/- basadas en formalismos de Argumentación Rebatible, teniendo en cuenta aspectos como la definición de relaciones de inferencia para estas ontologías aumentadas que tengan en cuenta los aspectos no-monótonos de la Argumentación Rebatible, o el impacto de tales relaciones en las conclusiones finales obtenidas y la complejidad de la obtención de las mismas.

**Palabras Clave:** Integración de Bases de Conocimiento, Revisión de Creencias, Representación de Conocimiento, Razonamiento, Argumentación Rebatible.

## 1. Contexto

Esta línea de investigación se lleva a cabo en el marco de los siguientes proyectos de investigación:

- **“Representación de conocimiento y razonamiento argumentativo: Herramientas inteligentes para la web y las bases de datos federadas”**. Director: Guillermo R. Simari. PGI 24/N030. Unidad coordinadora: Universidad Nacional del Sur.
- **“Combinación de Revisión de Creencias y Argumentación para mejorar las capacidades de Razonamiento y modelado de la Dinámica de Conocimiento en Sistemas Multi-agente”**. Director: Marcelo A. Falappa. PIP 112-20110101000. Unidad coordinadora: Consejo Nacional de Investigaciones Científicas y Técnicas.
- **“Desarrollo de Sistemas de Argumentación Masiva sobre Base de Datos Federadas”**. Director: Guillermo Simari. Co-director: Cristian Pacífico. PID-UNER 7041. Unidad coordinadora: Universidad Nacional de Entre Ríos.

Este último proyecto se enmarca dentro del ámbito de colaboración entre el Laboratorio de Investigación y Desarrollo (LIDIA) del Dep. de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur; y el Laboratorio de Investigación y Desarrollo en Inteligencia Artificial Concordia (LIDIA Concordia) de la Facultad de Ciencias de la Administración, Universidad Nacional de Entre Ríos.

## 2. Introducción

En los últimos tiempos la integración e interacción entre diferentes sistemas se ha vuelto muy común, especialmente en entornos colaborativos introducidos desde el arribo de la Web Semántica [BLHL01], *e. g.*, e-commerce.

Sin embargo, la colaboración entre sistemas no siempre puede realizarse de manera directa. Hay ocasiones donde problemas de inconsistencia (e incoherencia) aparecen cuando tomamos el conocimiento provisto por diferentes entidades como uno solo. En formalismos de representación de conocimiento como las ontologías estas inconsistencias suelen aparecer como violaciones a las

restricciones de integridad impuestas a los datos en las fuentes originales del conocimiento, que son violadas en el conocimiento integrado.

La resolución de incoherencias e inconsistencias en conocimiento es admitido como un problema importante que debe ser atacado [GCS10, HvHH<sup>+</sup>05, HvHtT05, BQL07], especialmente en procesos de integración de conocimiento proveniente de fuentes diversas [BHP09, BKM91, AK05]. En particular; el problema de incoherencia, si bien no ha sido estudiado en profundidad en Datalog+/-, si es conocido en otros formalismos de representación de conocimiento, especialmente en la comunidad de Lógicas Descriptivas (Description Logics - DL), donde ha sido analizado desde diferentes enfoques a través de los años [FHP<sup>+</sup>06, BB97, KPSH05, QH07].

La obtención de conocimiento coherente y consistente basado en distintas fuentes es un problema que aparece en muchos campos de la IA, *e. g.*, sistemas multi-agentes, y fue atacado desde diferentes enfoques. En particular, mucho trabajo al respecto fue hecho en el campo de Bases de Datos, principalmente en integración de esquemas, donde el objetivo es integrar muchas bases de datos en una sola, resolviendo inconsistencias de forma tal que la calidad de la información sea comprometida en la menor medida posible.

Otro área que ha hecho muchos avances en la integración de fuentes inconsistentes de información es la Revisión de Creencias a través de la definición de procesos de Integración a través de Integración de Creencias (Belief Merging [BHA<sup>+</sup>01]). A través de los años, diferentes enfoques para la integración se han desarrollado, *e. g.*, [BKM91, Cho98, LS98, FKIRS12].

En un proceso de integración se toman varias Bases de Conocimiento (Knowledge Bases - *KBs*) que pueden ser consistentes en sí mismas pero generan conflicto cuando son consideradas juntas, y se obtiene una nueva *KB* que está libre de conflictos y refleja la información de las bases originales tanto como sea posible.

Uno de los trabajos mas influyentes en Belief Merging es el de Konieczny and Pino-Pérez [KP02]. En él, los autores proponen ciertas propiedades lógicas que los operadores de Merging deberían satisfacer, estableciendo a su vez formas de conseguir definir operadores que cumplan estas propiedades, a través de la definición de Teoremas de Representación. Además de [KP02], muchos otros enfoques han sido desarrollados para la integración y revisión de bases de datos proposicionales (*e. g.*, [KP02, KM92, LS98, KP11, BMVW10, DJ12]), los cuales brindaron las bases para desarrollos que atacaron estos problemas para (fragmentos de) lógicas de primer orden, principalmente en la familia de lenguajes de Lógicas Descriptivas (Description Logics - DLs) [QLB06, BHP09, MLB05] y Programación Lógica (Logic Programming) [HPW09, DSTW09].

En esta línea de investigación nos enfocamos en la definición de procesos de integración como los presentados que se adecuen a su uso en formalismos de representación de conocimientos aptos para el uso en entornos colaborativos como la Web Semántica. En los últimos

años, las bases de conocimiento en forma de ontologías se han vuelto muy populares en estos entornos, ya que proveen formas de representar tanto los datos disponibles en sí como las restricciones impuestos a éstos. Además, el poder expresivo de las ontologías permite realizar tareas importantes en la integración de fuentes de conocimientos [Len02], y juega un rol preponderante en la Web Semántica [BLHL01].

Por lo tanto, es importante definir métodos que permitan la integración de ontologías resolviendo todos los problemas de incoherencia/inconsistencia que puedan aparecer. Esto podría ser el primer paso para la integración de otras fuentes de conocimiento que pueden ser expresadas mediante estas ontologías (*e. g.*, bases de datos relacionales).

En esta línea de investigación nos enfocamos en la integración de ontologías desarrolladas en un lenguaje de ontologías en particular denominado Datalog+/- [CGL12]. La familia de lenguajes de ontologías Datalog permite un estilo modular de representación de conocimiento mediante el uso de reglas de forma similar a la usada en Programación Lógica, y su decidibilidad le permite manejar los volúmenes masivos de datos que podemos encontrar en aplicaciones hoy en día, haciéndola útil en diferentes campos como la consulta de ontologías, extracción de datos en web o intercambio de datos [LMS12]. En particular, la representación de conocimiento en ontologías Datalog+/- se lleva a cabo mediante el uso de (a) una Base de Datos: un conjunto de átomos que representan hechos acerca del mundo, *e. g.*, alumno(pedro) (b) Tuple-generating Dependencies - TGDs: reglas que nos permiten obtener nuevos átomos mediante la activación de las mismas como ser: alumno(*X*)  $\rightarrow$  persona(*X*), (c) Equality-generating Dependencies: reglas que restringen la generación de átomos, por ejemplo: doctor(*D*, *P*)  $\wedge$  doctor(*D'*, *P*)  $\rightarrow$  *D* = *D'*; y (d) Negative Constraints - NCs: reglas que expresan relaciones que no pueden existir entre átomos, *e. g.*: alto(*X*)  $\wedge$  bajo(*X*)  $\rightarrow \perp$ .

Para la definición de los procesos de integración proponemos el desarrollo de un framework general para la integración de ontologías Datalog+/- basado en el uso de Kernel Contraction y los operadores de revisión asociados [Han94, Han97, Han01] surgidos del campo de la Revisión de Creencias. Utilizando estos formalismos se podrá solucionar los problemas de incoherencias a inconsistencias que tienden a aparecer en la integración de información proveniente de distintas fuentes.

### 3. Líneas de Investigación y Desarrollo

Esta línea de investigación se enfoca en la definición de procesos de integración de ontologías Datalog+/- a través del uso de formalismos enfocados en la resolución de incoherencias e inconsistencias provenientes de las áreas de Revisión de Creencias y Argumentación. Para ello distintos ejes deben ser investigados, que van desde la definición de incoherencias e inconsistencia en



el entorno de ontologías Datalog+/- hasta las posibles aplicaciones que un método automático de integración de estas ontologías podría tener.

### 3.1. Definición de métodos de identificación de Incoherencias e Inconsistencias en Datalog+/-

Datalog+/- se ha vuelto un lenguaje muy popular en los últimos años, y numerosos estudios se han realizado acerca de sus propiedades de decibilidad y la complejidad asociada a la respuesta de consultas en estas ontologías. Sin embargo, no ha habido mucho estudio acerca de los aspectos de Representación de Conocimiento en Datalog+/. Muy poco trabajo se ha hecho acerca de inconsistencias en ontologías Datalog+/. Peor aún es la situación respecto del concepto de incoherencia (*i. e.*, la imposibilidad de satisfacer cierto conjunto de TGDs sin violar una NC), que ha sido soslayado al punto tal de no haber actualmente una definición formal del mismo.

Uno de los ejes de esta línea de investigación es la definición formal del concepto de incoherencia en Datalog+/-, tomando como partida esfuerzos similares que han sido realizados para otros formalismos de representación de conocimiento, principalmente Description Logics. Adicionalmente, se procederá a identificar las propiedades que llevan a que un conjunto de TGDs sea incoherente, y las que hacen que una ontología Datalog+/- se vuelva inconsistente. De esta manera se podrán identificar tales casos, lo que será el primer paso para la posterior resolución de tales problemas.

### 3.2. Resolución de Incoherencias e Inconsistencias en Datalog+/-

Una vez que se tiene definidos e identificados los conjuntos incoherentes de TGDs y aquellos conjuntos de átomos que provocan inconsistencias en la unión de varias ontologías Datalog+/-, se debe proceder a la resolución de estos conflictos. En esta línea de investigación esto se hará mediante la aplicación de técnicas derivadas de la Revisión de Creencias denominadas Kernel Contraction.

Este tipo de técnicas resuelve conflictos de incoherencia/inconsistencia tomando los conjuntos conflictivos mínimos y eligiendo de alguna forma que elemento remover de los mismos para solucionar el problema. En el caso de integración de ontologías Datalog+/- esto es la remoción de ciertos átomos y ciertas TGDs de la unión de todas las ontologías que se está integrando. Adicionalmente, se puede pensar en la definición de métodos de debilitamiento de reglas, en lugar de la remoción de las mismas. Esto no es una tarea trivial, ya que hay muchos aspectos a definir, por ejemplo como elegir el mejor candidato entre los átomos o TGDs que pueden eliminarse, lo que a su vez lleva a definir formas (automáticas) de obtener órdenes entre los candidatos. Para esto se procederá a definir operadores de integración de ontologías Datalog+/-, así como se darán las propiedades esperadas de tales operadores y se definirán métodos de obtención

de operadores de tales características mediante Teoremas de Representación.

### 3.3. Posibles extensiones a Datalog+/- mediante Argumentación Rebatible

Otro aspecto a considerar dentro de la línea de investigación es como se beneficiaría Datalog+/- de otros formalismos de representación de conocimientos con propiedades diferentes a aquellas presentes en Datalog+/. Particularmente, proponemos analizar posibles extensiones a las ontologías Datalog+/- basadas en el uso de formalismos de Argumentación Rebatible como Programación Lógica Rebatible (Defeasible Logic Programming - DeLP) [GS04].

Para tales extensiones se deberán analizar diferentes aspectos. Por ejemplo, la relación de inferencia en estas ontologías aumentadas deberá tener en cuenta los aspectos no-monótonos de la Argumentación Rebatible, llevando a que se modifique la forma en que una consulta es respondida respecto del proceso actual en Datalog+/-, ya que se deberá tener en cuenta el análisis dialéctico llevado a cabo por DeLP antes de responder la misma.

Estas modificaciones en la forma en que información es inferida traerá aparejado un impacto en las inferencias finales de las ontologías en aquellos casos en donde las ontologías no son coherentes o consistentes, proveyendo otra forma de integrar las mismas (*i. e.*, simplemente considerarlas juntas sin importar problemas de incoherencia e inconsistencia, y dejar que el proceso argumentativo los resuelva).

Otro aspecto importante a analizar será el impacto de estos cambios en la decibilidad y (principalmente) la complejidad del proceso de resolución de consultas en (la extensión de) Datalog+/-.

### 3.4. Integración de otras fuentes de datos a través del Merging de Ontologías Datalog+/-

Finalmente, se analizará nuestro framework de integración de ontologías Datalog+/- como medio de integración de otras fuentes de datos. Principalmente nos enfocaremos en la integración automática de Bases de Datos Relacionales.

Para esto, primeramente definiremos métodos para expresar bases de datos relacionales a través de ontologías Datalog+/-, tanto los datos en sí como aspectos relacionados al esquema de las mismas, *e. g.*, las dependencias funcionales.

Una vez logrado esto se podría utilizar los métodos de integración de ontologías Datalog+/- para obtener una federación de las bases de datos expresadas, ya que la ontología final resultante de la integración brindaría un esquema integrador de las mismas así como los datos que serían parte de la federación, manteniendo a su vez la coherencia de las restricciones de integridad respecto del esquema unificado y la consistencia de los datos almacenados respecto de las dependencias funcionales.

## 4. Resultados y Objetivos

El objetivo general de este trabajo de investigación es el diseño y construcción de la infraestructura necesaria para la realización de métodos de integración de ontologías Datalog+/-, así como la identificación de posibilidades de transferencia del framework a la integración de otras fuentes de información como las bases de datos relacionales. Contar con estos métodos permitirá el uso de ontologías (u otras fuentes) de manera segura en el desarrollo de sistemas colaborativos. Por ejemplo, se podrían definir nuevas arquitecturas para Sistemas de Soporte de Decisiones (DSS) que acceda a información proveniente de diferentes entidades de manera transparente sin tener problemas de inconsistencias o incoherencias. Una extensión análoga podría definirse para los Sistemas de Recomendación (RS), en especial aquellos que basan su funcionamiento sobre una base de conocimiento y que utilizan tecnologías de Web Semántica, es decir, Sistemas de Recomendación Semánticos.

Respecto de los resultados y objetivos alcanzados, se ha avanzado en la definición e identificación de los casos de incoherencia e inconsistencia en Datalog+/- . A su vez, se ha comenzado la definición de operadores de integración de ontologías Datalog+/- basados en Kernel Contraction, estableciendo las propiedades de los mismos y como obtenerlos. Estos resultados se presentarán en congresos nacionales e internacionales.

Adicionalmente, los operadores definidos están siendo implementados actualmente, para así poder probar su eficacia mediante pruebas en la integración de ontologías ya desarrolladas adaptadas al estilo de representación de conocimiento de Datalog+/- . También se está avanzando en la definición de procesos alternativos de integración de ontologías que utilizan formalismos no-monótonos como la Argumentación Rebatible para la resolución de conflictos, habiendo establecido ya como los mismos podrían usarse sobre un conjunto de ontologías Datalog+/- . En lo subsecuente se procederá a analizar las propiedades de un framework de estas características, así como las posibilidades computacionales del mismo. Dentro de las propiedades a analizar del mismo se plantea como objetivo futuro realizar un estudio comparativo de nuestra propuesta respecto de un formalismo proveniente del área de Bases de Datos cuyo uso para la resolución de inconsistencias ha sido muy extendido: Consistent Query Answering (CQA) [ABC99]. En particular, se analizará la completitud y la sensatez de nuestra propuesta respecto de CQA, así como también las posibles extensiones en las inferencias que nuestro framework podría tener respecto del mismo.

## 5. Formación de Recursos Humanos

En la presente línea de investigación se enmarca el desarrollo de una tesis de posgrado en el Doctorado en Ciencias de la Computación del Departamento de Ciencias e Ingenierías de la Computación de la Universidad

Nacional del Sur.

## Referencias

- [ABC99] Marcelo Arenas, Leopoldo E. Bertossi, and Jan Chomicki. Consistent query answers in inconsistent databases. In Victor Vianu and Christos H. Papadimitriou, editors, *PODS*, pages 68–79. ACM Press, 1999.
- [AK05] Leila Amgoud and Souhila Kaci. An argumentation framework for merging conflicting knowledge bases: The prioritized case. In *Proc. of ECSQARU 2005*, pages 527–538, 2005.
- [BB97] Domenico Beneventano and Sonia Bergamaschi. Incoherence and subsumption for recursive views and queries in object-oriented data models. *Data Knowl. Eng.*, 21(3):217–252, 1997.
- [BHA<sup>+</sup>01] Isabelle Bloch, Anthony Hunter, Alain Appriou, André Ayoun, Salem Benferhat, Philippe Besnard, Laurence Cholvy, Roger M. Cooke, Frédéric Cuppens, Didier Dubois, Hélène Fargier, Michel Grabisch, Rudolf Kruse, Jérôme Lang, Serafín Moral, Henri Prade, Alessandro Saffiotti, Philippe Smets, and Claudio Sossai. Fusion: General concepts and characteristics. *Int. J. Intell. Syst.*, 16(10):1107–1134, 2001.
- [BHP09] Elizabeth Black, Anthony Hunter, and Jeff Z. Pan. An argument-based approach to using multiple ontologies. In *SUM*, pages 68–79, 2009.
- [BKM91] Chitta Baral, Sarit Kraus, and Jack Minker. Combining multiple knowledge bases. *IEEE Trans. Knowl. Data Eng.*, 3(2):208–220, 1991.
- [BLHL01] Tim Berners-Lee, James Hendler, and Ora Lassila. The semantic web. *Scientific American*, 284(5):3443, 2001.
- [BMVW10] Richard Booth, Thomas Andreas Meyer, Ivan José Varzinczak, and Renata Wassermann. Horn belief change: A contraction core. In *ECAI*, volume 215 of *Frontiers in Artificial Intelligence and Applications*, pages 1065–1066, 2010.
- [BQL07] David A. Bell, Guilin Qi, and Weiru Liu. Approaches to inconsistency handling in description-logic based ontologies. In *OTM Workshops (2)*, pages 1303–1311, 2007.
- [CGL12] Andrea Cali, Georg Gottlob, and Thomas Lukasiewicz. A general datalog-based framework for tractable query answering over ontologies. *J. Web Sem.*, 14:57–83, 2012.

- [Cho98] Laurence Cholvy. Reasoning about merged information. In Didier Dubois, Henri Prade, Dov M. Gabbay, and Philippe Smets, editors, *Belief Change*, volume 3 of *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, pages 233–263. Springer Netherlands, 1998.
- [DJ12] James P. Delgrande and Yi Jin. Parallel belief revision: Revising by sets of formulas. *Artif. Intell.*, 176(1):2223–2245, 2012.
- [DSTW09] James P. Delgrande, Torsten Schaub, Hans Tompits, and Stefan Woltran. Merging logic programs under answer set semantics. In Patricia M. Hill and David Scott Warren, editors, *ICLP*, volume 5649 of *Lecture Notes in Computer Science*, pages 160–174. Springer, 2009.
- [FHP<sup>+</sup>06] Giorgos Flouris, Zhisheng Huang, Jeff Z. Pan, Dimitris Plexousakis, and Holger Wache. Inconsistencies, negations and changes in ontologies. In *AAAI*, pages 1295–1300. AAAI Press, 2006.
- [FKIRS12] Marcelo Alejandro Falappa, Gabriele Kern-Isberner, Maurício Reis, and Guillermo Ricardo Simari. Prioritized and non-prioritized multiple change on belief bases. *Journal of Philosophical Logic*, 41(1):77–113, 2012.
- [GCS10] Sergio Alejandro Gómez, Carlos Iván Chesñevar, and Guillermo Ricardo Simari. Reasoning with inconsistent ontologies through argumentation. *Applied Artificial Intelligence*, 24(1&2):102–148, 2010.
- [GS04] Alejandro Javier García and Guillermo Ricardo Simari. Defeasible logic programming: An argumentative approach. *TPLP*, 4(1-2):95–138, 2004.
- [Han94] Sven Ove Hansson. Kernel contraction. *J. Symb. Log.*, 59(3):845–859, 1994.
- [Han97] Sven Ove Hansson. Semi-revision (invited paper). *Journal of Applied Non-Classical Logics*, 7(2), 1997.
- [Han01] Sven Ove Hansson. *A Textbook of Belief Dynamics: Solutions to Exercises*. Kluwer Academic Publishers, Norwell, MA, USA, 2001.
- [HPW09] Julien Hué, Odile Papini, and Eric Würbel. Merging belief bases represented by logic programs. In Claudio Sossai and Gaetano Chemello, editors, *ECSQARU*, volume 5590 of *Lecture Notes in Computer Science*, pages 371–382. Springer, 2009.
- [HvHH<sup>+</sup>05] Peter Haase, Frank van Harmelen, Zhisheng Huang, Heiner Stuckenschmidt, and York Sure. A framework for handling inconsistency in changing ontologies. In *Proc. of ISWC 2005*, pages 353–367, 2005.
- [HvHtT05] Zhisheng Huang, Frank van Harmelen, and Annette ten Teije. Reasoning with inconsistent ontologies. In *Proc. of IJCAI 2005*, pages 454–459, 2005.
- [KM92] Hirofumi Katsuno and Alberto O. Mendelzon. Propositional knowledge base revision and minimal change. *Artif. Intell.*, 52(3):263–294, 1992.
- [KP02] Sébastien Konieczny and Ramón Pino Pérez. Merging information under constraints: A logical framework. *J. Log. Comput.*, 12(5):773–808, 2002.
- [KP11] Sébastien Konieczny and Ramón Pino Pérez. Logic based merging. *J. Philosophical Logic*, 40(2):239–270, 2011.
- [KPSH05] Aditya Kalyanpur, Bijan Parsia, Evren Sirin, and James A. Hendler. Debugging unsatisfiable classes in owl ontologies. *J. Web Sem.*, 3(4):268–293, 2005.
- [Len02] Maurizio Lenzerini. Data integration: A theoretical perspective. In *PODS*, pages 233–246, 2002.
- [LMS12] Thomas Lukasiewicz, Maria Vanina Martinez, and Gerardo I. Simari. Inconsistency handling in datalog+/- ontologies. In *Proc. of ECAI*, pages 558–563, 2012.
- [LS98] Paolo Liberatore and Marco Schaerf. Arbitration (or how to merge knowledge bases). *IEEE Trans. Knowl. Data Eng.*, 10(1):76–90, 1998.
- [MLB05] Thomas Meyer, Kevin Lee, and Richard Booth. Knowledge integration for description logics. In *In Proceedings of the 7th International Symposium on Logical Formalizations of Commonsense Reasoning*, pages 645–650. AAAI Press, 2005.
- [QH07] Guilin Qi and Anthony Hunter. Measuring incoherence in description logic-based ontologies. In *ISWC/ASWC*, pages 381–394, 2007.
- [QLB06] Guilin Qi, Weiru Liu, and David A. Bell. Knowledge base revision in description logics. In Michael Fisher, Wiebe van der Hoek, Boris Konev, and Alexei Lisitsa, editors, *JELIA*, volume 4160 of *Lecture Notes in Computer Science*, pages 386–398. Springer, 2006.

## Modelo de Servicio de Razonamiento con Preferencias

Juan Carlos Teze<sup>1</sup>, Sebastian Gottifredi<sup>2</sup>,  
Alejandro Javier García<sup>3</sup> y Guillermo Ricardo Simari<sup>4</sup>

Laboratorio de Investigación y Desarrollo en Inteligencia Artificial

<sup>2,3,4</sup> Departamento de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur,  
Av. Alem 1253, (B8000CPB) Bahía Blanca, Argentina

Laboratorio de Investigación y Desarrollo en Inteligencia Artificial Concordia

<sup>1</sup>Facultad de Ciencias de la Administración, Universidad Nacional de Entre Rios,

<sup>1,2,3</sup> Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)

e-mail: {jct,sg,ajg,grs}@cs.uns.edu.ar

### Resumen

En la última década formalizar modelos para que agentes deliberativos puedan razonar en un sistema multi-agente se ha convertido en un importante área en AI. Muchos enfoques se han propuesto, uno de estos es el de Servicio de Razonamiento [15]; esta línea de investigación básicamente se focalizará en dicho servicio. Actualmente este servicio cuenta con un conjunto de operadores que modifican temporalmente la información almacenada en el servidor y un mecanismo de inferencia que es el que resuelve una consulta dada. No obstante, una instanciación de este servicio es el Servicio de Razonamiento basado en argumentación rebatible. En particular, en este servicio el mecanismo de inferencia está configurado para que utilice siempre los mismos criterios de preferencia al momento de resolver consultas cliente. El objetivo general de esta línea de investigación es intentar modelar un Servicio de Razonamiento con preferencias donde las mismas puedan manipularse de forma dinámica. Como resultado se buscará definir algunos operadores para lograr tales objetivos.

**Palabras clave:** Sistema Multi-agente, Agente Inteligente Deliberativo, Servicio de Razo-

namiento, Preferencias.

### Contexto

Esta línea de investigación se llevará a cabo dentro del ámbito de colaboración entre el Laboratorio de Investigación y Desarrollo (LIDIA) del Dep. de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur; y el Laboratorio de Investigación y Desarrollo en Inteligencia Artificial Concordia (LIDIA Concordia) de la Facultad de Ciencias de la Administración, Universidad Nacional de Entre Ríos. Está asociada a los siguientes proyectos de investigación:

- Desarrollo de Sistemas de Argumentación Masiva sobre Base de Datos Federadas. PID-UNER 7041.
- Representación de conocimiento y Razonamiento Argumentativo: Herramientas Inteligentes para la Web y las Bases de Datos Federadas. PGI 24/N030, financiados por la Universidad Nacional del Sur.



Además, este trabajo se realiza en el marco del desarrollo de una tesis doctoral para optar por el título de Doctor en Ciencias de la Computación del Departamento de Ciencias de la Computación de la Universidad Nacional del Sur.

## Introducción

Una característica esencial de los Sistemas Multi-Agente (SMAs) es la interacción entre agentes. Los agentes pertenecientes a un SMA interactúan con el objetivo de llevar a cabo tareas, ya sea individuales o colectivas. Los agentes deliberativos generalmente razonan utilizando dos tipos de conocimiento: uno público que es el que comparten con otros agentes y otro privado que surge en parte de su percepción del mundo.

En [15] se propone un modelo cliente-servidor que permite a los agentes compartir conocimiento con otros agentes y representar el conocimiento privado de cada uno. En este modelo se define el concepto de Servicio de Razonamiento. Este servicio permite representar información o conocimiento público y responder consultas de clientes utilizando dicho conocimiento público junto con información privada provista por el cliente. Básicamente un Servicio de Razonamiento está constituido por:

1. el conocimiento público que será utilizado como base para la resolución de las consultas de los clientes;
2. un mecanismo de inferencia que utilizará ese conocimiento para computar las respuestas a consultas de los clientes; y
3. un conjunto de operadores de tratamiento de contexto mediante los cuales se podrá, por ejemplo, agregar en forma temporal información para crear el contexto en el cual se procesará la consulta del cliente.

Aquellos contextos en donde la información que se maneja es incompleta o potencialmente contradictoria brindan un escenario ideal para sistemas argumentativos[10, 1, 3, 4] capaces de representar el conocimiento y definir el razonamiento de agentes inteligentes. El mecanismo de inferencia sobre el cual están basado, permite decidir entre conclusiones contradictorias y adaptarse fácilmente a entornos cambiantes.

En argumentación dos argumentos pueden estar en conflicto. Para determinar que argumento prevalece en una situación de conflicto se utilizan criterios de preferencia. El argumento que prevalece se dice que es un derrotador del otro argumento en conflicto. En la literatura existen numerosos criterios de preferencia [8, 2, 12] para comparar argumentos. En particular, en los Servicios de Razonamiento basado en argumentación rebatible, los criterios de preferencia se encontrarían embebido en el Mecanismo de Inferencia.

El modelo actual de Servicio de Razonamiento, puede ser utilizado en aplicaciones particulares en dominios de aplicación concretos con diferentes lenguajes de representación de conocimiento. Concretamente en [9], se propone un servidor basado en programación en lógica rebatible. Es importante mencionar que estos servicios están estrechamente relacionados con sistemas recomendadores [6, 7, 14, 11, 16]; esto se debe a que los sistemas recomendadores usualmente operan creando un modelo de las preferencias de los usuarios con el objetivo de poder anticiparse a las necesidades y deseos del usuario. Un Servicio de Razonamiento puede ser implementado como un servidor de recomendación.

En base a lo expuesto arriba, un paso importante para avanzar es el de formalizar un modelo de Servicio de Razonamiento basado en argumentación rebatible [5, 13] en donde

se puedan manejar diversas preferencias por parte del servidor y el tratamiento que puedan recibir sea dinámico. En la actualidad el Servicio de Razonamiento no cuenta con mecanismos para manejar preferencias. El uso de preferencias y operadores concretos aplicados a estas es un avance fundamental en estos tipos de servicios.

A continuación, se explica cómo esta línea de investigación integra la noción de preferencia a un Servicio de Razonamiento. Se repasarán los resultados obtenidos y el trabajo en progreso.

## Líneas de investigación y desarrollo

La línea principal de investigación de este artículo toma como base el concepto de Servicio de Razonamiento y busca avanzar sobre un tipo de servicio en particular; Servicio de Razonamiento basado en argumentación rebatible. Un primer avance va ser intentar extender un Servicio de Razonamiento para que pueda soportar varios criterios de preferencia sobre la información a computar. Además, se espera poder construir consultas específicas dirigidas a estos servicios en concreto.

Por lo tanto, nuestra línea de investigación está motivada por tres ejes: el primero se basa en adaptar los servicios mencionados en la introducción para que los criterios de preferencias sean un modulo independiente del Mecanismo de Inferencia; el segundo en definir consultas de clientes dirigidas a estos servicios en particular; y el tercero en definir nuevos operadores que permitan la manipulación de dichas preferencias.

### Servicio de razonamiento con preferencias

Un Servicio de Razonamiento responde una consulta recibida de un clien-

te. Para responder consultas, estos tienen almacenado conocimiento público el cual es utilizado junto al conocimiento privado del agente para crear un contexto adecuado a la consulta. Un Servicio de Razonamiento basado en argumentación rebatible podría contar de cuatro componentes: el conocimiento público, un mecanismo de inferencia, un conjunto de operadores de tratamiento de contexto y un conjunto de criterios de preferencia.

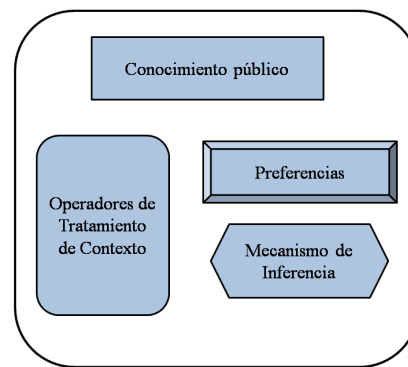


Figura 1: Servicio de razonamiento basado en argumentación rebatible.

Los Servicios de Razonamiento responden a un modelo cliente-servidor por consiguiente buscan resolver determinadas consultas de clientes. En nuestra línea de investigación las consultas siguen el formato de las consultas contextuales definidas en [15, 16], las cuales se detallarán en la siguiente sección.

### Consultas extendidas

La Consulta Contextual, permite a los clientes incluir en sus consultas conocimiento privado y especificar qué tratamiento debe darse a dicho conocimiento, al momento de integrarse con el conocimiento público almacenado en el Servicio de Razonamiento. Esta información es lo que se denomina contexto de la consulta y,

en general, representa conocimiento propio de los clientes.

Tener el conjunto de criterios de preferencias separadas del mecanismo de inferencia brinda muchas ventajas en un modelo de razonamiento cliente-servidor. A continuación mencionaremos algunas de las líneas de investigación a seguir en trabajos futuros.

En primer lugar como están definidas actualmente las consultas de los clientes, no permiten que la consulta contenga los criterios de preferencias que el cliente desea que el servidor utilice para computar su consulta. De esta manera, intuitivamente lo que se busca es configurar el servidor para que razona utilizando las preferencias del cliente.

Otra ventaja de poder incluir sus preferencias en las consultas es poder indicarle al servidor, de todas sus preferencias cuales utilizar al momento resolver la consulta. En definitiva, lo que se le intenta brindar al cliente es mayor capacidad de interacción con el servidor. Para llevar a cabo esta interacción es imprescindible definir operadores específicos para manipular preferencias.

### Operadores para manipular preferencias

Los operadores propuestos en [15], son utilizados para modificar temporalmente el conocimiento almacenado e indicar que tratamiento darle al conocimiento privado del cliente. Al agregar las preferencias como un componente más al Servicio de Razonamiento lo que proponemos es incluir operadores que puedan manipular dichas preferencias. A modo de ejemplo algunos de los operadores que proponemos se detallan a continuación:

- $\otimes$  Operador para reemplazar preferencias del servidor.
- $\succ$  Operador para dar prioridad a ciertas preferencias.

- $\oplus$  Operador para agregar preferencias al servidor.

Podemos definir infinidad de operadores, no obstante, estos dependen de la configuración en particular que el cliente le da al servidor al momento de la consulta.

## Resultados y Objetivos

Esta línea de investigación tiene como objetivo general formalizar un modelo de Servicio de Razonamiento basado en preferencias, de manera tal que dichas preferencias puedan manipularse de forma dinámica a través de operadores específicos. Para lograr los objetivos propuestos se tomarán como base los resultados obtenidos en los artículos [15, 9, 16].

## Formación de Recursos Humanos

Los temas de esta línea de investigación están estrechamente relacionados con el desarrollo de la tesis doctoral del 1º autor del presente artículo.

## Referencias

- [1] Alsinet, T., Chesñevar, C.I., Godo, L., Simari, G.R.: A logic programming framework for possibilistic argumentation: Formalization and logical properties. *Fuzzy Sets and Systems* 159(10), 1208–1228 (2008)
- [2] Antoniou, G., Maher, M.J., Billington, D.: Defeasible logic versus logic programming without negation as failure. *J. Log. Program.* 42(1), 47–57 (2000)
- [3] Capobianco, M., Chesñevar, C.I., Simari, G.R.: Argumentation and the dyna-

- mics of warranted beliefs in changing environments. *Autonomous Agents and Multi-Agent Systems* 11(2), 127–151 (2005)
- [4] Capobianco, M., Simari, G.R.: A proposal for making argumentation computationally capable of handling large repositories of uncertain data. In: *SUM*. pp. 95–110 (2009)
- [5] Chesñevar, C.I., Maguitman, A.G., Loui, R.P.: Logical models of argument. *ACM Computing Surveys* 32(4), 337–383 (2000)
- [6] Chesñevar, C.I., Maguitman, A.G., Simari, G.R.: Recommender system technologies based on argumentation 1. In: *Emerging Artificial Intelligence Applications in Computer Engineering*, pp. 50–73 (2007)
- [7] Chesñevar, C.I., Maguitman, A.G., Simari, G.R.: Argument-based user support systems using defeasible logic programming. In: *AIAI*. pp. 61–69 (2006)
- [8] Garcia, A., Simari, G.: Defeasible logic programming: An argumentative approach. *Theory and Practice of Logic Programming (TPLP)* 4, 95–138 (2004)
- [9] García, A.J., Rotstein, N.D., Tucacat, M., Simari, G.R.: An argumentative reasoning service for deliberative agents. In: *KSEM*. pp. 128–139 (2007)
- [10] García, A.J., Simari, G.R.: Defeasible logic programming: An argumentative approach. *Theory and Practice of Logic Programming* 4(1-2), 95–138 (2004)
- [11] Konstan, J.A.: Introduction to recommender systems: Algorithms and evaluation. *ACM Trans. Inf. Syst.* 22(1), 1–4 (2004)
- [12] Loui, R.P.: Defeat among arguments: a system of defeasible inference. *Computational Intelligence* 3, 100–106 (1987)
- [13] Prakken, H., Vreeswijk, G.: Logics for defeasible argumentation. In: Gabbay, D., Guenther, F. (eds.) *Handbook of Philosophical Logic*, vol. 4, pp. 218–319. Kluwer Academic Pub. (2002)
- [14] Resnick, P., Varian, H.R.: Recommender systems - introduction to the special section. *Commun. ACM* 40(3), 56–58 (1997)
- [15] Tucacat, M.: Grupos de Servicios de Razonamiento para el Procesamiento de Consultas Contextuales en Paralelo. Ph.D. thesis, Universidad Nacional del Sur, Bahía Blanca, Argentina (2000)
- [16] Tucacat, M., García, A.J., Simari, G.R.: Using defeasible logic programming with contextual queries for developing recommender servers. In: *AAAI Fall Symposium: The Uses of Computational Argument* (2009)



# Evaluación de la Calidad de la Información de Wikipedia en Español

Lian Pohn, Edgardo Ferretti, Marcelo Errecalde

Laboratorio de Investigación y Desarrollo en Inteligencia Computacional  
Departamento de Informática, Universidad Nacional de San Luis  
Ejército de los Andes 950 - (D5700HHW) San Luis - Argentina  
e-mails:lian\_pohn@hotmail.com,{ferretti,merreca}@unsl.edu.ar

## Resumen

Este artículo describe, brevemente, las tareas de investigación y desarrollo que se están llevando a cabo para evaluar la calidad de información en la Web en el marco del proyecto “Herramientos y mecanismos para la toma de decisiones en agentes inteligentes artificiales”. En particular, se ha tomado como caso de estudio primario la enciclopedia online Wikipedia en español. Este tema de trabajo permite la interacción de las dos líneas de investigación que contiene este proyecto y además se está abordando en forma conjunta con investigadores de Alemania, España y México, en el contexto de un proyecto FP7 financiado por la Unión Europea.

**Palabras clave:** Calidad de Información en la Web, Wikipedia, Sistemas Inteligentes.

## Contexto

El tema motivo de esta presentación se encuentra dentro de los alcances de la línea de investigación “Minería de Textos y de la Web” del proyecto “Herramientos y mecanismos para la toma de decisiones en agentes inteligentes artificiales”, Proyecto de Investigación consolidado de la Universidad Nacional de San Luis. En este proyecto, el objetivo principal es avanzar en la integración de

las investigaciones sobre herramientas para la extracción y análisis inteligente de contenido Web de calidad y desarrollo e integración de modelos y mecanismos efectivos para la toma de decisiones, y el aprendizaje automático.

Este proyecto posee además otra línea de investigación: “Agentes Inteligentes”; y la articulación entre ambas líneas está dada por la meta de implementar un sistema inteligente automático que evalúe la calidad de la información en Wikipedia y que siga el paradigma de agentes.

Las dos fuentes de financiamiento de este proyecto son la Universidad Nacional de San Luis y la Comisión Europea de Investigación e Innovación, a través del programa Marie Curie Actions: FP7 People 2010 IRSES.

## Introducción

En la actualidad, el acceso a la información es un tema clave en todos los aspectos relacionados con la vida moderna. En este sentido, la evaluación de la calidad de la información en la Web se ha convertido en una tarea crucial, dado que cada día son más las personas y entidades gubernamentales o privadas que toman decisiones basándose en información disponible en la Web. Asimismo, el notable incremento de información disponible en la Web ha potenciado la necesidad de eva-

luar su calidad de forma automática. Este hecho se debe entre otras cosas, a la creciente popularidad de sitios que permiten a usuarios comunes generar de forma muy sencilla contenido web y la inevitable divergencia en la calidad del contenido producido [8].

En este contexto, Wikipedia es un emprendimiento paradigmático. Esta enciclopedia de libre acceso generada a partir de las contribuciones de millones de usuarios, tiene esta característica como principal fortaleza en lo que respecta a su creciente popularidad, siendo (la versión inglesa) uno de los diez sitios más visitados en el mundo, en la actualidad. Sin embargo, esta característica es probablemente, el mayor desafío que Wikipedia enfrenta en cómo mejorar de forma sistemática la calidad informativa de sus archivos. Este aspecto no es casual si consideramos que los autores que contribuyen con Wikipedia son heterogéneos, en cuanto al nivel de educación, edad, cultura, habilidades del lenguaje y especialización en un área.

Una interpretación ampliamente aceptada de calidad de información, es que en sí mismo, es un concepto multi-dimensional que se define por ciertos aspectos de calidad (dimensiones); como por ejemplo: la exactitud, fiabilidad y relevancia [19]. Asimismo, la evaluación de la calidad de información requiere la consideración del contexto y casos de uso [23]. En particular, en el contexto de Wikipedia, es decir, el género de las enciclopedias, el ideal en lo que respecta a calidad de información ha sido formalizado bajo el concepto de artículo destacado (AD) (en inglés, *featured article*). Los ADs son artículos de Wikipedia que, como resultado de un arduo proceso de revisión entre pares, son considerados como artículos de alta calidad de acuerdo a criterios claramente definidos en la comunidad de Wikipedia. Estos criterios incluyen aspectos tales como la calidad de la escritura, la comprensibilidad, la existencia de un buen trabajo de investigación que lo sustente, neutralidad, estabilidad, entre otros.<sup>1</sup>

<sup>1</sup>[http://en.wikipedia.org/wiki/Wikipedia:Featured\\_article\\_criteria](http://en.wikipedia.org/wiki/Wikipedia:Featured_article_criteria).

A pesar de lo importante que es contar con una definición precisa sobre qué constituye un AD, como se indica en [3], en la actualidad menos de 0,1 % de los artículos en *Wikipedia.org* son AD. En el caso de Wikipedia en español, se estima que este porcentaje es de 0,11 %.<sup>2</sup>

En la siguiente sección, se describen los principales enfoques desarrollados por la comunidad científica en lo que respecta al estudio de calidad de información en Wikipedia.

## Líneas de Investigación y Desarrollo

Los métodos utilizados para determinar aspectos de calidad en los artículos de Wikipedia son muy variados; no obstante, en términos generales podrían identificarse tres líneas de investigación principales, que se describen a continuación.

### Identificación de Artículos Destacados

Dado que los artículos de Wikipedia se escriben de forma colaborativa y se mantienen principalmente por voluntarios, la lógica detrás de la idea: “mientras mayor sea el número de ediciones de un artículo y mayor sea el número de editores, mayor debería ser su calidad”, es muy razonable. De hecho, Wilkinson y Huberman [24] proveen evidencia que los artículos destacados pueden ser distinguidos de aquellos que no lo son considerando el número de ediciones que han tenido y de editores distintos que han contribuido.

De la misma manera, continuando con la intuición anterior, mientras más ediciones tenga un artículo tentativamente éstos tenderán a tener más contenido textual. Este hallazgo también ha sido explorado por Blumenstock [9] que demostró experimentalmente que una característica tan sencilla como el número de palabras en el documento puede competir con características sofisticadas para

<sup>2</sup>[http://es.wikipedia.org/wiki/Wikipedia:Articulos\\_destacados](http://es.wikipedia.org/wiki/Wikipedia:Articulos_destacados)

distinguir artículos destacados de aquellos que no lo son.

El hecho de que los artículos destacados tienden a ser más largos hace que probablemente éstos tengan más contenido factual. Basándose en esta suposición, Lex et al. [15] propusieron una medida llamada *densidad factual* para discriminar entre artículos destacados y los que no lo son. Básicamente, esta medida relaciona el número total de hechos en un artículo con respecto a su longitud. Esta medida probó ser muy útil para discriminar artículos destacados cuya longitud era en promedio similar a la de los artículos comunes. Este estudio demostró que los artículos destacados tienden a ser más informativos que los que no lo son, independientemente de su longitud.

De acuerdo con los resultados presentados en [7, 16], este fenómeno se debe a que la producción conjunta y cooperación entre miembros de una comunidad Wiki logra que surja contenido de alta calidad.

### **Desarrollo de métricas y características para la evaluación de la calidad**

Si bien los trabajos [7, 16], no tienen como objetivo particular la identificación de artículos destacados en Wikipedia, las conclusiones obtenidas de los mismos sirven de soporte para explicar los resultados obtenidos por Lex et al. [15], como se comentó anteriormente.

Un trabajo análogo a [16] es el llevado a cabo por Hu et al. [14], donde se desarrollan modelos para medir la calidad de los artículos editados en Wikipedia basados en el principio: “buenos autores escriben buenos artículos y los buenos artículos están escritos por buenos autores”, es decir, que existe una dependencia mutua entre calidad y autoría. Asimismo, en este trabajo también se reporta que a veces evaluar solamente la interacción entre editores ([7, 16]) no es suficiente para determinar la calidad de un artículo y que la longitud en los mismos tiende a aportar calidad informativa, sin que eso conlleve a que el mero hecho de que un artículo sea extenso implicará que sea de buena calidad.

En [20], Stvilia et al. proponen siete métricas computacionales específicas de calidad de información en Wikipedia. Probaron la validez de las mismas en la discriminación de artículos destacados de los que no lo son, con una exactitud cercana al 90%. También, toman la iniciativa en la construcción de métricas de calidad de información que siendo originalmente pensadas para Wikipedia, se pueden adaptar para su aplicación en contextos similares de creación de contenido por parte de los usuarios.

### **Detección de Defectos de Calidad**

A diferencia de las líneas de investigación descritas precedentemente, la detección de defectos de calidad (en inglés: *quality flaws*) intenta identificar imperfecciones o fallas de calidad específicas en los artículos de Wikipedia, convirtiéndose en los últimos tiempos en una de las tareas seleccionadas para su evaluación en la competencia PAN del CLEF 2012 Evaluation Labs and Workshop [2]. Si bien se pueden mencionar algunos trabajos que han seguido esta línea de investigación considerando pequeñas muestras de artículos [21] o analizaron solamente un conjunto restringido de deficiencias de calidad [5, 13], los primeros en realizar un análisis detallado de los defectos de calidad en Wikipedia fueron investigadores de la Universidad de Weimar encabezados por Anderka [1, 3, 6]. Este análisis revela que un 27,52% de los artículos de Wikipedia en inglés contiene al menos un defecto de calidad y que el 70% de las deficiencias se relacionan con la verificabilidad del artículo. Este análisis está basado en artículos etiquetados manualmente por lo que se supone que el número real de fallas es más alto. Por lo tanto, es altamente probable que muchos artículos con deficiencias no hayan sido identificados aún.

Respecto a la competencia llevada a cabo en el contexto del PAN [2], los organizadores plantearon la predicción de defectos de calidad en Wikipedia como un problema de clasificación de una clase [22] (*one-class classification problem*), como ya fuera propuesto

en [4, 6]. En este contexto, el problema es: dado un conjunto de artículos de Wikipedia etiquetados con un defecto de calidad particular, decidir si un artículo (no etiquetado) sufre de este defecto. En este problema, y como se resalta en [2], el desafío clave es la ausencia de datos de entrenamiento “negativos”, es decir, artículos etiquetados como “no conteniendo” una falla particular. Esto hace que ciertas técnicas clásicas de clasificación basadas en discriminación (binaria o multiclase) sean inaplicables. Por lo tanto, la ingeniería de características (*feature engineering*), es decir, el desarrollo de modelos de documentos que discriminan artículos conteniendo una cierta falla desde todos los otros artículos, se convierte en un factor crucial.

De hecho, el equipo ganador [11] de esta competencia implementó un sub-conjunto de las features propuestas en [6] para la representación de los artículos de Wikipedia y enfatizó el estudio experimental con distintas variantes algorítmicas de PU Learning [18] relacionadas principalmente con la evaluación de distintas estrategias de muestreo de documentos no etiquetados y diferentes enfoques para la selección de negativos confiables.

## Resultados y Objetivos

De acuerdo con lo expuesto anteriormente, puede observarse que si bien existe un área de investigación muy activa relacionada a la calidad de la información en Wikipedia, estos estudios están centrados casi mayoritariamente en idioma inglés, no existiendo de acuerdo a nuestro conocimiento, estudios similares con Wikipedia en español.

Los antecedentes de nuestro grupo en esta temática [11, 15] se han basado en la versión inglesa de Wikipedia. Es por eso, que se tiene como objetivo general realizar una primera aproximación al problema del análisis de la calidad de la información de Wikipedia en español. En particular, nos centraremos en aspectos vinculados al soporte para la categorización automática de ADs y de algunos defectos de calidad más frecuentes.

En este contexto, un objetivo parcial a cumplir será el estudio y análisis comparativo del estado de Wikipedia en español con respecto a otra más desarrollada como es Wikipedia en inglés. También se estudiará la factibilidad de crear un pequeño corpus de entrenamiento con artículos destacados para verificar si alguna de las conclusiones obtenidas con técnicas sencillas como las utilizadas en [9] y [17] también se verifican con los artículos de Wikipedia en español. Finalmente, como un objetivo secundario, se prevé la realización de un estudio preliminar tendiente a analizar y comparar la situación de los defectos de calidad de Wikipedia en español con respecto a su versión en inglés y, en caso de ser posible realizar un trabajo inicial con alguno de los defectos más frecuentes.

## Formación de Recursos Humanos

Trabajos de tesis vinculados con las temáticas descritas previamente:

- 1 tesis de Licenciatura en ejecución.

## Referencias

- [1] M. Anderka and B. Stein. A Breakdown of Quality Flaws in Wikipedia. In Castillo et al. [10], pages 11–18.
- [2] M. Anderka and B. Stein. Overview of the 1th international competition on quality flaw prediction in wikipedia. In Forner et al. [12].
- [3] M. Anderka, B. Stein, and M. Busse. On the Evolution of Quality Flaws and the Effectiveness of Cleanup Tags in the English Wikipedia. In *Wikipedia Academy 2012*. Wikipedia, July 2012.
- [4] M. Anderka, B. Stein, and N. Lipka. In *20th ACM International Conference on Information and Knowledge Management (CIKM 2011)*. ACM.



- [5] M. Anderka, B. Stein, and N. Lipka. Towards automatic quality assurance in wikipedia. In *20th International Conference on World Wide Web*, 2011.
- [6] M. Anderka, B. Stein, and N. Lipka. Predicting Quality Flaws in User-generated Content: The Case of Wikipedia. In *35th Annual SIGIR Conference*. ACM, 2012.
- [7] D. Anthony, S. Smith, and T. Williamson. Reputation and reliability in collective goods: The case of the online encyclopedia wikipedia. *Rationality & Society*, 21(3), 2009.
- [8] R. Baeza-Yates. User generated content: how good is it? In *3rd Workshop on information credibility on the Web*, 2009.
- [9] J. E. Blumenstock. Size matters: word count as a measure of quality on wikipedia. In *17th International Conference on World Wide Web*, 2008.
- [10] C. Castillo, Z. Gyongyi, A. Jatowt, and K. Tanaka, editors. *Proceedings of the 2nd Joint WICOW/AIRWeb Workshop on Web Quality*. ACM, 2012.
- [11] E. Ferretti, D. H. Fusilier, R. Guzmán-Cabrera, M. M. y Gómez, M. Errecalde, and P. Rosso. On the use of pu learning for quality flaw prediction in wikipedia. In Forner et al. [12].
- [12] P. Forner, J. Karlgren, and C. Womser-Hacker, editors. *CLEF 2012 Evaluation Labs and Workshop, Online Working Notes, Rome, Italy, September 17-20, 2012*, 2012.
- [13] L. Gaio, M. den Besten, A. Rossi, and J. Dalle. Wikibugs: using template messages in open content collections. In *5th Symposium on wikis and open collaboration*, 2009.
- [14] M. Hu, E. Lim, A. Sun, H. Lauw, and B. Vuong. Measuring article quality in wikipedia: models and evaluation. In *CIKM*. ACM, 2007.
- [15] E. Lex, M. Voelske, M. Errecalde, E. Ferretti, L. Cagnina, C. Horn, B. Stein, and M. Granitzer. Measuring the quality of web content using factual information. In Castillo et al. [10].
- [16] A. Lih. Wikipedia as participatory journalism: reliable sources? metrics for evaluating collaborative media as a news resource. In *5th international symposium on online journalism*, 2004.
- [17] N. Lipka and B. Stein. Identifying Featured Articles in Wikipedia: Writing Style Matters. In *19th International Conference on World Wide Web*, 2010.
- [18] B. Liu, Y. Dai, X. Li, W. S. Lee, and P. S. Yu. Building text classifiers using positive and unlabeled examples. In *3rd IEEE International Conference on Data Mining*, 2003.
- [19] T. Redman. *Data Quality for the Information Age*. Artech House, 1996.
- [20] B. Stvilia, M. Twidale, L. Smith, and L. Gasser. Assessing information quality of a community-based encyclopedia. In *ICIQ*, 2005.
- [21] B. Stvilia, M. Twidale, L. C. Smith, and L. Gasser. Information quality work organization in wikipedia. *JASIST*, 59(6):983–1001, 2008.
- [22] D. Tax. *One-class classification*. PhD thesis, Technische Universiteit Delft, 2001.
- [23] R. Y. Wang and D. M. Strong. Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, 12(4), 1996.
- [24] D. Wilkinson and B. Huberman. Cooperation and quality in wikipedia. In *3rd International symposium on wikis and open collaboration*. ACM, 2007.

# Integración de Modelos de Razonamiento Práctico, Tecnologías de Acuerdo y Aprendizaje

Cecilia Sosa Toranzo, Marcelo Errecalde, Edgardo Ferretti

Laboratorio de Investigación y Desarrollo en Inteligencia Computacional  
Departamento de Informática, Universidad Nacional de San Luis  
Ejército de los Andes 950 - (D5700HHW) San Luis - Argentina  
e-mails: {csosatoranzo, merreca, ferretti}@unsl.edu.ar

## Resumen

Este artículo describe, en forma breve, los trabajos de investigación y desarrollo que se están llevando a cabo en la línea de investigación “Agentes Inteligentes” en la problemática vinculada a la integración de tecnologías de acuerdo y aprendizaje dentro de arquitecturas de razonamiento práctico. Se abordan temas vinculados a arquitecturas de agentes inteligentes (en particular, arquitecturas BDI y Procesos de Decisión Markov), mecanismos de votación, argumentación y negociación como tecnologías de acuerdo, y aprendizaje automático.

**Palabras clave:** sistemas multi-agente, tecnologías de acuerdo, toma de decisión, aprendizaje automático.

## Contexto

La línea de investigación “Agentes Inteligentes” forma parte del proyecto “Herramientas y mecanismos para la toma de decisiones en agentes inteligentes artificiales”, Proyecto de Investigación consolidado de la Universidad Nacional de San Luis. En este proyecto el objetivo principal es avanzar en la integración de las investigaciones sobre herramientas para la extracción y análisis inte-

ligente de contenido Web de calidad y desarrollo e integración de modelos y mecanismos efectivos para la toma de decisiones, y el aprendizaje automático.

Las dos fuentes de financiamiento de este proyecto son la Universidad Nacional de San Luis y la Comisión Europea de Investigación e Innovación, a través del programa Marie Curie Actions: FP7 People 2010 IRSES.

## Introducción

Aunque no existe en la actualidad un consenso en la definición de *agente artificial inteligente* (AAI), tal vez la acepción más comúnmente aceptada es la dada por Wooldridge y Jennings [28], donde un agente es un sistema computacional basado en hardware o software que posee las siguientes características:

- Autonomía: opera sin la intervención directa de humanos u otros agentes, y tiene algún tipo de control sobre sus acciones y estado interno.
- Sociabilidad: puede interactuar y comunicarse, cuando es apropiado, con otros agentes o humanos.
- Reactividad: percibe su ambiente y responde de manera oportuna a los cambios que se producen en él.

- Pro-actividad: puede exhibir comportamiento dirigido por un objetivo y tomar la iniciativa cuando sea apropiado.

Otra característica de agencia a considerar en ambientes no determinísticos, desconocidos y altamente dinámicos que presentan la mayoría de los problemas de la vida real, es la capacidad de **aprendizaje y adaptación** que tienen los agentes artificiales “embebidos” en este tipo de ambientes. Como es bien sabido, este tipo de capacidades son cruciales para hacer frente a la potencial falta de conocimiento inicial de los agentes como así también para la actualización del conocimiento que se tiene sobre un ambiente y/o agentes que cambian a lo largo del tiempo. De hecho, en la bibliografía fundamental relacionada a los agentes artificiales inteligentes [18], se considera incluso que la capacidad de aprendizaje automático de los agentes directamente determina su grado de autonomía, y que un agente será autónomo en la medida que tenga la capacidad de aprender a determinar cómo tiene que compensar el conocimiento incompleto o parcial que le fue proporcionado inicialmente por su diseñador.

Proveer a un agente de las capacidades necesarias para exhibir el comportamiento flexible descrito anteriormente, no es una tarea trivial. De hecho, una de las áreas de investigación más activa en el ámbito de agentes ha sido la definición de modelos y arquitecturas de agentes que intentan dar una respuesta a este problema. Existen arquitecturas que se han concentrado en el aspecto de la reactividad [1, 6] y otras en cambio que han privilegiado los mecanismos de deliberación y planificación necesarios para proveer un comportamiento pro-activo [15]. Sin embargo, existe actualmente un consenso generalizado de que cualquier arquitectura realista de agente, debería proveer un soporte adecuado para todas estas capacidades. Las arquitecturas híbridas [9, 10] y las basadas en comportamientos [14, 16] han intentado lograr un balance adecuado entre pro-actividad y reactividad. Sin embargo, las arquitecturas que mayor atención han recibido para este pro-

pósito son las denominadas **arquitecturas BDI** [5]. Similarmente, aquellas personas que se han orientado a los enfoques basados en utilidad, han tomado a los **Procesos de Decisión Markov** [3] y los juegos estocásticos [4], como los modelos principales para diseñar e implementar agentes inteligentes.

En todas las arquitecturas mencionadas previamente, el fin último de los procesos internos y de razonamiento de los agentes, es seleccionar las acciones que éste realizará sobre su ambiente. En este contexto, estos procesos orientados hacia las acciones suelen ser referenciados como de *razonamiento práctico* [27], para diferenciarlo del razonamiento teórico, que sólo afecta las creencias del agente sobre el mundo. Sin embargo, desde un punto de vista más general, estas propuestas también pueden ser consideradas como modelos y arquitecturas para la *toma de decisiones de agentes inteligentes*.

Por otra parte, en un *sistema multi-agente* (SMA) [27] un conjunto de agentes interactúa para conseguir algún objetivo difícil de lograr de forma individual. Usualmente, en este tipo de sistemas, cada agente posee información incompleta sobre su entorno y sobre sí mismo, teniendo además capacidades limitadas. El entorno es dinámico y cambiante, viéndose afectado por las acciones de todos los agentes. Por esta razón, no puede predecirse con certeza el estado futuro de ese entorno. El control del sistema es distribuido, los datos están descentralizados, y la computación es asincrónica. Las características mencionadas previamente, que aumentan la complejidad del sistema, proveen también las siguientes ventajas: eficiencia (debido al paralelismo que se logra al disponer de múltiples agentes), robustez y confiabilidad (gracias a la redundancia), escalabilidad y reusabilidad.

Actualmente, ha surgido un nuevo paradigma para la próxima generación de sistemas distribuidos y abiertos, donde las interacciones entre agentes computacionales están basadas en el concepto de *acuerdo*. En el pasado, dicho concepto, fue dominio de estudio para filósofos y sociólogos, y fue aplicado

solamente a la sociedad humana. Sin embargo, hoy en día, el acuerdo y todos los procesos y mecanismos involucrados en lograr acuerdos entre diferentes agentes son también tema de análisis desde perspectivas orientadas a la tecnología, dando lugar a sistemas computacionales donde las componentes interactúan utilizando protocolos de argumentación [2, 11], votación [8, 14, 25], negociación [17, 12] y mercado [19], entre otros. Dicho enfoque constituye la visión denominada **Tecnologías de Acuerdo** (en inglés, *agreement technologies (AT)*) [13] que engloba métodos, plataformas y herramientas para definir, especificar y verificar tales sistemas. Este nuevo paradigma ha quedado de manifiesto en el surgimiento de consorcios de investigación europeos especializados en la temática<sup>1</sup> como así también en la Primera Conferencia Internacional de Tecnologías de Acuerdo (AT-2012) realizada en Octubre de 2012 en Croacia<sup>2</sup> y la segunda (AT-2013) a realizarse en Agosto del corriente año en China.<sup>3</sup>

Si bien las tecnologías de acuerdo han recibido una atención considerable en el área de SMA, su aplicación dentro de las arquitecturas y modelos de agentes individuales para la toma de decisiones racional es un área que no ha recibido, hasta el momento, la necesaria atención. Más allá de trabajos aislados sobre el uso de votación en sistemas basados en comportamiento [14, 16] y algunos esfuerzos por incorporar argumentación y/o votación en arquitecturas BDI [20, 21, 25] existe una falencia notoria de propuestas tendientes a sistematizar el uso de estas tecnologías en modelos de toma de decisión ya establecidos, como las arquitecturas BDI y las arquitecturas basadas en utilidades.

Este no es un dato secundario si consideramos que el diseño modular de los agentes, o bien la existencia de múltiples criterios o preferencias personales que pueden entrar en conflicto, conducen naturalmente a situaciones que podrán ser modeladas como un SMA

que funciona “dentro de la cabeza del agente”.

Por otra parte, si bien el aprendizaje automático (en inglés, *Machine Learning (ML)*) ha sido ampliamente desarrollado en modelos como los MDPs [26], no existe una clara especificación de cómo estos procesos podrían llevarse a cabo en enfoques basados en arquitecturas BDI, o de qué manera ML y AT se podrían complementar y/o asistir mutuamente. De igual manera, y a excepción de algunos trabajos recientes en el área [22, 23, 24], tampoco existen suficientes investigaciones dedicadas a analizar la relación entre los MDPs y las arquitecturas BDI, lo que permitiría en primer lugar, integrar las ventajas y atenuar o solucionar los problemas que cada una de ellas exhibe por separado.

Esta idea de integración también ha sido planteada por el Massachusetts Institute of Technology [7] en respuesta al estancamiento producido en el área de Inteligencia Artificial respecto a algunos objetivos planteados inicialmente en la disciplina: “*En lugar de buscar la bala de plata para resolver el problema, estamos considerando un amplio rango de modelos, probando integrarlos y agregarlos*”.

## Líneas de Investigación y desarrollo

En base a lo anterior, se puede apreciar la disposición de enfoques y modelos muy potentes para el desarrollo de agentes inteligentes como los MDPs, las arquitecturas BDI, las AT y ML. Por otra parte, si bien la integración de algunos de estos enfoques ha comenzado a realizarse recientemente, no existe, de acuerdo a nuestro conocimiento, propuestas que avancen en una integración de todos ellos en forma simultánea y consideramos, que esto permitiría modelos efectivos, eficientes y flexibles para la toma de decisiones de alto nivel de AAI. Por lo tanto, es en esta problemática donde se pretende generar las principales contribuciones, teniendo como ejes principales de investigación: 1) Arquitecturas de Razonamiento Práctico, particularmente los MDPs y las arquitecturas

<sup>1</sup><http://www.agreement-technologies.org>

<sup>2</sup><http://at2012.tel.fer.hr/>

<sup>3</sup><http://www.ia.urjc.es/at2013/>



BDI; 2) Tecnologías de Acuerdo; y 3) Aprendizaje automático.

## Resultados y objetivos

Las principales motivaciones surgen de la intención de dar respuestas a las falencias descritas previamente, y plantear contribuciones en el área a partir del cumplimiento de los siguientes objetivos:

1. Lograr un entendimiento más profundo de los distintos enfoques basados en AT y aprendizaje automático para su uso en distintas componentes y procesos de modelos de razonamiento práctico, como las arquitecturas BDI y algoritmos basados en MDP y sus variantes. Este aprendizaje de los enfoques, debería incluir la determinación y comparación de las propiedades formales y los aspectos computacionales que involucran estas técnicas y modelos.
2. Obtener, mediante un proceso adecuado de generalización, una arquitectura abstracta para la incorporación de AT y aprendizaje en modelos basados en MDPs y BDI. La idea es lograr un modelo y una arquitectura genérica donde estas componentes puedan ser fácilmente integradas e intercambiadas. En este sentido, nuestra idea es lograr un marco de trabajo genérico para la integración de AT dentro de las arquitecturas de razonamiento práctico al igual que el soporte para el aprendizaje.
3. Determinar, mediante la instanciación e implementación de la arquitectura propuesta en el punto anterior, en qué medida se adecúa a la resolución de problemas concretos del mundo real. Como potenciales dominios de aplicación se prevé trabajar con robots o dispositivos de comunicación móviles (agentes de hardware) y problemas vinculados al área de Web Intelligence.

En este contexto, ya se han logrado algunos avances en estos temas, relacionados

al uso de votación en los procesos internos del agente [8, 25], mientras que en [20, 21] se desarrolló la primera arquitectura concreta de agente que integra al modelo BDI con servicios Web y razonamiento argumentativo en un mismo framework.

## Formación de Recursos Humanos

Trabajos de tesis vinculados con las temáticas descritas previamente:

- 1 tesis Doctoral en ejecución.

## Referencias

- [1] R. C. Arkin. *Behaviour-Based Robotics*. The MIT Press, 1998.
- [2] J. Bentahar, R. Alam, and Z. Maaamar. An argumentation-based protocol for conflict resolution. In *Workshop on Knowledge Representation for Agents and Multi-Agent Systems*, 2008.
- [3] C. Boutilier, T. Dean, and S. Hanks. Decision-theoretic planning: Structural assumptions and computational leverage. *J. Artif. Intell. Res. (JAIR)*, 1999.
- [4] M. Bowling and M. Veloso. An analysis of stochastic game theory for multiagent reinforcement learning. Technical Report 00-165, Computer Science Department, Carnegie Mellon University, 2000.
- [5] M. Bratman, D. Israel, and M. Pollack. Plans and resource bounded reasoning. *Computational Intelligence*, 1988.
- [6] R. A. Brooks. A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, 1986.
- [7] D. Chandler. *Rethinking artificial intelligence*. MIT News Office, 2009.
- [8] M. Errecalde, G. Aguirre, and F. González. Agentes y mecanismos de votación. In *X CACIC*, 2004.

- [9] E. Gat. Integrating planning and reacting in a heterogeneous asynchronous architecture for mobile robots. *SIGART Bulletin*, 1991.
- [10] E. Gat. On three-layer architectures. In *Artificial Intelligence and Mobile Robots*, 1998.
- [11] A. Kakas, N. Maudet, and P. Moraitis. Layered strategies and protocols for argumentation-based agent interaction. In *Argumentation in Multi-Agent Systems*, pages 64–77, 2005.
- [12] H. J. Müller. Negotiation principles. In *Foundations of distributed artificial intelligence*. John Wiley & Sons, 1996.
- [13] S. Ossowski. *Agreement Technologies*. Law, governance and technology series. Springer, 2013.
- [14] P. Pirjanian. *Multiple Objective Action Selection and Behaviour Fusion using Voting*. PhD thesis, Department of Medical Informatics and Image Analysis, Institute of Electronic Systems, Aalborg University, 1998.
- [15] M. E. Pollack. Planning technology for intelligent cognitive orthotics. In *Proceedings of 6th International Conference on AI Planning and Scheduling*, 2002.
- [16] J. Rosenblatt. Damn: A distributed architecture for mobile navigation - thesis summary. In *JETAI*, 1995.
- [17] J. S. Rosenschein and G. Zlotkin. *Rules of Encounter - Designing Conventions for Automated Negotiation among Computers*. The MIT Press, 1998.
- [18] S. Russel and P. Norvig. *Artificial Intelligence - A Modern Approach*. Prentice Hall, 2010.
- [19] T. Sandholm. Distributed rational decision making. In *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*, chapter 5. The MIT Press, 1999.
- [20] F. Schlesinger, M. Errecalde, and G. Aguirre. An approach to integrate web services and argumentation into a BDI system. In *9th Intl. Conf. on Autonomous Agents and Multiagent Systems*, 2010.
- [21] F. Schlesinger, E. Ferretti, M. Errecalde, and G. Aguirre. An Argumentation-based BDI Personal Assistant. In *23rd IEA-AIE*, volume 6096 of *LNAI*. Springer, 2010.
- [22] M. Schut, M. Wooldridge, and S. Parsons. On partially observable mdps and bdi models. In *Selected papers from the UKMAS Workshop on Foundations and Applications of Multi-Agent Systems*. Springer-Verlag, 2002.
- [23] G. I. Simari. *Markov Decision Processes and the Belief-Desire-Intention Model: Bridging the Gap for Autonomous Agents*. Springerbriefs in Computer Science. Springer New York, 2011.
- [24] G. I. Simari and S. Parsons. On the relationship between MDPs and the BDI architecture. In *AAMAS*, 2006.
- [25] C. Sosa, F. Schlesinger, E. Ferretti, and M. Errecalde. Integrating a voting protocol within an argumentation-based BDI system. In *XVI CACIC*, 2010.
- [26] R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. Adaptive Computation and Machine Learning Series. Bradford Book, 1998.
- [27] M. Wooldridge. *An Introduction to MultiAgent Systems*. John Wiley & Sons, Chichester, England, 2002.
- [28] M. Wooldridge and N. R. Jennings. Intelligent agents: Theory and practice. *Knowledge Engineering Review*, 10:115–152, 1995.

# Uso de técnicas metaheurísticas avanzadas para resolver problemas de optimización combinatoria

Gabriela Minetti<sup>1</sup>, Carolina Salto<sup>1</sup>, Hugo Alfonso<sup>1</sup> y Fernando Sanz Troiani<sup>2</sup>

Laboratorio de Investigación en Sistemas Inteligentes (LISI)

Facultad de Ingeniería - Universidad Nacional de La Pampa

Calle 110 Esq. 9 (6360) General Pico - La Pampa - Rep. Argentina

Te. / Fax: (02302) 422780/422372, Int. 6302

e-mail: <sup>1</sup>{minettig, saltoc, alfonsoh@ing.unlpam.edu.ar}, <sup>2</sup>fstnando@gmail.com

**Resumen** La finalidad de esta línea de investigación es el estudio y resolución de problemas de optimización combinatoria mediante la utilización de métodos aproximados. Particularmente, nuestro trabajo se enfoca en el análisis y desarrollo de algoritmos metaheurísticos basados en trayectoria y en población, así como también híbridos, que permitan resolver eficientemente problemas genéricos como es el caso de QAP y problemas específicos y del mundo real como FAP y TSP. También consideramos la posibilidad de distribuir y/o paralelizar estos métodos dependiendo de la complejidad del problema a resolver.

**Palabras claves:** Bioinformática, ADN, metaheurísticas, secuenciamiento de un genoma, ensamblado de fragmentos, métodos de búsqueda híbrida y distribuida, optimización combinatoria.

## CONTEXTO

El proyecto de investigación en el cual se enmarca esta línea de investigación se denomina "Resolviendo problemas complejos con técnicas metaheurísticas avanzadas". Este proyecto es dirigido por la Dra. Carolina Salto y llevado a cabo en el Laboratorio de Investigación de Sistemas Inteligentes (LISI), de la Facultad de Ingeniería de la Universidad Nacional de La Pampa. Los integrantes de este laboratorio mantienen desde hace varios años una importante vinculación con investigadores de la Universidad Nacional de San Luis (Argentina) y de la Universidad de Málaga (España), con quienes se han realizado varias publicaciones conjuntas.

## I. INTRODUCCIÓN

Como es bien conocido, el problema de asignación cuadrática (QAP, del Inglés Quadratic Assignment Problem) es un problema de optimización combinatoria NP-duro que forma parte del núcleo de muchos problemas del mundo real [1]. Diversas aplicaciones en áreas tan diferentes como Investigación Operativa, Optimización Combinatoria y Computación Paralela y Distribuida [2], [3], [4], [5], [6], [7], [8], [9], [10] pueden modelarse por medio de este problema. En particular, el interés de nuestra línea de investigación se centra en los problemas combinatorios, tales como: el problema del viajante de comercio (TSP del Inglés Travelling Salesman Problem), el problema de ensamblado de fragmentos de ADN

(FAP del Inglés Fragment Assembly Problem), diversos problemas de planificación de tareas, entre otros.

En general, la dificultad de este tipo de problemas radica en buscar la solución óptima en un espacio de soluciones de elevada complejidad. En el caso de los métodos de búsqueda exactos, esta dificultad se traduce en tiempos de cómputos inviables para hallar una solución óptima. En cambio, los métodos aproximados, especialmente las metaheurísticas, reducen considerablemente la complejidad temporal para obtener soluciones óptimas o quasi-óptimas. En consecuencia, esto implica una reducción significativa de los costos de cualquier organización a la hora de resolver esta clase de problemas.

Las metaheurísticas obtienen soluciones que cumplen con la calidad requerida y los tiempos de demora impuestos en el campo industrial, además permiten estudiar clases genéricas de problemas en lugar de instancias de problemas particulares. En general, las técnicas que obtienen mejores resultados (en precisión y esfuerzo) al resolver problemas complejos y del mundo real usan metaheurísticas. Sus campos de aplicación van desde la optimización combinatoria, bioinformática, telecomunicaciones a la economía, software ingeniería, etc, que necesitan soluciones rápidas con alta calidad [11].

Existen diferentes formas de clasificar y describir las técnicas metaheurísticas [12], [13], [14]. Dependiendo de las características que se seleccionen se pueden obtener diferentes taxonomías: basadas en la naturaleza y no basadas en la naturaleza, con memoria o sin ella, con una o varias estructuras de vecindario, etc. Una de las clasificaciones más populares las divide en *metaheurísticas basadas en trayectoria* y *basadas en población*. Las primeras realizan búsquedas orientadas a la explotación del espacio de soluciones al manipular en cada paso un único elemento de dicho espacio. En tanto que, las segundas son métodos orientados a la exploración del espacio de búsqueda, ya que trabajan sobre un conjunto de elementos (población). Esta taxonomía se muestra de forma gráfica en la figura 1, que además incluye las principales metaheurísticas.

En los últimos años, el interés en combinar metaheurísticas (*metaheurísticas híbridas*) ha aumentado considerablemente en el área de optimización [15]. Combinaciones de algorit-

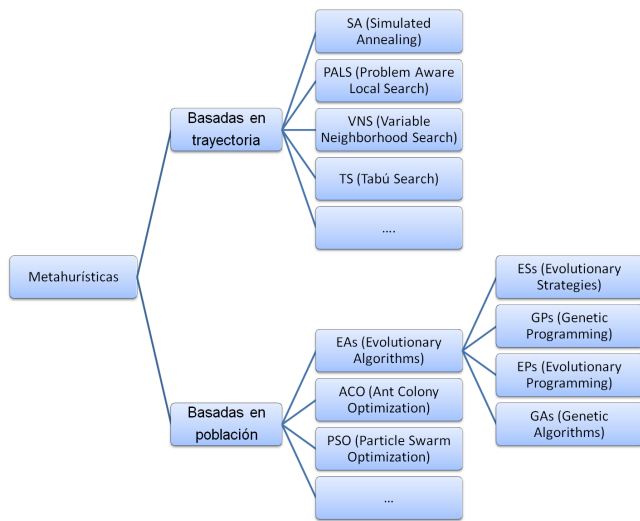


Fig. 1. Clasificación de las metaheurísticas.

mos tales como metaheurísticas basadas en población y/o en trayectoria, programación matemática, programación con restricciones y técnicas de aprendizaje automático (machine learning) proveen algoritmos de búsqueda muy poderosos.

Los objetivos de esta línea de investigación son: analizar, diseñar y desarrollar algoritmos metaheurísticos basados en trayectoria y en población, así como también híbridos, que permitan resolver eficientemente problemas genéricos como es el caso de QAP y problemas específicos y del mundo real como FAP y TSP.

## II. DESARROLLO

En esta sección describimos los desarrollos que se llevan a cabo en esta línea de investigación, pero primero introducimos una breve descripción de cada uno de los problemas a resolver.

El Problema de Asignación Cuadrática es un problema clásico de optimización combinatorio, en el cual se encuentra un vasto número de problemas de diseño y de distribución de recursos en diferentes campos, donde la decisión a tomar es una asignación de elementos de un conjunto en otro. El QAP es considerado como un problema complejo y dificultoso de resolver y puede establecerse como un conjunto de  $n$  elementos distintos que deben ser localizados (asignados) en  $n$  localidades distintas minimizando el costo. En cada localidad sólo puede haber un elemento y todos los elementos deben ser asignados a una localidad. Por cada par de localidades el costo es calculado como el producto de la distancia entre las localidades y el flujo asociado a los elementos en las localidades. El costo total es la suma de todos los costos asociados a cada par de localidades.

En tanto que, FAP es un problema resuelto en las primeras fases del proyecto del genoma y por lo tanto muy importante, ya que los demás pasos dependen de su precisión. El proceso de ensamblado de fragmentos consiste en: una primera fase de superposición (calcula el *puntaje de solapamiento* entre los fragmentos), una segunda de distribución (encuentra el

orden de los fragmentos basado en el puntaje de similitud computado) y una tercera de consenso (deriva la secuencia de ADN a partir de la distribución anterior). Una resolución óptima del problema se produce cuando el algoritmo es capaz de ensamblar un determinado conjunto de fragmentos en un solo *contig*. Un *contig* es una secuencia en la que el solapamiento entre los fragmentos adyacentes es mayor o igual a un umbral predefinido (parámetro de corte denominado *cutoff*).

El problema del viajante de comercio, es uno de los problemas de optimización combinatoria NP-duros más ampliamente estudiado. Su declaración es engañosamente simple: un viajante busca el camino más corto para pasar por  $m$  ciudades. En otras palabras, una persona debe visitar un conjunto de  $m$  ciudades, comenzando en una ciudad determinada y finalizando en la misma ciudad; luego de haber visitado todas ellas sólo una vez. Esto significa que nunca regresa a una ciudad ya visitada, excepto la primera.

Desde el punto de vista de la optimización combinatoria, la construcción de un consenso (en FAP) es similar a la de un recorrido en un problema del viajante de comercio. Esto es porque cada fragmento tiene una ubicación específica en la formación de una secuencia en la etapa de consenso. Aunque los puntos terminales de un recorrido de TSP sean irrelevantes ya que su solución es un recorrido circular de ciudades, en el caso de FAP estos puntos son importantes ya que ellas representan los extremos opuestos de la secuencia original de ADN. En TSP el ordenamiento de las ciudades es la solución final al problema. En cambio para FAP, el ordenamiento de fragmentos es sólo un resultado intermedio que será utilizado en la fase de consenso. Por otra parte, estos dos problemas son formulados como casos especiales de QAP.

Por un lado, nuestro trabajo consiste en profundizar el estudio sobre la robustez de las metaheurísticas poblacionales basados en colonia de hormigas (*Ant Colony Optimization* -ACO-) [16] para resolver TSP. Con el propósito de estudiar la robustez de una metaheurística se analizan las diferencias entre las soluciones encontradas para las instancias sin y con ruido. Si no se detectan diferencias (estadísticamente significativas), la metaheurística muestra un comportamiento neutro (insensible, indistinto) a pequeñas variaciones en los datos de entrada. Consecuentemente, esta metaheurística se considera robusta para resolver instancias ruidosas. De esta forma, si ACO resultara una metaheurística robusta para TSP podríamos adaptarla fácilmente para resolver las instancias con ruido de FAP.

Por otra parte analizamos el comportamiento de los ensambladores metaheurísticos que hasta el momento han mostrado un mejor desempeño: ISA (*Inversion Simulated Annealing*), PALS (*Problem Aware Local Search*) y GAG<sub>50</sub> (*Genetic Algorithm with a Greedy seeding strategy*) [17], [18], [19]. Los ensambladores ISA, PALS y GAG<sub>50</sub> han logrado una mejor calidad de los resultados, ya sea medida por medio del *fitness* o por el número final de contigs. Además, han sido capaces de resolver eficazmente un conjunto completo de instancias



con un esfuerzo computacional mínimo [17], [19], [20], [21]. Teniendo en cuenta las virtudes de estos ensambladores desarrollamos una nueva metaheurística híbrida para resolver instancias de mayor tamaño de FAP, denominada SAX. El objetivo perseguido con SAX es combinar la potencialidad de ISA como ensamblador de fragmentos con operadores de recombinación genéticos como intensificadores de la búsqueda.

Finalmente, dado que SAX y PALS demostraron resolver eficientemente el problema de ensamblado de fragmentos [22], [18] analizaremos el comportamiento de estas dos metaheurísticas en la resolución del problema genérico QAP.

### III. RESULTADOS OBTENIDOS/ESPERADOS

En esta sección presentamos los resultados obtenidos de nuestra investigación en el transcurso del año 2012 y los esperados en el 2013.

Con respecto al análisis de robustez de ACO, frente a la presencia de ruido en instancias de TSP, esperamos encontrar evidencia empírica que demuestre fehacientemente dicha característica. Estos resultados nos permitirán una mejor adaptación de esta técnica metaheurística a la resolución de instancias con o sin ruido de FAP.

Por otra parte, desarrollamos un nuevo ensamblador, denominado SAX, para resolver FAP [22]. Este ensamblador es la combinación de una heurística basada en trayectoria (ISA) con un operador genético de cruce basado en el orden (OX, *Order Crossover*). El componente ISA le permite a SAX utilizar el procedimiento por inversión para generar dos nuevas soluciones y así extender la exploración hacia regiones de búsqueda prometedoras, además de, utilizar el criterio de Boltzman para escapar de óptimos locales. El componente OX incrementa la explotación de las regiones prometedoras en el espacio de búsqueda. De esta forma, SAX encuentra distribuciones óptimas con altos puntajes de superposición para todas las instancias de FAP.

Finalmente esperamos encontrar resultados que demuestren empíricamente que SAX y PALS son aplicables al grupo de problemas derivados de QAP. Es importante destacar que, tanto en esta investigación como en las dos anteriores, las conclusiones obtenidas del análisis de resultados son verificadas rigurosamente a través de los correspondientes tests estadísticos.

### IV. FORMACIÓN DE RECURSOS HUMANOS

Durante el año 2012, integrantes del proyecto han realizado diversos cursos de postgrado directamente relacionados con la temática del proyecto, con el objetivo de llevar a cabo carreras de postgrado en un futuro cercano.

En tanto que, en el LISI se trabaja con alumnos avanzados en la carrera Ingeniería en Sistemas en temas relacionados a la resolución de problemas de optimización usando técnicas inteligentes, con el objeto de guiarlos en el desarrollo de sus tesis de grado y, también, de formar futuros investigadores.

### REFERENCES

- [1] M. Garey and D. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman, 1979.
- [2] R. S. Bhaba, E. W. Wilbert, and L. H. Gary, "Locating sets of identical machines in a linear layout," *Ann. Oper. Res.*, vol. 77, pp. 183–207, 1998.
- [3] S. P. Coy, B. L. Golden, G. C. Runger, and E. A. Wasil, "See the forest before the trees: fine-tuned learning and its application to the traveling salesman problem," *IEEE Transactions on Systems, Man, and Cybernetics, Part A*, pp. 454–464, 1998.
- [4] D. Chiang and L. C. Potter, "Minimax non-redundant channel coding for vector quantization," in *Proceedings of the 1993 IEEE international conference on Acoustics, speech, and signal processing: image and multidimensional signal processing - Volume V*, ser. ICASSP'93. Washington, DC, USA: IEEE Computer Society, 1993, pp. 617–620. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1947148.1947321>
- [5] D. Gamboa, C. Rego, and F. Glover, "Implementation analysis of efficient heuristic algorithms for the traveling salesman problem," *Comput. Oper. Res.*, vol. 33, no. 4, pp. 1154–1172, Apr. 2006. [Online]. Available: <http://dx.doi.org/10.1016/j.cor.2005.06.014>
- [6] A. Mason and M. Rönnqvist, "Solution methods for the balancing of jet turbines," *Computers & Operations Research*, vol. 24, pp. 153–167, 1998.
- [7] J. W. Pepper, B. L. Golden, and E. A. Wasil, "Solving the traveling salesman problem with annealing-based heuristics: a computational study," *Trans. Sys. Man Cyber. Part A*, vol. 32, no. 1, pp. 72–77, Jan. 2002. [Online]. Available: <http://dx.doi.org/10.1109/3468.995530>
- [8] G. G. Polak, "On a special case of the quadratic assignment problem with an application to storage-and-retrieval devices," *Annals of Operations Research*, vol. 138, pp. 223–233, 2005.
- [9] R. C., "Relaxed tours and path ejections for the traveling salesman problem," *European Journal of Operational Research*, vol. 106, no. 2, pp. 522–538, 1998.
- [10] Q. Zhang, J. Sun, and E. Tsang, "An evolutionary algorithm with guided mutation for the maximum clique problem," *IEEE Transactions on Evolutionary Computation*, vol. 9, no. 2, pp. 192–200, April 2005.
- [11] E. Alba, G. Luque, and S. Nesmachnow, "Parallel metaheuristics: recent advances and new trends," *International Transactions in Operational Research*, vol. 20, no. 1, pp. 1–48, 2013. [Online]. Available: <http://dx.doi.org/10.1111/j.1475-3995.2012.00862.x>
- [12] E. Cantú-Paz, "Migration, selection pressure, and superlinear speedups," in *Efficient and Accurate Parallel Genetic Algorithms*. Kluwer Academic Publishers, 2000, pp. 97–120.
- [13] T. G. Crainic and M. Toulouse, "Parallel strategies for metaheuristics," in *Handbook of Metaheuristics*. Kluwer Academic Publishers, 2003, pp. 475–513.
- [14] C. Blum and A. Roli, "Metaheuristics in combinatorial optimization: Overview and conceptual comparison," *ACM Comput. Surv.*, vol. 35, pp. 268–308, September 2003.
- [15] E.-G. Talbi, "A taxonomy of hybrid metaheuristics," *Journal of Heuristics*, vol. 8, pp. 541–564, September 2002. [Online]. Available: <http://portal.acm.org/citation.cfm?id=594957.595102>
- [16] M. Dorigo, "Optimization, learning and natural algorithms," Ph.D. dissertation, Politecnico di Milano, Italy, 1992.
- [17] G. Minetti, G. Luque, E. Alba, and G. Leguizamón, "A new Hybrid SA for Solving the DNA Fragment Assembly Problem," in *XXVIII International Conference of the Chilean Computing Science Society (SCCC)*, November 2009, pp. 109–116.
- [18] E. Alba and G. Luque, "A new local search algorithm for the dna fragment assembly problem," in *Evolutionary Computation in Combinatorial Optimization, EvoCOP'07*, ser. Lecture Notes in Computer Science. Valencia, Spain: Springer, 2007, vol. 4446, pp. 1–12.
- [19] G. Minetti, E. Alba, and G. Luque, "Seeding strategies and recombination operators for solving the DNA fragment assembly problem," *Information Processing Letters*, vol. 108, no. 3, pp. 94–100, October 2008.
- [20] G. Minetti and E. Alba, "Metaheuristic assemblers of DNA strands: Noiseless and noisy cases," in *2010 IEEE Congress on Evolutionary Computation (CEC)*, July, pp. 1–8.
- [21] G. Minetti, G. Leguizamón, and E. Alba, "Assembling DNA Sequences Containing Noisy Information With Metaheuristic Algorithms," *Journal of Information Sciences, Elsevier (in evaluation)*, 2011.

- [22] G. F. Minetti, M. G. Leguizamón, and E. A. Torres, "SAX: a new and efficient assembler for solving DNA Fragment Assembly Problem," in *41JAIIO Jornadas Argentinas de Informática*, S. A. de Informática (SADIO), Ed. SADIO, 2012, pp. 177–188.

# Agentes inteligentes en ambientes dinámicos: Simplificación de Texto

Enzo Ferrarini

Sandra Roger

email: {ferrarini.enzo.nqn,giuvago}@gmail.com

*Grupo de Investigación en Lenguajes e Inteligencia Artificial*

Departamento de Teoría de la Computación

Facultad de Informática

UNIVERSIDAD NACIONAL DEL COMAHUE

Buenos Aires 1400 - (8300)Neuquén - Argentina

## Resumen

La meta fundamental de este proyecto está puesta tanto en los formalismos de planificación y razonamiento rebatible para la creación y control de agentes inteligentes, como en el impacto que tienen las tecnologías del lenguaje humano (TLH) en la inclusión social. En estos escenarios, el razonamiento, la toma de decisiones, la planificación de acciones y el aprendizaje ocurren bajo restricciones de tiempo críticas y en intensa interacción con el usuario.

Una de las líneas de investigación es la simplificación de texto. El objetivo de esta es el desarrollo de tecnologías para facilitar el acceso a la información para todas aquellas en el que la inteligibilidad del texto es de gran importancia.

**Palabras Clave:** AGENTES INTELIGENTES, LENGUAJE NATURAL, SIMPLIFICACIÓN DE TEXTO, ACCESIBILIDAD.

## Contexto

Este trabajo está parcialmente financiado por la Universidad Nacional del Comahue, en el contexto del proyecto de investigación *Agentes inteligentes en ambientes dinámicos*. El

proyecto de investigación tiene prevista una duración de cuatro años, desde enero del 2013 hasta diciembre de 2016.

## 1. Introducción

La simplificación de texto es una metodología utilizada en muchos idiomas para conocer las necesidades de los lectores que tienen diferentes niveles educativos y de capacidad de comprensión. Además proporciona una solución para los creadores de contenidos interesados en mejorar la sencillez y, por tanto, la accesibilidad de su contenido.

La forma en la que están estructuradas las sentencias, sean estas por su longitud y complejidad, ocasionan un gran problema tanto a nivel de comprensión como a nivel de entendimientos del texto en sí mismo. Como se menciona en [2]: “Si la complejidad del texto puede hacerse más simple, las oraciones resultarán más fáciles de procesar tanto para las personas como para los programas”.

Las características claves en el proceso de comprensión de lectura de un texto son: el texto, el lector y las circunstancias en que se celebre la reunión. El análisis de estudios publicados muestra que la comprensión lectora implica varios factores que se pueden dividir en tres

grupos: i) en el texto, ii) para el lector, y iii) en relación con intervención pedagógica. Entre los factores relacionados con el texto se destacan, tradicionalmente, la legibilidad (formato del texto) y la inteligibilidad (uso frecuente de palabras y estructuras sintácticas menos complejas). Es bien sabido que las sentencias largas, con varios niveles de subordinación, cláusulas incrustadas (relativa), las frases en voz pasiva, el uso de no-canónico para que los componentes de una frase, y el uso de palabras de baja frecuencia aumentan la complejidad de un texto para lectores con problemas de comprensión, por ejemplo, analfabetos funcionales, la afasia y la dislexia [6].

En la actualidad, también hay una preocupación con la macroestructura del texto más allá de la microestructura. En esta última se observan otros factores, los cuales son vistos como facilitadores en la comprensión de texto, tal como la organización, la cohesión, la coherencia, o conceptos del texto sensibles al lector. Este último presenta características que pueden facilitar la comprensión como aproximación de la anáfora, el uso de marcadores discursivos entre las oraciones o una preferencia por definiciones explícitas [5].

En este trabajo se pone principal énfasis, en primera instancia, en la recopilación organizada y metodológica que provee la revisión sistemática sobre la simplificación de texto. Posteriormente se trabajará sobre las implementaciones necesarias que se desprendan de la revisión descripta anteriormente.

## 2. Líneas de investigación y desarrollo

El proyecto de investigación *Agentes inteligentes en ambientes dinámicos* tiene varios objetivos generales. Por un lado, el de desarrollar conocimiento especializado en el área de Inteligencia Artificial. Además, se estudian técnicas de representación de conocimiento y razonamiento, junto con métodos de planificación y tecnologías del lenguaje natural aplicadas al desarrollo de sistemas multiagentes.

La producción de textos con diferentes longitudes y complejidades pueden ser abordadas por la tarea de adaptación del texto (TA), una práctica muy conocida en los ambientes educativos. Young [7] y Burstein [1] mencionan dos técnicas diferentes para TA: Simplificación de texto y elaboración del texto. La primera puede ser definida como cualquier tarea que reduce la complejidad léxica o sintáctica de un texto al intentar preservar el significado e información. Esta se puede subdividir en simplificación sintáctica, simplificación léxica, resumen automático, y otras técnicas. En cuanto a la elaboración del texto, su objetivo es aclarar y explicar la información al hacer conexiones explícitas en un texto, por ejemplo, proporcionando definiciones cortas o sinónimos de palabras conocidas a sólo unos pocos hablantes de una lengua.

### 2.1. Revisión sistemática

La revisión sistemática comprende una metodología específica y ordenada con la cual es llevada a cabo un estudio. No es lo mismo hablar de una revisión sistemática que de una revisión clásica o narrativa. La diferencia principal entre ambas, es que en esta última, existe la falta de un protocolo que defina los pasos específicos que utilizó el revisor para actualizar el tópico en cuestión. Por ejemplo, estrategias utilizadas para la búsqueda y selección de la bibliografía y métodos para determinar la validez de los mismos. Se puede definir una revisión sistemática como un resumen de la evidencia existente en una determinada fecha usando una metodología bien definida que intenta disminuir, en lo posible, el sesgo de las revisiones narrativas.

La revisión sistemática se obtiene a través de una metodología compuesta de ocho pasos bien definidos [3, 4]. A partir de los cuales, se obtienen revisiones bibliográficas independientes a diferencia de otros sistemas no metodológicos.

Pasos:

1. Propósito de la Revisión de la Literatura: El primer paso en cualquier revisión requiere que el revisor identifique claramente



te el propósito y objetivos previstos de la misma. El propósito debe ser explícito hacia sus lectores.

2. Protocolo y Formación: Toda revisión que implique a más de un crítico, debe fundamentar el procedimiento detallado a seguir. Esto requiere de un documento escrito, el protocolo detallado y la formación para todos los revisores para asegurar la consistencia en la ejecución de la revisión.
3. Búsqueda de la Literatura: El revisor debe ser explícito en la descripción de los detalles de la búsqueda en la literatura, y las necesidades de explicar y justificar la misma.
4. Pantalla Práctica: también conocido como screening para la inclusión; este paso requiere que el revisor sea explícito acerca de los estudios que fueron considerados para la revisión, y cuales fueron eliminados o descartados sin más examen.
5. Evaluación de la calidad: este paso es también conocido como la detección de la exclusión, el revisor debe de forma explícita exponer los criterios para juzgar que artículos son de calidad suficiente para ser incluido en la síntesis de la revisión. Todos los artículos incluidos necesitan ser marcados por su calidad, dependiendo de las metodologías de investigación empleadas por los artículos.
6. Extracción de datos: Después de que todos los estudios que se deben incluir en la revisión han sido identificados, los revisores deben extraer de forma sistemática la información correspondiente de cada estudio.
7. Síntesis de los estudios: Este paso es conocido también como análisis; implica la combinación de los hechos extraídos de los estudios que utilicen técnicas apropiadas, ya sea cualitativa, cuantitativa o ambas.
8. Redacción de la revisión: El proceso de revisión sistemática de la literatura tiene q

ser reportado detalladamente para que los resultados de la revisión puedan reproducirse de forma independiente.

### 3. Resultados esperados

Uno de los objetivos de esta línea es poder mostrar que es posible realizar una revisión sistemática en el área. La revisión sistematizada se han realizado en áreas como la medicina por ejemplo. Existe evidencia que muestran que en área de la computación, esta metodología solo es aplicada al campo de la ingeniería del software. Se pretende mostrar que es posible aplicarla en esta línea de investigación.

Por otro lado, partiendo de los resultados obtenidos a partir de la revisión, se pretende desarrollar herramientas que faciliten el tema de la inclusión en el ámbito de la simplificación de texto para el idioma español. Actualmente, se está trabajando en la primer etapa de la revisión.

### 4. Formación de Recursos Humanos

Durante la realización de este sistema se espera lograr, como mínimo, la culminación de 2 tesis de grado dirigidas y/o co-dirigidas por los integrantes del proyecto y una tesis de maestría.

Finalmente, es constante la búsqueda de nuevos investigadores como también, la consolidación de investigadores dentro del grupo.

### Referencias

- [1] J. Burstein. Opportunities for natural language processing research in education. In *Proceedings of CICLing*, pages 6–27, 2009.
- [2] R. Chandrasekar, C. Doran, and B. Srinivas. Motivations and methods for text simplification. In *In Proceedings of the Sixteenth International Conference on Computational Linguistics (COLING '96*, pages 1041–1044, 1996.

- [3] B. Kitchenham. Kitchenham, 2004 procedures for performing systematic reviews, 2004.
- [4] B. Kitchenham and S. Charters. Guidelines for performing systematic literature reviews in software engineering no. ebse 2007-001., 2007.
- [5] C. E. Scarton and S. M. Aluísio. Análise da inteligibilidade de textos via ferramentas de processamento de língua natural: adaptando as métricas do Coh-Metrix para o português. *linguamatica*, 2(1):45–61, Abril 2010.
- [6] A. Siddharthan. Preserving discourse structure when simplifying text. In *In Proceedings of the 2003 European Natural Language Generation Workshop*, pages 103–110, 2003.
- [7] D. J. Young. Linguistic simplification of sl reading material: effective instructional practice. In *The Modern Language Journal* 83(3), pages 350–366, 1999.

## Agentes BDI con Función de Utilidad: Un enfoque hacia la opinión y confianza en la toma de decisiones.

Rodrigo René Cura, Romina Stickar, Carlos E. Buckle.

Depto. de Informática, Fac. De Ingeniería, Universidad Nacional de la Patagonia San Juan Bosco.  
Puerto Madryn, Argentina.  
+54 280-4472885 – Int. 117.  
rodrigo.renecura@gmail.com, romistickar@yahoo.com.ar, cbuckle@unpata.edu.ar

### Resumen

El presente trabajo se enfoca al desarrollo de un modelo general de decisión aplicable a agentes BDI (*Beliefs, Desires, Intentions*) que contempla aspectos sociales en la formación de una opinión mediante mecanismos de preferencias y comunicación. Para cuantificar el grado de preferencia se basa en el cálculo de la función de Utilidad. El modelo permite generar agentes prácticos para ser utilizados en simulación social. El trabajo, además de las metas de investigación y desarrollo experimental, persigue objetivos académicos de formación de recursos humanos y de vinculación con otras especialidades. Se desarrolla en el marco de una Tesina de Grado de Licenciatura en Informática.

**Palabras Clave:** Simulación Social Basada en Agentes, Toma de Decisiones, BDI, Utilidad.

### 1. Contexto

El presente trabajo se realiza en el marco de la tesina de grado de uno de sus participantes y como desarrollo inicial para el comienzo del proyecto de investigación “*Simulación Social para el apoyo a la toma de decisiones de gobierno.*” del grupo de estudio de Inteligencia Artificial, perteneciente a la Facultad de Ingeniería de la Universidad Nacional de la Patagonia San Juan Bosco (UNPSJB) Sede Puerto Madryn.

### 2. Introducción

La evaluación de escenarios hipotéticos y la predicción del impacto a futuro, en la introducción de un cambio dentro del entorno en que se plantea, puede alcanzar gran complejidad cuando se tiene la intención de considerar la mayor cantidad de aspectos que se verían alterados. Más aun cuando el ambiente es grande, las consecuencias y el costo alto, el cambio irreversible y los efectos secundarios poco claros, por ejemplo, una decisión política que afecte a los habitantes de una ciudad. Las ciencias de la información pueden colaborar con herramientas que ayuden al análisis de este tipo de problemas. La computación basada en agentes, la simulación y la ciencia social nos llevan a una intersección donde los tres paradigmas se combinan en un subconjunto llamado *Simulación Social Basada en Agentes* [5], en

esta perspectiva se enfoca este trabajo. Este tipo de simulaciones requiere la colaboración tanto de profesionales en las ciencias de la información como en las ciencias sociales, lo que conlleva a un conjunto de actividades que deben desarrollarse en la fase inicial del proyecto, por ejemplo, para un escenario urbano, es necesario el modelado de la ciudad, los transportes, las personas y su comportamiento, la explotación de información geográfica, los protocolos de comunicación entre las entidades del sistema, etc. Los agentes inteligentes resultan muy adecuados para este tipo de modelado porque ofrecen una forma familiar y comprensible para enfrentar sistemas complejos, siendo particularmente útil para el entendimiento entre especialistas en diferentes disciplinas.

Este proyecto se centra en el desarrollo de un modelo teórico que permita incorporar inteligencia para la toma de decisiones en los agentes, por ejemplo, en la elección de un medio de transporte, basada no solo en la experiencia sino también en la influencia de la opinión de terceros.

El modelo planteado combina la arquitectura BDI (*Beliefs, Desires, Intentions*) [2, 3] con la Teoría de Utilidad y el concepto de *preferencia* [1, 8] generando agentes con una potente capacidad de razonamiento.

La arquitectura BDI se emplea en el diseño de los agentes, para dotarlos de *Creencias, Deseos e Intenciones*, aptitudes que definen sus distintos estados

cognitivos (*mental state*). Las *creencias* son utilizadas para representar tres factores claves: las *opciones* disponibles, el *peso* o grado de relevancia de los factores a evaluar en una opción y la *confianza* hacia los demás agentes.

En la evaluación de las opciones disponibles es posible utilizar la *Teoría de Utilidad* y el concepto de *función de utilidad* que permite cuantificar el grado de preferencia de una opción desde la perspectiva de un agente. Después de haber tomado una línea de acción y que el agente haya percibido el *grado de desempeño* obtenido en cada factor a evaluar para la misma, hace empleo de la función de utilidad para combinarlos con su propio conjunto de pesos y conseguir, según su criterio, la utilidad real de desempeño. Para futuras referencias este valor de utilidad, obtenido de la ejecución de la acción se define como *Utilidad Percibida*.

Los agentes se relacionan mediante *actos comunicativos* [4, 6, 7] entre los ubicados en un área común. Un agente informa al resto cual es, conforme su juicio, la mejor *opción* y la *utilidad* asociada a ella. Cada agente almacena estas opiniones y, ponderadas por la confianza, las utiliza en el proceso de decisión. En este punto se introduce al modelo una nueva creencia, la *utilidad*, como una terna *opción, valor, fuente*, donde el valor es la utilidad de la opción informada por la fuente. El objetivo de este mecanismo es detectar los cambios en el desempeño de las opciones que el agente no cree útiles y por lo tanto no seleccionaría, a menos que alguien que la considere, lo informe. Se define esta utilidad almacenada con el nombre de *Utilidad Esperada*.

Por último debe tenerse en cuenta la actualización de los valores de confianza y utilidad esperada, para lo cual se definen *parámetros de sensibilidad*, encargados de regular la estabilidad o dinamismo en el cambio de los valores, evitando reacciones erráticas o demasiado rígidas.

A continuación se realiza la definición formal del modelo, se explica el algoritmo del proceso de razonamiento de los agentes y al final se explica un caso de estudio para la implementación y evaluación del modelo.

## 2.1 Definición Formal

Sea  $M = \{A, O, F, \alpha, \beta\}$  el modelo, con

$A = \{a_1, a_2, \dots, a_n\}$  un conjunto de agentes,

$O = \{o_1, o_2, \dots, o_m\}$  las opciones generales en el modelo y

$F = \{f_1, f_2, \dots, f_l\}$  los factores relevantes para la toma de decisiones.

$\alpha$  y  $\beta$  los parámetros de sensibilidad en confianza y utilidad respectivamente.

Se define  $\Psi_i$  como el estado cognitivo de  $a_i \in A$

con  $i = 1, 2, \dots, n$  siendo  $\Psi_i = \{S, W, T\}$  donde

$S \subseteq O$  las opciones disponibles para  $a_i$ ,

$W = [w_1 : f_1, w_2 : f_2, \dots, w_l : f_l]$  con cada  $f_j \in F, w_j \geq 0$  el grado de relevancia con respecto a cada uno de los factores,

$T = \{t_1, t_2, \dots, t_n\}$  la confianza relativa a cada agente  $a_j \in A$  con  $j = 1, 2, \dots, n$ .

Para cada  $o_i \in O$  con  $i = 1, 2, \dots, m$  se establece

$D = [d_1 : f_1, d_2 : f_2, \dots, d_l : f_l]$  con cada  $f_j \in F, d_j \geq 0$  el grado de desempeño relativo en cada uno de los factores.

Por último se define la función de utilidad

$$U: A, O \rightarrow \mathbf{R}$$

$$U(a_i, o_j) = \sum_{i=1}^k (w_i d_i) \text{ con } a_i \in A \text{ y } o_j \in O.$$

## 2.2 Algoritmo

1. Inicialización
2. Loop {
3.     Selección
4.     Acción
5.     Actualización
6.     Comunicación
7. }

### Inicialización

En este punto se establece el conjunto inicial de *creencias* del agente tal que:

- $w_i = G_i(x)$  para todo  $i = 1, 2, \dots, l; x \in [0, 1]$  aleatorio y  $G_i: [0, 1] \rightarrow [0, 1]$  la función de distribución de la relevancia de  $f_i$ .
- $S \subseteq O$  formado por todas las  $o_i \in S$  tal que  $P_i(x) = 1$ , siendo  $x \in [0, 1]$  aleatorio y  $P_i: [0, 1] \rightarrow 0$  si  $x < p_i$  y 1 si  $x \geq p_i$ , donde  $p_i$  es la probabilidad de existencia de  $o_i$ , para todo  $i = 1, 2, \dots, m$ .
- Para cada  $o_i \in S$  una  $U(a_i, o_i) = 0$ , siendo  $a$  el agente mismo.
- $T = \{t_a = 1\}$  siendo  $a$  el agente mismo.

### Selección:

En este punto, y luego de reiteradas iteraciones el agente posee en su base de creencias un conjunto de valores de utilidad asociados a opciones (ver paso de Comunicación más adelante). Llamaremos a estos valores *Utilidades Esperadas*, cada una indicada como  $U_E$ .

Para cada opción disponible  $o_i \in S$ , buscará en el conjunto de  $U_E$  el valor de utilidad informado por el agente  $a_i$  tal que el valor de confianza  $t_i \in T$  asociado a  $a_i$  sea máximo.



Una vez obtenido el subconjunto de  $S$ , seleccionará como la mejor opción aquella que maximice su valor de utilidad esperada  $U_E$ .

#### **Acción:**

El agente encarga la acción a realizar según la opción seleccionada al ambiente, el cual debe poseer mecanismos para informar, terminada la ejecución exitosa, la utilidad real calculada en base a los parámetros de desempeño de la opción  $d_i \in D$ . Llamaremos a esta utilidad calculada por el ambiente *Utilidad Percibida* y la denotaremos como  $U_P$ .

#### **Actualización:**

Realiza la comparación de la Utilidad Esperada y la Utilidad Percibida, y actualiza los valores de confianza del agente del cual se tomo la opción y Utilidad Esperada de la opción seleccionada a partir del propio agente, según los siguientes criterios:

- Si  $U_P \geq U_E \rightarrow t_s = \alpha t_s + (1 - \alpha)$  con  $\alpha \in [0,1]$  (significa que el agente es confiable)
- Si  $U_P < U_E \rightarrow t_s = \alpha t_s$  con  $\alpha \in [0,1]$  (el agente no es confiable)
- $U_E(a, o) = \beta U_E + (1 - \beta) U_P$  con  $\beta \in [0,1]$

Por este motivo a  $\alpha$  y  $\beta$  se los denomina parámetros de sensibilidad en la actualización de las creencias del agente, afectando a la confianza y utilidad esperada respectivamente.

#### **Comunicación:**

La comunicación es el paso donde los agentes intercambian opiniones acerca de las opciones que consideran más factibles, y donde se establecen los lazos de confianza iniciales.

Primero el agente necesita saber que agentes pueden escuchar su opinión. Esta tarea es delegada al ambiente, el cual devuelve una lista de agentes cercanos, una vez obtenida esta lista el agente emisor realiza un acto comunicativo con una fuerza ilocucionaria acertiva, que informa al grupo receptor una utilidad esperada asociada a una opción de la forma:

$$U(a, o) = U_E$$

donde  $a$  es el agente transmisor,  $o$  es la mejor opción según el criterio del agente transmisor y  $U_E$  es la utilidad esperada almacenada en la base de creencias de  $a$ .

## **2.3 Caso de Estudio**

El modelo anteriormente planteado será evaluado en un caso de estudio que pretende simular el movimiento de personas, mediante un grupo de distintos medios de transporte disponibles, a través de las calles de una ciudad y poder observar patrones en la opinión general de los ciudadanos.

### **Ambiente**

El ambiente sobre el que interactuarán los ciudadanos es una representación simplificada de un entorno urbano interconectado por calles, el cual implementa la acción:

$viajar(a_i, o_j) \rightarrow U_P$  donde  $a_i \in A$ ,  $o_j \in S$  y  $U_P$  es la utilidad percibida por  $a_i$  al finalizar exitosamente la acción.

### **Agentes**

Los agentes poseen una base de creencias formada por:

- *opción( $o_i$ )*. Para cada  $o_i \in S$ .
- *factor( $w_i, f_i$ )*. Para cada  $w_i \in W$  y  $f_i \in F$ .
- *confianza( $t_i, a_i$ )*. Para cada  $t_i \in T$  y  $a_i \in A$  que haya comunicado su opinión.
- *utilidad( $U_E, o_i$ ) $[a_i]$* . Para cada  $o_i \in S$  y  $a_i \in A$  que haya comunicado su opinión.

### **Opciones**

Las opciones  $o_i$  están implementadas dentro del ambiente y cada una posee:

- *desempeño( $d_i, f_i$ )*. Para cada  $d_i \in D$  y  $f_i \in F$ .
- $U: A \rightarrow [0,1]$  que realiza el calculo de utilidad de la elección de la opción.

### **Inicialización**

La inicialización de los agentes de realizará aleatoriamente en base a distribuciones estadísticas establecidas por el usuario. Los factores se definirán por completo, las opciones en base a una probabilidad de existencia, la utilidad esperada se iniciará en 0 para cada una de las opciones disponibles y por último la confianza será 1 solo para si mismo. Por último se establecerá un conjunto de número variable de destinos que visitar, a los cuales el agente se dirigirá cíclicamente.

En cuanto a los parámetros de desempeño y sensibilidad de actualización, estos serán fijados por el usuario para darle la posibilidad de aplicar técnicas de análisis a la simulación.

### **Observador**

El sistema contará con un agente observador que se encargará de acumular y llevar registro de los indicadores para la evaluación del sistema. Los indicadores claves serán, para cada opción, la cantidad de agentes que la han seleccionado y el patrón con el que dicha selección se ha expandido por el ambiente. Otros indicadores a tenerse en cuenta son: La media y desviación estándar de la utilidad esperada para cada opción  $U_E(o_i)$ , la media y desviación estándar de los valores de confianza  $t_i$ .

### 3. Líneas de Investigación y Desarrollo

Esta línea de investigación está orientada a desarrollar un modelo que permita reflejar el impacto de la opinión popular en las creencias de un agente, influyendo indirectamente en las decisiones del mismo. Esto proporcionará agentes con un comportamiento, que al ser utilizados en la simulación social, generarán información útil para el apoyo a las decisiones políticas.

### 4. Resultados y Objetivos

El proyecto se encuentra en un estado de avance en el que ya se ha implementado el modelo mental completo del agente ciudadano y gran parte del ambiente que se encarga de la acción de viajar, si bien el sistema parece responder a simple vista aún no se han obtenido estadísticos precisos pues aún queda pendiente la implementación del agente observador. Los pasos a seguir son la implementación de dicho agente, verificación, recolección de datos, análisis de sensibilidad y experimentación.

Los objetivos propios de la investigación y desarrollo de este trabajo se pueden resumir en:

- Investigar la aplicabilidad de agentes inteligentes a la simulación social.
- Determinar la factibilidad y aplicabilidad de la simulación social basada en agentes en la obtención de información práctica para la predicción del impacto de cambios en un entorno urbano.
- Diseñar un modelo genérico que pueda utilizarse como mecanismo de decisión social en agentes BDI.
- Proponer mejoras a los mecanismos de estimación del comportamiento social ante un cambio del entorno.

Por otra parte, los objetivos académicos son:

- Iniciar un grupo de estudio en el área de

Computación Basada en Agentes.

- Desarrollar el trabajo de tesina de grado de un alumno de Licenciatura en Informática.
- Trabajar en forma conjunta con especialistas de otras ciencias como por ejemplo las Ciencias Sociales.

### 5. Formación de Recursos Humanos

Este trabajo se lleva a cabo en el contexto de la realización de la tesina de grado de uno de sus autores, en la carrera Licenciatura en Informática. Por otra parte, pretende iniciar formación en una rama de las ciencias de la computación con gran potencial para investigación interdisciplinaria, particularmente en lo relacionado al comportamiento social.

### 6. Referencias

- [1] von Neumann, J. and Morgenstern, O. (1944). *Theory of games and economic behavior*.
- [2] Bratman M. E. (1987). *Intentions, Plans and Practical Reason*.
- [3] Rao A. S. and Georgeff M. P. (1991). *Modeling Rational Agents within a BDI-Architecture*.
- [4] Wooldridge, M. (2002). *An introduction to Multiagents Systems*.
- [5] Davidsson P. (2002). *Agent Based Social Simulation: A Computer Science View - Journal of Artificial Societies and Social Simulation vol. 5, no. 1*.
- [6] ANA MAS (2005). *Agentes de Software y Sistemas Multi-Agente: Conceptos, Arquitecturas y Aplicaciones*.
- [7] Bordini R. H., Hüber J. M., Wooldridge, M. (2007). *Programming multiagents systems in AgentSpeak using Jason*.
- [8] Shoham, Y. and Leyton-Brown, K. (2009). *MULTIAGENT SYSTEMS: Algorithmic, Game-Theoretic, and Logical Foundations*.

# Desarrollo de API de consulta a fuentes de información en la web para Sistemas de Argumentación Rebatible

**Cristian D. Pacifico, Luis I. Pelayo Novello, Leandro J. Tisocco Pertus**

Facultad de Ciencias de la Administración - Universidad Nacional de Entre Ríos  
Av. Tavella 1425, (E3202KAC) Concordia - ER, Argentina - Tel: (0345) 423-1400  
e-mail: {cripac, inovello}@fcad.uner.edu.ar, ltisocco@gmail.com

**Alejandro J. Garcia, Guillermo R. Simari**

Laboratorio de Investigación y Desarrollo en Inteligencia Artificial  
Departamento de Ciencias e Ingeniería de la Computación - Universidad Nacional del Sur  
Av. Alem 1253, (B8000CPB) Bahía Blanca - BA, Argentina - Tel: (0291) 459-5135  
e-mail: {ajg, grs}@cs.uns.edu.ar

## Resumen

Es cada vez más necesario automatizar los procesos que integran la información que se encuentra disponible en diferentes formatos, descentralizada, con orígenes heterogéneos y administraciones diferentes. Uno de los grandes desafíos al integrar fuentes de datos heterogénea, es poder manejar la inconsistencia informacional y la posible incompletitud que puede trasladarse de las fuentes locales. Los formalismos de Argumentación son adecuados para manejar este tipo de problemas. En particular, pensar en formalismos de *argumentación rebatible* (como DeLP) para la definición y automatización de la integración de fuentes de datos heterogéneas, inconsistentes entre sí e incompletas, es una opción aceptable.

El trabajo de esta línea de investigación propone desarrollar interfaces para vincular DeLP con fuentes de datos en formatos web. En especial, fuentes de sitios web descriptas en HTML y en Servicios Web. La información obtenidas de estas fuentes, podrán ser utilizadas como *hechos* o *presunciones* en la base de conocimiento que utiliza el sistema argumentativo.

**Palabras clave:** Argumentación, Integración de Información, Defeasible Logic Programming, World Wide Web, Web Services

## 1. Contexto

Esta línea de investigación se llevará a cabo dentro del proyecto de I+D:

**PID-UNER 7041** “*Desarrollo de Sistemas de Argumentación Masiva sobre Bases de Datos Federadas*”

Este proyecto está encuadrado en el convenio de colaboración entre el Laboratorio de Investigación y Desarrollo en Inteligencia Artificial (LIDIA) del Departamento de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur y el Área de Agentes y Sistemas Inteligentes de la Facultad de Ciencias de la Administración, Universidad Nacional de Entre Ríos.

## 2. Introducción

Es cada vez más necesario automatizar los procesos que integran la información que se encuentra disponible en diferentes formatos, descentralizada, con orígenes heterogéneos y administraciones diferentes. Desde hace décadas existen enfoques tradicionales para integrar bases de datos relacionales para conformar bases federadas. Estos enfoques proponen desde integraciones ad-hoc para constituir almacenes particulares [11, 10]; hasta la utilización de formalismos lógicos [12] y sistemas basados en marcos

[19] para describir, en forma abstracta, el marco de integración. Teniendo presente el acceso a Internet, también se crearon mecanismos para integrar información en los formatos que le son propios; por ejemplo de documentos [2], de servicios web [14], de descripciones ontologías en formalismos de web semántica [15, 13], etc.

Un gran desafío al integrar diversas fuentes de datos, es manejar eficientemente la inconsistencia informacional y la posible incompletitud que puede trasladarse de cada fuente local. Ya se sabe que los formalismos de Argumentación son adecuados para manejar estas problemáticas. En particular, pensar en formalismos de *argumentación rebatible* para la definición y automatización de la integración de fuentes de datos heterogéneas, inconsistentes entre sí e incompletas, es una opción aceptable.

La *argumentación rebatible* (*Defeasible Argumentation*) [17] es un mecanismo de razonamiento no-monótono en donde la aceptación o el rechazo de una proposición dependen de un análisis entre argumentos a favor y en contra de esa proposición [5, 16]. Usualmente es utilizada bajo diferentes formalizaciones para capturar aspectos del razonamiento del sentido común y la representación de información incompleta y potencialmente inconsistente [1, 3, 18].

La investigación propuesta aquí, se basa en la *Programación Lógica Rebatible (DeLP)* [9] siendo este es un formalismo que combina programación lógica y argumentación rebatible. Como rasgo particular, DeLP utiliza argumentación para decidir entre información contradictoria a través de un análisis dialéctico, lo que justifica la conclusiones obtenidas.

### 3. Línea de investigación

Esta investigación se vincula con el desarrollo de razonadores de argumentación rebatible para lograr integración de información masiva, cuyo marco teórico fue ya establecido; se ha presentado un framework (DBI-DeLP) en [6] que integra DeLP con bases de datos para lograr la ejecución de procesos de argumentación sobre fuentes de datos masivas externas y actualizables [7]. En concreto y sobre este framework se han desarrollado sistemas recomendadores [4].

El trabajo de esta línea de investigación propone *desarrollar interfaces para vincular DeLP*

*con fuentes de datos en formatos web*. En especial, fuentes residentes en sitios web descriptas en HTML y en Servicios Web soportados con SOAP y RESTFull. La información obtenidas de estas fuentes, podrán ser utilizadas como *hechos* o *presunciones* [8, 16] en la base de conocimiento que utiliza el sistema argumentativo.

Para llevar a cabo este objetivo, se propone desarrollar una interfaz de programación (API). Se opta por lenguajes de Programación Lógica para desarrollo de las APIs, con el ánimo de que sean integrables con DeLP. La API contendrán una estructura común que se subdivide en tres módulos:

1. Módulo **main**: contendrán los predicados/funciones que permiten recuperar, actualizar y consultar información de fuentes en la web. Estos predicados/funciones deben ser abstractos al formato de información y a los protocolos de transporte de las fuentes.
2. Módulo **sources**: contendrá los predicados/funciones que permiten definir y establecer la conexión particular con cada fuente. Para el caso de páginas web, se asocia a cada URI parametrizable, la manera de recuperación y filtrado de documentos HTML para la obtención de un dato en particular. Para el caso de servicios web, se define la estructura de la petición SOAP o RESTFull, y el filtrado de la correspondiente respuesta. Las relaciones definidas en este módulo deben configurarse para cada fuente en particular, pudiendo coexistir varias configuraciones.
3. Módulo **relations**: contendrá los predicados/funciones que vinculan relaciones semánticas con operaciones de recuperación y consulta en las fuentes. Es decir, definir la obtención de presunciones estableciendo las relaciones entre los predicados del módulo **source** para lograrlas, utilizando a éstos como funciones primitivas. Este módulo funciona como controlador de los dos anteriores.

### 4. Resultados y Objetivos

Actualmente, se ha implementado parte de la estructura de la API en SWI-Prolog utilizan-



do sus APIs: `http_client`, `http_sgml_plugin` y `xpath`. Como caso de testeo se han desarrollado configuraciones para tomar información de sitios de datos de películas ([www.imdb.com](http://www.imdb.com), [www.decine21.com](http://www.decine21.com), [www.dvdempire.com](http://www.dvdempire.com) y otras). Se han definido relaciones para integrar la información de directores, títulos, actores y ratings de una determinada película.

Se han definido predicados para realizar búsquedas en el sitio y obtener conjuntos resultados, por nombre de actor, director, o título. De igual forma, se desarrollo una configuración para trabajar sobre servicios web, para tratar los mismos datos.

El trabajo siguiente propone como desafío mejorar la performanse de la API, con técnicas de cacheo. Contextualizar la API en un ejemplo de argumentación real, con vinculación a DeLP. Por último, describir una metodología para describir fuentes de datos, establecer relaciones y programas DeLP en base a esta API.

## 5. Formación de Recursos Humanos

Esta investigación se lleva a cabo dentro del proyecto de I+D **PID-UNER 7041** “Desarrollo de Sistemas de Argumentación Masiva sobre Bases de Datos Federadas”.

Ha servido como tema de plan de tesis para 2 (dos) becas de Formación de Recursos Humanos UNER destinadas a alumnos de grado.

Está previsto que el trabajo continúe siendo la base para el plan de 1 (una) de Iniciación a la Investigación en Proyectos UNER.

La fase de desarrollo de la API, sirve como disparador para conformar temas de Tesis de grado en la Licenciatura en Sistemas de la FCAD-UNER.

## Referencias

- [1] Bench-Capon, T.J.M. y P.E. Dunne: *Argumentation in artificial intelligence*. Artificial Intelligence, 171(10-15), 2007. Argumentation in Artificial Intelligence.
- [2] Bergamaschi, S., S.Castano y M.Vincini: *Semantic Integration of Semistructured and Structured Data Sources*. SIGMOD Record, 28(1), 1999.
- [3] Besnard, Philippe y Anthony Hunter: *Elements of Argumentation*. MIT Press, 2008.
- [4] Briguez, C.E., M.C. Budán, C.A.D. Deagustini, A.G. Maguitman, M. Capobianco y G.R. Simari: *Towards an Argument-based Music Recommender System*. Volumen 245 de *Frontiers in Artificial Intelligence and Applications*. IOS Press, 2012.
- [5] Chesñevar, C.I., A.G. Maguitman y R.P. Loui: *Logical Models of Argument*. ACM Computing Surveys, 32:337–383, 2000.
- [6] Deagustini, C.A.D., S.E. Fulladoza Dalibón, S. Gottifredi y G.R. Simari: *DBI-DeLP: a framework for defeasible argumentation over databases*. XII Workshop Agentes y Sistemas Inteligentes, 2011.
- [7] Deagustini, C.A.D., S.E. Fulladoza Dalibón, S. Gottifredi, M.A. Falappa y G.R. Simari: *Consistent Query Answering Using Relational Databases through Argumentation*. En *Database and Expert Systems Applications*, volumen 7447, páginas 1–15. Springer Berlin Heidelberg, 2012.
- [8] Garcia, A., C.Chesñevar y G.Simari: *An argumentative framework for reasoning with inconsistent and incomplete information*. Workshop on practical reasoning and rationality. 13th biennial European Conference on Artificial Intelligence, 1998.
- [9] García, A.J. y G.R. Simari: *Defeasible logic programming: an argumentative approach*. Theory Pract. Log. Program., 4:95–138, 2004.
- [10] Heimbigner, D. y D. McLeod: *A federated architecture for information management*. ACM Trans. Inf. Syst., 3(3), 1985.
- [11] Inmon, W. H.: *Building the data warehouse (2nd ed.)*. John Wiley & Sons, Inc., 1996.
- [12] Lakshmanan, L. y N. Shiri: *A Parametric Approach to Deductive Databases with Uncertainty*. IEEE Trans. on Knowl. and Data Eng., 13(4), 2001.
- [13] McGuinness, D. L.: *Ontologies Come of Age*. En *Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential*. MIT Press, 2002.

- [14] Mcilraith, S.A., T.C. Son y Honglei Zeng: *Semantic Web services*. Intelligent Systems, IEEE, 16, 2001.
- [15] Meersman, R.: *Semantic Ontology Tools in IS Design*. En *11th International Symposium on Foundations of Intelligent Systems*. Springer-Verlag, 1999.
- [16] Nute, Donald: *Defeasible Logic*. En *Handbook of Logic in AI and Logic Programming*. Oxford University Press, 2001.
- [17] Pollock, John L.: *Defeasible reasoning*. Cognitive Science, 11:481-518, 1987.
- [18] Rahwan, Iyad y Guillermo R. Simari: *Argumentation in Artificial Intelligence*. Springer, 2009.
- [19] S.Heymans et al.: *Ontology Reasoning with Large Data Repositories*. En *Ontology Management*, volumen 7, páginas 89-128. Springer US, 2008.

# ESTUDIO COMPARATIVO DE LA EFICIENCIA ENTRE CONTROLADORES DIFUSOS DEL TIPO MANDANI Y SUGENO “UN CASO DE ESTUDIO EN LA NAVEGACIÓN AUTÓNOMA DE ROBOT”

Kornuta, Cristian

Marinelli, Marcelo

Departamento de Informática, Facultad de Ciencias Exactas Químicas y Naturales.  
Universidad Nacional de Misiones

Felix de Azara 1552, Posadas, Misiones

Te: 376-422186

cristian.kornuta@yahoo.com.ar, marcelomarinelli@gmail.com

## Resumen

En los sistemas de navegación autónoma de robot móviles, se presenta el problema de evadir obstáculos fijos utilizando alguna técnica perteneciente a la inteligencia artificial para determinar su trayectoria. En el presente trabajo se rediseño el controlador difuso para la navegación de un robot móvil desarrollado en [Marinelli Marcelo, 2009] “Diseño De Un Controlador Difuso Para Un Sistema De Navegación De Robot Con Tracción Diferencial”. Se presenta un nuevo modelo utilizando la herramienta FIS de MATLAB, posteriormente se desarrolló un módulo a partir de este modelo diseñado para la realización de simulaciones sobre un robot con el objetivo de evaluar su desempeño en entornos reales con obstáculos.

## Palabras clave:

Navegación; Vehículos autónomos; Control predictivo; Robots móviles; Control difuso.

## Contexto

Este proyecto se enmarca en el “Programa de Investigación en Computación” del Centro de Investigación y Desarrollo Tecnológico de la Facultad de Ciencias Exactas Químicas y Naturales de la Universidad Nacional de Misiones; también vinculado con el Doctorado en Ingeniería de Sistemas y Computación que funciona en la Universidad de Málaga.

Dentro del proyecto se desempeñan docentes, tesis, doctorando y becarios de las carreras de

Analista en Sistemas de Computación, Licenciatura en sistemas de Información y Profesorado en Física

## Introducción

Los robots autónomos son entidades físicas programables con capacidad de percepción sobre su entorno y de actuar sobre el mismo en base a dichas percepciones, sin necesidad de supervisión o intervención humana, por tal motivo la navegación en un robot autónomo es un tema crucial.

La navegación eficiente y eficaz de un robot en un entorno dinámico o desconocido, dependerá de su capacidad de decisión con respecto a sus acciones frente a los obstáculos desconocidos. En la implementación de una solución a estos problemas es deseable en muchos casos utilizar alguna técnica que se aproxime al razonamiento humano dado la capacidad que presenta el ser humano frente a estas mismas situaciones. La lógica difusa proporciona un medio para capturar la capacidad de la mente humana, esta técnica frecuentemente es utilizada en problemas que presentan estas características de incertidumbre [5] [6].

En la navegación autónoma de un robot, la capacidad para evitar la colisión con obstáculos imprevistos o el seguimiento de una ruta marcada, son tareas vitales para un robot autónomo; problemas que son estudiados frecuentemente. Debido a que al momento de ser implementado estas capacidades, no existen en el mundo real, mapas fiables en los que se

encuentren marcados los obstáculos que pudiera encontrarse el robot al desempeñarse en el entorno, debido a la naturaleza cambiante de los entornos, conjuntamente hay que considerando que los datos capturados por los sensores en ocasiones no son del todo fiables por diferentes causas. A estos problemas, la lógica difusa nos proporciona una solución robusta y fiable que contempla estas situaciones y la posible entrada de datos imprecisos. En el presente trabajo presentamos una solución eficiente a este problema aun vigente.

La teoría de la lógica difusa propuesta por Lotfi A. Zadeh en 1965 que da origen a la teoría de conjuntos difusos; fue creada para emular la lógica humana y poder tomar decisiones acertadas a partir información imprecisa, lo cual se aproxima a la forma que el cerebro humano piensa o razona[7].

La Lógica difusa presenta un conjunto de principios matemáticos basados en grados de membresía o pertenencia que permite modelar la información.

## Objetivos

El objetivo de este trabajo es el desarrollar un controlador difuso para la navegación de robots móviles utilizando las herramientas de la lógica difusa de MATLAB. También es realizar las simulaciones en el FIS de MATLAB y una vez perfeccionada la base de conocimientos y las reglas difusas, diseñar un módulo para realizar experiencias en la modalidad de evasión de obstáculos con un robot en tiempo real.

## Metodología

La metodología utilizada en el presente trabajo consistió en una primera instancia de un relevamiento profundo y amplio sobre documentos, bibliografías, papers, revistas especializadas, tesis, foros y páginas en general sobre el tema, en búsqueda de los temas que guiarán nuestra investigación y desarrollo con el fin de poder estudiar y analizar el estado actual del tema. Luego del análisis de la información recopilada, se estudió a partir de la información recopilada la metodología de construcción de controlador difuso que nos guiaría en el

rediseño, el cual fue los lineamientos propuesto por Cox en su libro [5]; a partir de los lineamientos propuestos por el autor se prosiguió a analizar y rediseñar el controlador difuso del tipo Mandani propuesto en [Marinelli Marcelo, 2009] en forma de bosquejos para luego implementar el nuevo modelo en la herramienta FIS de MATLAB. Una vez perfeccionada la base de conocimientos y las reglas difusas para lograr un mejor comportamiento y respuesta reactiva a partir del análisis de diferentes situaciones planteadas; se diseñó un nuevo controlador difuso a partir de la base de conocimientos y las reglas difusas del controlador difuso Mandani diseñado anteriormente pero con la particularidad que este fuese del tipo Sugeno para poder comparar su comportamiento y respuesta reactiva entre los dos.

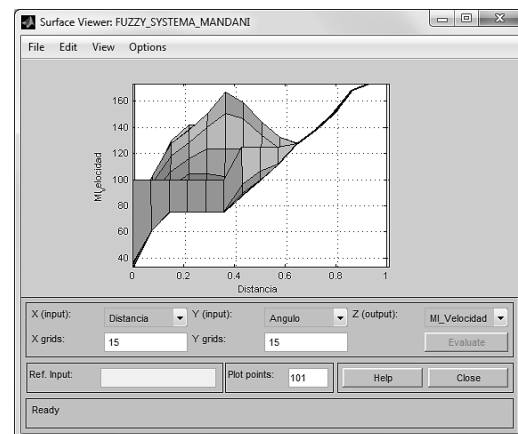


Figura 1. Simulación del modelo Mandani

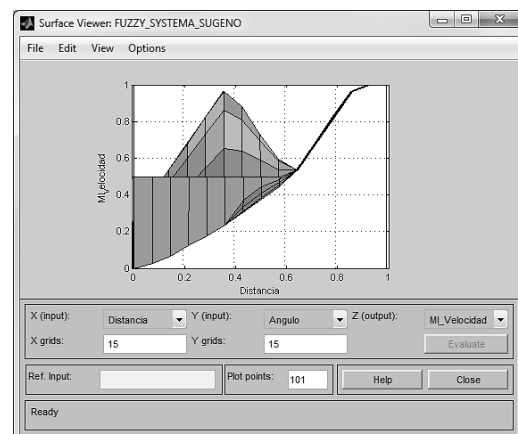


Figura 2. Simulación del modelo Sugeno

Posteriormente de las realizaciones de pruebas conjuntas con los dos controladores difusos y haber estudiado sus comportamientos a



partir del ajuste de las base de conocimiento se y de comprobar que el modelo Sugeno presenta una mejor respuesta; se decidió codificar los dos controladores difusos por separado, en el lenguaje MATLAB para poder obtener una aproximación real en código del modelo implementado en la herramienta, en esta instancia se siguió una metodología de software con un enfoque evolutivo, en una segunda instancia se exporto este mismo el código de los controladores difusos al lenguaje Arduino para ser utilizado con el robot en las pruebas reales con la modalidad evasión de obstáculos.



**Figura 3.** Robot N6 con las unidades sensoras

Una vez tenido el código en el robot se diseñó las experiencias que posibiliten evaluar el modelo en diferentes situaciones con el fin de ajustar el modelo.

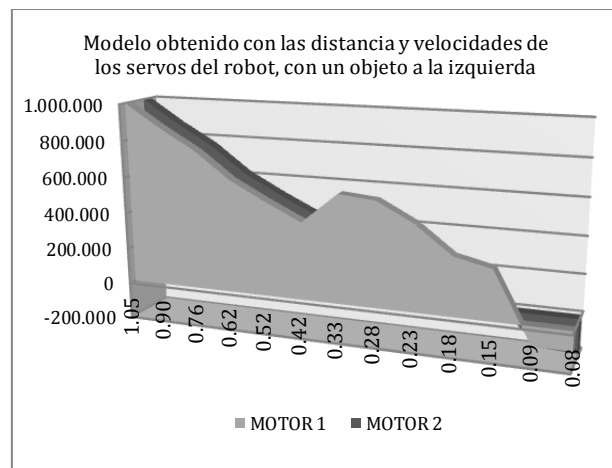
En última instancia se realizaron las experiencias y toma de datos.

### Resultados obtenidos

A pesar de haber realizado las simulaciones de los dos controladores difusos diseñados sobre la herramienta y de concluir que el modelo Sugeno presenta una mejor respuesta, se implementó los dos controladores difusos sobre la placa DuinoBot presente el robot Multiple N6 junto a él se utilizaron conjuntamente los sensores presente en el robot; lo que demostró esta experiencia es que el controlador difuso del tipo Sugeno se comporta de forma más eficientemente computacionalmente, que el modelo Mandani. Además al momento de ser implementado el modelo Mandani sobre la placa

se nos presentó el problema del tamaño del programa para guardarlo en la memoria de la tarjeta, esto nos perjudico al momento de tratar de depurar, algo que con el modelo Sugeno no tuvimos este problema, al no tener un método de deborización con lo cual se ahorra muchas líneas de código.

En resumen debido a que es compacto y computacionalmente más eficiente que su representación del modelo Mandani y principalmente se comporta mucho mejor el modelo Sugeno para este caso de estudio, se concluye que el modelo Sugeno se presta más para este caso de estudio por utilizar técnicas lineales lo que garantiza la continuidad de la superficie de salida.



De las pruebas realizadas podemos concluir que el robot se comporta de una forma muy aceptable, con un grado de aleatoriedad entre prueba y prueba con respecto a su trayectoria normal, producido por las variables de giro intervinientes, en el momento de realizar los giros del robot que se encuentra condicionadas por la distancia de los objetos y el ángulo en ese momento determinado; se produjeron errores de lectura por parte de los sensores que causaron choques imprevistos que fueron causas por la velocidad del robot y el tiempo de lectura los cuales son aceptables, problemas que estamos corrigiendo.

A partir de los resultados obtenidos de las diversas pruebas realizadas, pudimos concluir que el controlador difuso del tipo Sugeno

presenta una muy buena eficiencia y respuesta reactiva en la evasión de obstáculos en 98 %.

## Líneas de investigación y desarrollo

Dentro de las líneas de investigación que contienen el “Programa de Investigación en Computación”, este proyecto se enmarca en el área de inteligencia artificial en donde se aplican técnicas de lógica difusa, redes neuronales, algoritmos bioinspirados, sistemas expertos, etc., estas técnicas se utilizarán en los siguientes trabajos:

- Controladores difusos aplicados al proceso de la elaboración de yerba mate.
- Redes neuronales aplicados a la navegación autónoma de robot.
- Sistemas de control de navegación para robots utilizando arreglos de sensores de ultra sonido.
- Controladores difusos aplicados a la navegación autónoma de robot.
- Telemetría con tecnología bluetooth para el control de navegación de robot.
- Desarrollo de aplicaciones con Java móvil y tecnología bluetooth, con el objeto de usar los dispositivos móviles para el control de robots domésticos.

## Formación de Recursos Humanos

En esta línea de investigación se prevé desarrollar cuatro Tesis de grado de la carrera de Licenciatura en Sistemas de Información y dos Tesis doctorales del Doctorado en Ingeniería de Sistemas y Computación del Departamento de Lenguajes y Ciencias de la Computación de la Universidad de Málaga.

## Referencias

[1]. Marinelli Marcelo J.; Kuna Horacio D.; Korol Fernando J.; Plenc Hugo A. Diseño De Un Controlador Difuso Para Un Sistema De Navegación De Robot Con Tracción Diferencial - II Jornadas

Científico-Tecnológicas de la FCEQyN , 2009, posadas.

- [2]. Marinelli, M. (2006). Evaluación de la Efectividad para Detectar Obstáculos y Representarlos en V.R.M.L. de un Sistema de Navegación para Robot mediante Sonar. U.Na.M., Tesis de Maestría.
- [3]. Ahmad M. Ibrahim (2004). FUZZY LOGIC for Embedded Systems Applications. Estados Unidos: Elsevier Science.
- [4]. Hao Ying (2000). FUZZY CONTROL AND MODELING. Estados Unidos: IEEE Press.
- [5]. Cox E. (1994). The Fuzzy Systems Handbook - A Practitioner's Guide to Building, Using, and Maintaining Fuzzy Systems, ACADEMIC PRESS.
- [6]. Elmer P. Dadios (2012). Fuzzy Logic – Controls, Concepts, Theories and Applications. Croatia: InTech.
- [7]. Ponce Cruz, Pedro (2010). Inteligencia artificial, con aplicaciones a la ingeniería. Mexico: Alfaomega.
- [8]. Jan Jantzen (2007). Foundations of Fuzzy Control. Estados Unidos: John Wiley & Sons.
- [9]. Sohail Iqbal (2012). Fuzzy Controllers – Recent Advances in Theory and Applications. Croatia: Janeza Trdine.
- [10]. Chennakesava R. Alavala (2008). Fuzzy Logic and Neural Networks, Basic Concepts and Applications. Estados Unidos: New Age International Pvt.
- [11]. García Martínez (1997). Sistemas Autónomos. Aprendizaje Automático. Argentina: Nueva Librería.

## Técnicas de Soft Computing aplicadas a Biometría, Predicción y Ruteo de Vehículos

Laura Lanzarini<sup>1</sup>, Franco Ronchetti<sup>2</sup>, César Estrebou<sup>3</sup>, Juan Maulini<sup>4</sup>

Javier López<sup>5</sup>, Juan Martín Iglesias Caride<sup>6</sup>, Luciana Lens<sup>7</sup>

Instituto de Investigación en Informática LIDI (III-LIDI)<sup>8</sup> -Facultad de Informática. UNLP

{laural, fronchetti, [cesarest](mailto:cesarest@lidi.info.unlp.edu.ar), [jmaulini](mailto:jmaulini@lidi.info.unlp.edu.ar)}@lidi.info.unlp.edu.ar,

jlopez@telefe.com.ar, jmartinigulesias@yahoo.com.ar, lucianasollens@gmail.com

Aurelio Fernández Bariviera<sup>9</sup>

Depto.de Gestión de Empresas, Universitat Rovira i Virgili, Reus, España

aurelio.fernandez@urv.net

### Contexto

Esta presentación corresponde al Subproyecto “Sistemas Inteligentes” perteneciente al Proyecto “Procesamiento paralelo y distribuido. Fundamentos y aplicaciones en Sistemas Inteligentes y Tratamiento de imágenes y video” acreditado por la UNLP.

### Resumen

Esta línea de investigación se centra en el estudio y desarrollo de técnicas de *Soft Computing*. El énfasis está puesto en Redes Neuronales, Lógica Difusa y Optimización por cúmulo de partículas aplicadas al procesamiento de señales (audio e imágenes) para efectuar reconocimiento biométrico, predicción de índices de bolsa y toma de decisiones en problemas del tipo VRP (Vehicle Routing Problem).

En lo referido a biometría se ha desarrollado un nuevo reconocedor basado en una red neuronal competitiva difusa y una red neuronal feedforward que no sólo mejora la tasa de acierto

con respecto a soluciones previas sino que además, incorpora el concepto de clase de rechazo.

En el área de predicción se han utilizado redes neuronales feedforward para operar sobre series temporales económicas. Se han obtenido resultados levemente mejores que la conocida estrategia de comprar y esperar (*buy-and-hold*). También se han comparado los distintos niveles de memoria a largo plazo que introduce el índice de Hurst con la respuesta obtenida a partir de diferentes redes neuronales *feedforward*.

En lo que respecta a la toma de decisiones en problemas del tipo VRP se ha completado la presentación de una tesis de Doctorado en Ciencias Informáticas la cual incluye el diseño e implementación de varias alternativas del método PSO que fueron aplicadas en la optimización de funciones complejas así como al ruteo de vehículos para mejorar el servicio en una empresa de emergencias médicas.

**Palabras claves:** Redes Neuronales, Lógica Difusa, Optimización mediante Cúmulos de Partículas.

<sup>1</sup> Profesor Titular. Facultad de Informática. UNLP

<sup>2</sup> Becario de Postgrado UNLP. Ayudante Alumno. Facultad de Informática.

<sup>3</sup> Jefe de Trabajos Prácticos SD. Facultad de Informática. UNLP

<sup>4</sup> Ayudante Diplomado. Facultad de Informática. UNLP

<sup>5</sup> Mg.en Administración de Empresas. Tesista del Doctorado en Cs.Informáticas. Fac.Informática UNLP

<sup>6</sup> Lic.en Administración de Empresas. Tesista de la Maestría en Explotación de Datos y Descubrimiento del Conocimiento. Fac.de Ingeniería. UBA.

<sup>7</sup> Tesista de la Lic. en Sistemas. Fac.de Ingeniería. UNPSJB.

<sup>8</sup> Calle 50 y 120 2do Piso, (1900) La Plata, Argentina, TE/Fax +(54) (221) 4227707. <http://weblidi.info.unlp.edu.ar>

<sup>9</sup> Profesor Titular. Depto de Gestión de empresas. Universitat Rovira i Virgili

## 1. Introducción

Las técnicas de *Soft Computing* han sido utilizadas exitosamente para resolver numerosos problemas que manejan información incierta, incompleta e imprecisa.

En el Instituto de Investigación en Informática LIDI se está trabajando en este tema desde hace varios años. Inicialmente se desarrollaron estrategias basadas en Redes Neuronales y Técnicas de Optimización. En especial, se han propuesto nuevos métodos para determinar la arquitectura y entrenar Redes Neuronales competitivas. La incorporación de la Lógica Difusa a estos modelos ha permitido mejorar su desempeño.

A continuación se detallan en forma breve los avances realizados últimamente.

### 1.1. Reconocimiento de Patrones Biométricos

Dentro de este tópico se viene trabajando en el III-LIDI sobre dos líneas de investigación que incluyen el procesamiento y reconocimiento de señales de audio e imagen

En cuanto a la primera línea de investigación, se ha trabajado sobre el problema de identificación de locutor. El tratamiento digital de señales de audio es una de las técnicas biométricas que puede utilizarse para realizar las tareas de identificación y verificación. El reconocimiento de voz es una herramienta sumamente útil en el área de seguridad ya que permite validar que la persona es quien dice ser o bien identificar por algún motivo específico que persona es la que está hablando.

Se han desarrollado varios métodos para lograr un modelado que resuelva el problema de identificación. En esta línea de investigación se ha propuesto y desarrollado un algoritmo de tres etapas, llamado ProbSOM [1], que genera un único modelo para identificar varias personas por medio del análisis de voz con mejores resultados que los alcanzados en [2].

Por otro lado, desde 2008 se ha trabajado en el reconocimiento de rostros analizando si la imagen del rostro de una persona se corresponde o no con alguna de las imágenes existentes en una base de datos. Este problema es difícil de resolver automáticamente debido a los cambios que distintos factores, como la expresión facial, el envejecimiento e incluso la iluminación, produ-

cen en la imagen. Para resolverlo se han utilizado los descriptores SIFT definidos en [3] y seleccionado utilizando una técnica de optimización. Los resultados obtenidos fueron publicados en [4] y [5].

Actualmente se está desarrollando un reconocedor biométrico general que puede utilizarse tanto para voz como para imágenes. Esto se debe a que la caracterización de la señal de entrada la convierte en un conjunto de vectores que luego pueden ser procesados directamente. El nuevo reconocedor está basado en la combinación de una red SOM difusa con un multiperceptrón. El reconocimiento tiene dos partes: la red SOM difusa identifica la persona de la base de datos que posee la señal más parecida a la entrada y el multiperceptrón decide si se trata efectivamente de esa persona. De esta forma, eventualmente el reconocedor podrá indicar si la señal de entrada corresponde a una persona desconocida o no.

### 1.2. Predicción de series temporales económicas

Esta línea de investigación se centra en la aplicación de distintas arquitecturas de redes *backpropagation* para la predicción del movimiento de índices de bolsa. En particular en [7] se ha explorado la capacidad predictiva en mercados alcistas, bajistas o constantes (*bull, bear and constant markets*). Los resultados obtenidos permiten afirmar que las Redes Neuronales tienen un poder predictivo que permite superar la estrategia pasiva de comprar y esperar. Sin embargo el poder predictivo dista de ser perfecto. (entre el 2,5% y el 36% de una predicción perfecta). Además se ha encontrado que la red se comporta mejor para mercados estables o bajistas.

En [8] se ha analizado la relación entre el poder predictivo de la Red neuronal y el nivel de memoria de largo plazo medido por el exponente de Hurst. En dicho trabajo se realizó una prueba empírica sobre dos series temporales, una de un índice de acciones y otra de un índice bonos de empresas privadas. El primer índice presenta un exponente de Hurst estable y cercano a 0,5 (ausencia de memoria de largo plazo). El segundo índice presenta una etapa de ausencia de memoria de largo plazo y una se-



gunda etapa con un exponente de Hurst superior a 0,5 lo que indicaría memoria de largo plazo. Se ha probado la performance de la red en ambos subperíodos para los dos índices. Se ha encontrado que la capacidad predictiva de la red es mejor en ausencia de la memoria de largo plazo, aunque la diferencia no es significativa. En forma consistente con [7], se ha encontrado que la red predice mejor las bajadas del índice que las subidas.

En base a los resultados obtenidos, actualmente se está analizando el uso de arquitecturas alternativas para encontrar un modelo óptimo de predicción haciendo énfasis en la incorporación del exponente de Hurst como variable predictiva de la red, con el objetivo de mejorar el resultado total.

### 1.3 Optimización utilizando PSO

Esta línea de trabajo está basada en la investigación y aplicación de metaheurísticas poblacionales para la resolución de problemas mono-objetivo y multi-objetivo.

En el III-LIDI se ha investigado en profundidad la metaheurística PSO (Particle Swarm Optimization) definida originalmente en [6] y se han propuesto variantes del algoritmo alcanzando resultados satisfactorios. En [9] se incorporó a PSO el concepto de población variable. Este aspecto permite terminar con la relación que existe entre el tamaño de la población y la calidad de los resultados obtenidos. En [10] se redujo el tiempo de convergencia de PSO mediante la detección de las oscilaciones que habitualmente se producen alrededor del óptimo a medida que la partícula se frena al lograr la convergencia. En [13] se ha definido una nueva versión de PSO Binario combinando una representación discreta (o binaria) con otra continua que al ser utilizada para resolver problemas de optimización de funciones ha demostrado ser capaz de brindar resultados superadores.

Existe un tipo especial de problemas de optimización caracterizados por requerir la minimización o maximización simultánea de más de un objetivo a la vez, estando generalmente estos objetivos en conflicto. Se han desarrollado variantes de las diferentes metaheurísticas existen-

tes para abordar la solución de estos problemas de optimización multiobjetivo.

La solución en este tipo de problemas está constituida por un conjunto de soluciones óptimas, conocido como el frente de Pareto, y no por una única solución como en los problemas mono-objetivo. Existen diferentes versiones de la metaheurística PSO adaptadas para la resolución de problemas con más de un objetivo (MOPSO – Multi Objective Particle Swarm Optimization).

Dentro del ámbito del trabajo realizado por el III-LIDI se han propuesto versiones originales de algoritmos del tipo MOPSO, aplicando las mismas a la resolución de problemas de laboratorio y del mundo real. En [11] se propone una variante del algoritmo PSO con población variable, basado en la propuesta realizada en [9]. Esta versión del algoritmo demostró la obtención de muy buenos resultados, superando a técnicas de optimización referentes en este campo de investigación.

En [12] se evalúan un conjunto de técnicas multi-objetivo aplicadas a un caso de negocio de la industria de la salud, la optimización del proceso de asignación de móviles a emergencias médicas. En este caso, se persigue la minimización simultánea del tiempo de atención a cada prestación y la reducción en la utilización de móviles de terceros. Ofrecer al responsable de la toma de decisiones el conjunto de soluciones óptimas encontradas (Frente de Pareto) le permite entender el problema, visualizar el compromiso entre los diferentes objetivos, y aprender acerca de la interacción entre los criterios en juego. Es importante destacar que el trabajo realizado no se limitó solo a la investigación y evaluación de alternativas, si no que la solución propuesta se implementó para resolver el problema en el mundo real.

## 2. Líneas de investigación y desarrollo

- Estudio de representaciones vectoriales numéricas aplicables a señales biométricas.
- Desarrollo e implementación, a partir de los métodos existentes, de estrategias adaptativas capaces de construir y mantener modelos adecuados en entornos de información dinámicos.

- Diseño y adaptación de redes neuronales aplicables a series temporales financieras.
- Estudio del índice de Hurst como memoria a largo plazo y su relación con el aprendizaje de las redes neuronales.
- Optimización mono y multi objetivo. Desarrollo de nuevas variantes de algoritmos tipo PSO y evaluación de rendimiento.
- Estudio, evaluación y aplicación de metaheurísticas evolutivas multi-objetivo a la resolución del problema de ruteo de vehículos y en especial a la asignación de ambulancias de una empresa de emergencias médicas.
- Estudios de performance de los algoritmos desarrollados. Análisis de eficiencia en la resolución de problemas concretos.
- Análisis de los distintos tipos de Redes Neuronales competitivas dinámicas.

### 3. Resultados obtenidos

- Diseño y desarrollo de un reconocedor biométrico aplicable a señales de audio e imagen haciendo énfasis en los siguientes aspectos:
  - Representación adecuada de la señal de entrada (coeficientes cepstrales para audio y descriptores SIFT para imagen).
  - Diseño y entrenamiento de la red neuronal difusa encargada de sugerir una respuesta.
  - Diseño y entrenamiento de una red feed-forward que a partir de un cambio del espacio de entrada realizado por la RN difusa decide si la respuesta sugerida es correcta o no.
- Diseño y entrenamiento de una red neuronal que, operando sobre series temporales económico-financieras, posee una capacidad de estimación ligeramente mejor que la estrategia *buy-and-hold*.
- Se comprobó que la memoria a largo plazo del índice de Hurst no incide significativamente en la capacidad de predicción de una red neuronal feedforward.
- Desarrollo e implementación de estrategias basadas en cúmulos de partículas (PSO) aplicables a la optimización de funciones y a

la asignación de recursos. Se trabaja sobre los siguientes conceptos

- Población variable.
  - Movimiento de partículas discretas (representación binaria) utilizando vectores de velocidad continuos.
  - Aplicación de PSO a problemas multiobjetivos
- Investigación, desarrollo, y puesta en marcha de optimización multi-objetivo en un caso real de negocios en la industria de la salud.

### 4. Formación de Recursos Humanos

Dentro de los temas involucrados en esta línea de investigación se están desarrollando actualmente 2 tesis de doctorado, 1 de maestría y al menos 3 tesinas de grado de Licenciatura. También participan en el desarrollo de las tareas becarios y pasantes del III-LIDI.

### Referencias

- [1] "Voice recognition based on probabilistic SOM". Estrebow C., Lanzarini L., Hasperué W. Conferencia Latinoamericana de Informática. CLEI 2010. Paraguay. Octubre 2010.
- [2] "Voice Recognition based on vote-SOM". Estrebow C., Hasperué W., Lanzarini L. I Chilean Workshop on Pattern Recognition: Theory and Applications (CWPR). Jornadas Chilenas de Computación 2009. pag.89-93. Santiago de Chile. Noviembre de 2009.
- [3] "Distinctive Image Features from Scale-Invariant Keypoints". Lowe D. International Journal of Computer Vision. Springer Netherlands. ISSN 0920-5691. Vol. 60, nro. 2/2004. Pages 91-110.
- [4] "Face recognition using SIFT and binary PSO descriptors". Lanzarini L., La Battaglia J., Maulini J., Hasperué W. 32th International Conference on Information Technology Interfaces. Publicado por IEEE Computer Society Press. June 2010. Association. Pags.SA-002-1 a SA-002-5. ISBN: 978-602-99359-0-5. Junio 2011. Surabaya and Bali Island Indonesia.

- [5] "Face Recognition using SIFT descriptors and Binary PSO with velocity control". Maulini, Lanzarini. Publicado en el Libro *Computer Science & Technology Series – XVII Argentine Congress of Computer Science Selected Papers*. ISBN 978-950-34-0885-8. Pags. 43-53. EDULP, 2012.
- [6] "Particle Swarm Optimization". Kennedy J., Eberhart R. In Proceedings of the IEEE International Joint Conference on Neural Networks, pp 1942-1948, IEEE Press, 1995.
- [7] "Are Technical Trading Rules Useful to Beat the Market? Evidence from the Brazilian Stock Market". Laura Lanzarini, Martín Iglesias y Aurelio Fernández. IFSA 2011 - World Congress of International *Fuzzy Systems*
- [8] "The effect of changes of the hurst exponent in return predictability: the case of the dutch market". Lanzarini, Fernández Bariviera, Guercio, Tomás-Monterde. Publicado en el libro *Methods for Decision Making in an uncertain environment*. World Scientific Proceedings Series on Computer Engineering and Information Science. Proceedings of the XVII SIGEF Congress. Volume 6, pags.384-393. ISBN 978-981-4415-76-7. Reus -Cambrils. España. Junio 2012.
- [9] "Particle Swarm Optimization with Variable Population Size". Lanzarini L., Leza V., De Giusti A. Lecture Notes in Computer Science. Vol 5097/2008. Artificial Intelligence and Soft Computing. Pags. 438-449. Junio de 2008. ISBN 987-3-540-69572-1. Springer.
- [10] "Particle Swarm Optimization with Oscillation Control". López J., Lanzarini L., De Giusti A. ACM Special Interest Group on Genetic and Evolutionary Computation. Pages: 1751-1752. 2009. ISBN:978-1-60558-325-9
- [11] "Evolutionary Multiobjetive Optimization for Emergency Medical Services". López J., Lanzarini L., De Giusti A. Genetic and Evolutionary Computation Conference, GECCO 2011. Dublin – Irlanda. July 12-16, 2011.
- [12] "Variable Population MOPSO applied to Medical Visits". López, Lanzarini, Fernández Bariviera. Journal *Fuzzy Economic Review*. Vol XVII, nro 1. International Association for Fuzzy Set Management and Economy (SIGEF). Mayo 2012. Pags.3-14. ISSN 1136-0593
- [13] "A new Binary PSO with velocity control". Lanzarini L., López J., Maulini J., De Giusti A. Advances in Swarm Intelligence. Lecture Notes in Computer Science. 2011, Volume 6728/2011, 111-119. Springer 2011.

# Metaheurísticas secuenciales y distribuidas: Adaptación de parámetros y entornos de ejecución

Carolina Salto, Gabriela Minetti, Natalia Stark, Carlos Bermudez, Hugo Alfonso, German Dupuy  
 Laboratorio de Investigación en Sistemas Inteligentes (LISI)  
 Facultad de Ingeniería - Universidad Nacional de La Pampa  
 Calle 110 Esq. 9 (6360) General Pico - La Pampa - Rep. Argentina  
 Te. / Fax: (02302) 422780/422372, Int. 6302

## Resumen

La contribución de esta línea de investigación es proponer mecanismos adaptativos que puedan guiar el cambio de ciertos parámetros de los algoritmos evolutivos, tanto secuenciales como paralelos, usando información del estado del proceso de búsqueda. Otro aspecto importante es el referido al entorno de ejecución de estos algoritmos, a tales efectos se analiza la incidencia del uso de plataformas heterogéneas, tanto en la calidad de los resultados finales como en los tiempos de ejecución.

## CONTEXTO

Esta línea de investigación se desarrolla en el marco del proyecto de investigación "Resolviendo problemas complejos con técnicas metaheurísticas avanzadas", dirigido por la Dra. Carolina Salto, y llevado a cabo en el Laboratorio de Investigación de Sistemas Inteligentes (LISI), de la Facultad de Ingeniería de la Universidad Nacional de La Pampa. Este proyecto mantiene desde hace varios años una importante vinculación con investigadores de la Universidad Nacional de San Luis (Argentina) y de la Universidad de Málaga (España) con quienes se han realizado varias publicaciones conjuntas.

## I. INTRODUCCIÓN

La eficiencia de un algoritmo evolutivo (AE) depende de muchos parámetros, tales como tamaño de la población, representación de los individuos, estrategias de selección, probabilidades de los operadores, número total de generaciones, entre otros. Determinar cuál es el valor más apropiado de un parámetro para un determinado problema es en sí mismo un problema complejo, debido a que existe una interacción no lineal entre los parámetros y no existe un modelo matemático de tal interacción.

Entre los parámetros críticos para el rendimiento de un AE podemos mencionar la probabilidad de mutación  $p_m$ . Valores grande de  $p_m$  transforman al AE en un algoritmo de búsqueda puramente aleatorio, pero por otro lado es necesario algún grado de mutación para prevenir la convergencia prematura del AE a soluciones subóptimas. En la literatura se han sugeridos muchos valores para  $p_m$ , los cuales se derivan de la experiencia o prueba y ensayo [4], [6], [8], [13]. Es difícil, sino imposible, hallar un seteo apropiado para el valor de  $p_m$  a fin de alcanzar un rendimiento óptimo, debido a que las probabilidades varían con el problema que se intenta resolver,

y también están influenciadas con las diferentes etapas del proceso genético.

Esta problemática de seteo de parámetros también está presente en los AEs distribuidos. En particular, los AEs distribuidos introducen otro operador responsable de la estructura del algoritmo y del esquema de intercomunicación entre las subpoblaciones o islas (política de migración), conocido como el operador de migración. Por lo tanto, se necesitan parámetros adicionales a los de un AE secuencial; esta situación aumenta la complejidad en el uso de la técnica. Como parámetros que definen la política de migración podemos encontrar:

- periodo de migración: el periodo de evolución aislada de las subpoblaciones
- tasa de migración: el número de individuos que se intercambian.
- elección / reemplazo de los migrantes: estrategias para la selección de los individuos que se envían en una operación de migración y para la sustitución de los individuos locales con los entrantes.
- topología: este parámetro define el vecindario de cada subpoblación.

Aunque muchos parámetros regulan la migración, la mayoría de ellos son estándares y ampliamente usados, como una topología de intercomunicación en anillo, enviar uno o pocos individuos, reemplazar el peor individuo de la población receptora, y usar comunicación asíncrona para obtener buenos valores de speedup.

Para determinar el valor óptimo de los parámetros del AE, tanto secuencial como distribuido, se pueden considerar al menos dos opciones: ajuste (*tuning*) de parámetros y control de parámetros [7]. La primera opción está relacionada con hacer una experimentación exhaustiva para determinar, de una manera empírica, el valor del parámetro que permite obtener los mejores resultados para el problema en cuestión. Esta tarea implica un tiempo de seteo extra, la cual es computacionalmente muy costosa y raramente reportada en la literatura. En esta línea, se han reportado varias alternativas, tal como diseño de experimentos [14], F-RACE [5] y REVAC [12]. La segunda opción consiste en hacer cambios en los valores de los parámetros durante la ejecución del AE. Esto requiere valores iniciales para los parámetros y estrategias de control, las cuales pueden ser determinísticas, adaptativas o



auto-adaptativas. Estas estrategias manejan información que influyen los valores de los parámetros durante la búsqueda y definen un conjunto de criterios para producir los cambios. La línea de investigación que se describe en este trabajo utiliza la técnica de control de parámetros, en particular se plantean estrategias adaptativas.

Cuando se estudia un algoritmo evolutivo paralelo, es importante tener en cuenta sobre qué plataforma de computación se ha implementado, ya que la arquitectura de hardware impacta notablemente en el tiempo requerido para realizar las operaciones, comunicaciones, sincronizaciones y compartimiento de datos. Hasta la última década, las propuestas tradicionales de metaheurísticas se enfocaban en supercomputadores y clusters de computadoras. Actualmente, las novedosas arquitecturas de computadoras paralelas emergentes tal como procesadores multicore, unidades de procesamiento gráfico (GPU), o entornos grid, proveen nuevas oportunidades para desarrollar técnicas de computación paralelas para mejorar la resolución de problemas y disminuir el tiempo de ejecución requerido.

Los ambientes de computación paralelos y distribuidos se han tornado populares en las últimas décadas como una forma de proveer el poder de cómputo necesario para resolver problemas complejos, representando una estrategia efectiva para la ejecución de AEs distribuidos. La mayoría de los resultados reportados sobre AEs distribuidos asumen que el ambiente de hardware subyacente tiene características idénticas (ambiente homogéneo) teniendo en cuenta no sólo componentes de hardware (procesadores, memoria, redes) sino también software (sistema operativo) [1]. Por otra parte, esta clase de homogeneidad en hardware es muy difícil de mantener en el transcurso del tiempo. Una de las principales razones son las roturas de las máquinas y la dificultad de construir un sistema igual teniendo en cuenta el mismo procesador, arquitectura interna y configuración. Por otra parte, el rápido desarrollo de la tecnología en el diseño de procesadores, redes y almacenamiento de datos sumado al constante decremento de la relación costo/rendimiento, posibilita el uso de nuevos recursos disponibles. Como consecuencia, la coexistencia de equipamientos nuevos y viejos en un ambiente de computación ha dado lugar a la emergencia de plataformas paralelas heterogéneas, las cuales son actualmente muy comunes en laboratorios, compañías, instituciones, campus, etc. Uno de los primeros trabajos que trata con ambientes de computación heterogéneos y AEs distribuidos se puede hallar en [2]. Trabajos más recientes relacionados con ambientes heterogéneos se pueden encontrar en [3], [11].

Las GPUs, originalmente diseñadas como dispositivos específicos para procesamiento gráfico, se han transformado en plataformas muy poderosas a costos accesibles para utilizar con propósitos de cómputo general [9]. Por esta razón, el estudio de implementaciones de AEs usando GPUs [10] ha crecido a pasos agigantados, ya que ayuda a reducir el tiempo de ejecución de estos algoritmos mediante la explotación del paralelismo masivo de tales dispositivos.

## II. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

En esta sección se describen la línea de investigación y desarrollo, junto con las sublíneas que se desprenden.

La definición del valor apropiado para el periodo de migración en un AE distribuido requiere considerar varios valores (desde bajos a altos) siguiendo un estudio similar a un análisis factorial, realizar la experimentación previa correspondiente, para finalmente analizar cuál de esos periodos será usado en la resolución del problema. Esta tarea es computacionalmente costosa y el valor apropiado para el periodo de migración depende de las características del problema a resolver. En esta línea de investigación se propone una estrategia diferente: un AE distribuido proactivo que adaptativamente define el valor del periodo cuando la entropía (medida de diversidad genética) de cada subpoblación satisface una determinada condición; el objetivo es mantener la diversidad genética en valores aceptables. Por consiguiente, cuando la entropía cae debajo de cierto umbral (se detecta pérdida de diversidad), esas subpoblaciones proactivamente solicitan nuevo material genético al reducir el valor del periodo de migración de la subpoblación que le provee inmigrantes, es decir, se producen migraciones más frecuentes. De esta manera, si la topología de la estrategia de migración es un anillo, la isla  $i$  modifica el periodo de migración de la isla precedente  $i-1$  solicitando migraciones más frecuentes. Análogamente, la propuesta también reacciona cuando detecta mucha diversidad, en este caso incrementa el periodo de migración de la isla  $i-1$  para alargar el periodo de aislamiento entre esas islas. La continuación que se propone a esta línea de investigación es el uso de diferentes métricas para la toma de decisiones, como así también la incorporación de más parámetros de migración al mecanismo proactivo del AE distribuido.

En esta misma línea, se está trabajando en la adaptación de otro parámetro de un AE secuencial como lo es la probabilidad de mutación, esta variación se realiza dependiendo de la diversidad genética presente en la población. El nuevo método adaptativo pretende mejorar el comportamiento del AE, evitando la convergencia prematura y la pérdida de diversidad. Una de las ventajas de este método es que no se introducen grandes cambios en la formulación original del algoritmo. El objetivo de nuestra estrategia adaptativa es incrementar el valor de  $p_m$  si la diversidad genética (medida por el valor de la entropía de la población) se pierde gradualmente, con la intención de mantener una población distribuida en el espacio de búsqueda. El valor de  $p_m$  se decrementa cuando se observa un incremento en la diversidad de la población. De esta manera, estos cambios en el valor de la probabilidad es una fuente adicional para mantener el balance entre la exploración y explotación del espacio de búsqueda. En este caso, la acción futura está orientada a perfeccionar el mecanismo auto-adaptativo considerando nuevos criterios e incorporando otros parámetros del AE como pueden ser las probabilidades de cruce. Por otra parte, para fusionar esta línea de investigación con la anterior, se podría considerar este AE auto-adaptativo en un entorno

distribuido y así avanzar en la propuesta de AEs distribuidos heterogéneos.

Acompañando la propuesta adaptativa también se puede considerar dentro de esta línea la plataforma de computación en la cual son implementadas las distintas propuestas. En este sentido hemos considerado la ejecución un algoritmo evolutivo distribuido en un conjunto de procesadores con rendimientos dispares, desde modernos a viejos. Las diferentes subpoblaciones son evolucionadas por procesadores con diferentes velocidades, en consecuencia en un determinado momento cada una se encuentra en diferentes estados de la evolución. Tan pronto una isla ejecutando en un procesador rápido llega al intercambio de soluciones, envía una solución a sus vecinos y continua con su procedimiento de optimización de una manera asíncrona. Las soluciones desde sus vecinos más lentos deberán tomar más tiempo para diseminarse a otras islas. La metodología consiste en calcular las diferencias relativas entre las velocidades de cada procesador. Con esta información, se puede establecer un ordenamiento que puede ayudar a tomar decisiones más informadas de los valores de los parámetros del algoritmo a fin de contar con un mecanismo que permita al AE distribuido tratar con diferencias en las características en las velocidades de los procesadores. El trabajo futuro está orientado a incluir más niveles de heterogeneidad, por tal motivo consideraremos unidades de procesamiento gráficos y procesadores multicore.

### III. RESULTADOS OBTENIDOS/ESPERADOS

En esta sección se detallan los resultados obtenidos en las distintas líneas de investigación en desarrollo, presentadas en la sección anterior.

Los resultados del AE adaptativo que usa una estrategia de control para modificar la probabilidad de mutación, denominado APmAE, fue comparado con los producidos por las distintas variantes de AE utilizando probabilidades de mutación estáticas, con el fin de mostrar la robustez del algoritmo adaptativo como un método de búsqueda alternativo [16]. Como estamos interesados en la auto-adaptación de la probabilidad de mutación, en el caso del AE se presenta un conjunto de valores para la probabilidad de mutación pertenecientes a  $\{0.001, 0.005, 0.01, 0.05, 0.1, 0.25, 0.5\}$ , con el fin de caracterizar el efecto de la probabilidad de mutación en la calidad de las soluciones obtenidas. Como resultado general obtuvimos resultados muy prometedores. En particular, el APmAE obtuvo mejor calidad de soluciones que las distintas variantes de AE con probabilidad de mutación fija, con velocidades de convergencia similares. Un punto importante a resaltar es que nuestra propuesta no adiciona tiempo de cómputo al aplicar la estrategia de control diseñada para adaptar los valores de la probabilidad de mutación. Finalmente, como el APmAE determina en forma dinámica la probabilidad de mutación durante el proceso de búsqueda, este método ayuda a aliviar el tiempo de seteo requerido, al menos en un 85 %, para determinar la probabilidad de mutación más adecuada.

Para evaluar los resultados obtenidos con el AE distribuido y proactivo, hemos realizado una comparación con varios algoritmos AEs distribuidos (no proactivos) que mantienen el parámetro de migración fijo con valores para la frecuencia de migración en  $\{1, 64, 128, 256, 512\}$ , es decir, desde un máximo de ensamble entre las islas hasta un proceso de búsqueda prácticamente en aislamiento. Los resultados indicaron que el algoritmo proactivo es capaz de encontrar soluciones al problema planteado de calidad similar al mejor algoritmo AE distribuido homogéneo [15]. El algoritmo proactivo ha funcionado especialmente bien sobre las instancias más difíciles del problema, donde claramente se impone al mejor homogéneo. Un punto interesante de nuestra propuesta es que presenta una mejor velocidad de convergencia, es decir, ha obtenido sus mejores soluciones en un número más bajo de generaciones que cualquiera de las versiones homogéneas.

En cuanto a la sublínea referida al entorno de ejecución, hemos dado un primer paso referente a la paralelización de un AE básico para ser ejecutado sobre una GPU. Los resultados iniciales han sido alentadores, ya que se ha logrado resolver un conjunto de instancias de grandes dimensiones de un problemas de optimización, mejorando los óptimos conocidos en tiempos de cómputo razonables. Por otra parte, hemos desarrollado una propuesta sobre un procedimiento metodológico, y el subsecuente diseño algorítmico, para tratar con ambientes paralelos heterogéneos. Resultados preliminares muestran que el nuevo mecanismo tiene un rendimiento similar a un conjunto de máquinas modernas ejecutando el mismo procedimiento de optimización, esto permite a los investigadores usar todos los recursos de procesamiento disponibles en los laboratorios.

### IV. FORMACIÓN DE RECURSOS HUMANOS

La actividad de formación de recursos humanos está orientada a la formación de becarios de iniciación en la investigación. En este marco, un alumno de la carrera Ingeniería en Sistemas se encuentra desarrollando las actividades planificadas para una Beca de Estímulo a las Vocaciones Científicas, otorgada por el CIN para el periodo 2012-2013. Por otra parte, es interesante resaltar que en el LISI se trabaja con alumnos avanzados en la carrera Ingeniería en Sistemas en temas relacionados a la resolución de problemas de optimización usando técnicas inteligentes, con el objeto de guiarlos en el desarrollo de sus tesis de grado y, también, de formar futuros investigadores.

### REFERENCES

- [1] E. Alba. *Parallel Metaheuristics: A New Class of Algorithms*. Wiley, 2005.
- [2] E. Alba, A.J. Nebro, and J.M. Troya. Heterogeneous computing and parallel genetic algorithms. *Journal of Parallel and Distributed Computing*, 62:1362–1385, 2002.
- [3] J. Dominguez E. Alba. A methodology for comparing the execution time of metaheuristics running on different hardware. In Jin-Kao Hao and Martin Middendorf, editors, *Evolutionary Computation in Combinatorial Optimization*, volume 7245 of *Lecture Notes in Computer Science*, pages 1–12. Springer Berlin Heidelberg, 2012.
- [4] Thomas Bäck. Self-adaptation in genetic algorithms. In *Proceedings of the First European Conference on Artificial Life*, pages 263–271. MIT Press, 1992.

- [5] M. Birattari, T. Stutzle, L. Paquete, and K. Varrentrapp. A racing algorithm for configuring metaheuristics. In *GECCO 2002*, pages 11 – 19, 2002.
- [6] K. DeJong. *An analysis of the behavior of a class of genetic adaptive systems*. PhD thesis, University of Michigan, Ann Arbor, MI, 1975.
- [7] A. Eiben, R. Hinterding, and Z. Michalewicz. Parameter control in evolutionary algorithms. *IEEE Transactions Evolutionary Computation*, 3(2):124–141, 1999.
- [8] J. J. Greffentette. Optimization of control parameters for genetic algorithms. *IEEE Transaction on System Man and Cybernetic*, 16(1):122–128, 1986.
- [9] T. Wong K. Fok and M. Wong. Evolutionary computing on consumer graphics hardware. *Intelligent Systems*, 22(2):69–78, 2007.
- [10] W. Langdon. Graphics processing units and genetic programming: an overview. *Soft Computing - A Fusion of Foundations, Methodologies and Applications*, pages 1–13, 2011.
- [11] S. Mostaghim, J. Branke, A. Lewis, and H. Schneck. Parallel multi-objective optimization using master-slave model on heterogeneous resources. In *Evolutionary Computation, 2008. CEC 2008. (IEEE World Congress on Computational Intelligence). IEEE Congress on*, pages 1981 –1987, 2008.
- [12] V.Nannen and A. Eiben. Relevance estimation and value calibration of evolutionary algorithm parameters. *Proceedings of the Joint International Conference for Artificial Intelligence (IJCAI)*, pages 975–980, 2006.
- [13] Gabriela Ochoa, Inman Harvey, and Hilary Buxton. On recombination and optimal mutation rates. In *in Proceedings of Genetic and Evolutionary Computation Conference*, pages 488–495. Morgan Kaufmann, 1999.
- [14] E. Ridge and D. Kudenko. Tuning an algorithm using design of experiments. In *Experimental Methods for the Analysis of Optimization Algorithms*, pages 265 – 286. Springer, 2010.
- [15] C. Salto, F. Luna, and E. Alba. Heterogeneity through proactivity: Enhancing distributed eas. *Seventh International Conference on Parallel, Grid, Cloud and Internet Computing (3PGCIC)*, pages 279–284, 2012.
- [16] N. Stark, G. Minetti, and C. Salto. A new strategy for adapting the mutation probability in genetic algorithms. *XVIII Congreso Argentino de Ciencias de la Computación (CACIC 2012)*, pages 51–59, 2012.

# Argumentación rebatible en agentes adaptativos de escenarios dinámicos

Juan Manuel Torres, Diego C. Martínez, y Guillermo R. Simari

Laboratorio de Investigación y Desarrollo en Inteligencia Artificial (LIDIA)  
Departamento de Ciencias e Ingeniería de la Computación (DCIC)  
Universidad Nacional del Sur (UNS)  
Av. Alem 1253, (B8000CBP), Bahía Blanca, Argentina  
Tel: (0291) 459-5135 / Fax: (0291) 459-5136  
{juan.torres, dcm, grs}@cs.uns.edu.ar

## Resumen

Un agente inteligente es una entidad computacional que se desenvuelve en un entorno mediante sensores y efectores, capaz de actuar sin intervención de otros agentes o humanos. Un agente adaptativo es aquel que tiene especial preparación para reaccionar a variaciones inesperadas en su entorno. En este contexto, el área de Juegos Generales se enfoca en la creación de agentes capaces de aceptar las reglas formales de un juego estratégico y utilizarlas para aprender a jugar. Esto potencia el avance en conceptos como aprendizaje automatizado, representación de conocimiento, razonamiento y reconocimiento de patrones.

La argumentación rebatible es un mecanismo de razonamiento no monótono que permite capturar aspectos del razonamiento del sentido común y la representación de información incompleta y potencialmente inconsistente. Los procesos argumentativos son inherentemente dialécticos y permiten modelar adecuadamente interacciones sociales entre agentes inteligentes.

El objetivo general de esta investigación es la integración de formalismos de argumentación rebatible en agentes adaptativos en entornos dinámicos. Particularmente, el diseño y desarrollo de técnicas y tecnologías argumentativas para asistir a agentes capaces de enfrentar juegos generales como modelo de problema de razonamiento.

**Palabras clave:** *Agentes adaptativos, Argumentación, Juegos Generales.*

## 1. Contexto

Esta línea de investigación se encuadra dentro de los lineamientos de los Proyectos de Investigación “*Agentes Inteligentes y Creibles en Ambientes Interactivos Digitales (24/ZN22)*” y “*Representación de Conocimiento y Razonamiento Argumentativo: herramientas inteligentes para la web y las bases de datos federadas (24/N030)*”, desarrollados en el Laboratorio de Investigación y Desarrollo en Inteligencia Artificial (LIDIA) del Departamento de Ciencias e Ingeniería de la Computación de la Universidad Nacional del Sur.

## 2. Introducción

A partir del surgimiento del concepto de agente inteligente en las Ciencias de la Computación, su estudio ha tomado diferentes perspectivas, dependiendo del rol que cumple un agente en diversos escenarios o dominios de aplicación [1,2]. Actualmente los agentes tienen un campo de aplicación muy amplio y existen muchos tipos de agentes diferentes (por ejemplo: agentes reactivos, deliberativos, inteligentes, de interface, colaborativos, etc.), los cuales a su vez están orientados a distintos entornos de aplicación.

El consenso general es que un agente inteligente es una entidad autónoma que es capaz de percibir su entorno y de actuar en él acordemente, posiblemente alterándolo en



diversos grados. Decir que es autónoma significa que tiene control sobre su propio comportamiento y que puede actuar sin la intervención de otros agentes o humanos. En particular, los agentes adaptativos [3–5] son una clase especial de agentes que tiene especial preparación para reaccionar a variaciones inesperadas en su entorno. Un buen ejemplo de condiciones cambiantes que requieren adaptación son los juegos estratégicos, desde el ajedrez hasta los juegos interactivos de estrategia en tiempo real.

En agentes adaptativos, el énfasis es puesto en mecanismos de detección, comprensión y asimilación de variaciones o cambios en el entorno en el que el agente se desenvuelve. Estos cambios requieren una revisión del comportamiento del agente para adaptarse al nuevo escenario. Puede implicar cambios en sus prioridades, en la toma de decisiones, en las heurísticas de comportamiento e incluso en sus objetivos intrínsecos. Actualmente el desarrollo científico y tecnológico en agentes adaptativos suele ser presentado y testeado en contextos de juegos o simulaciones dinámicas [3–5]. Este tipo de agentes resultan de especial interés en el desarrollo de dispositivos autónomos (como los UAV - Unmanned Aerial Vehicles), en simulaciones, en economía y en las aplicaciones de entretenimiento interactivas, entre otras disciplinas.

La Inteligencia Artificial ha tenido una estrecha relación con los juegos, pues son un desafío natural para las expresiones de vida inteligente. Históricamente, los agentes capaces de jugar competitivamente fueron diseñados para algún juego en particular [1]. Este es el caso de la computadora Deep Blue de IBM, aunque con algoritmos no sofisticados técnicamente, fue capaz de vencer a un maestro de ajedrez. El programa Chinook [6] fue diseñado para jugar a las damas, siendo el primero en ganar torneos contra humanos. Sin embargo, aun cuando estos programas eran capaces de vencer oponentes expertos, no eran capaces de jugar cualquier otro juego.

Los juegos generales (originalmente en inglés, General Game Playing o GGP) [7, 8] son un área novedosa dentro de la Inteligencia Artificial. Aquí el objetivo es el diseño y desarrollo de agentes autónomos capaces de jugar juegos cuyas reglas no son conocidas con anticipación. En los juegos generales, las reglas son formalizadas declarativamente y los agentes deben comprender el juego y aprender estrategias para jugar exitosamente sin intervención humana. Las propiedades de los juegos generales no son conocidas por el desarrollador del agente, sino que deben ser des-

cubiertas por el agente en tiempo de ejecución. Luego, el agente no es una entidad construida para jugar algún juego en particular, sino cualquier juego general apropiadamente formalizado. En General Game Playing, los agentes no son provistos con conocimiento específico del dominio del juego, sino que deben ser capaces de aprender estrategias y razonar en forma abstracta. Esta demanda de máxima flexibilidad requiere el uso e integración de varias técnicas de Inteligencia Artificial. Si bien los juegos generales son un tópico de interés creciente en sí mismos, la motivación superior es que los avances científicos y tecnológicos en el área pueden ser utilizados en una gran variedad de dominios, tales como la administración de empresas, el comercio electrónico, la simulación de emergencias civiles y operaciones militares, y la toma de decisiones en escenarios con grados de incertidumbre.

Uno de los precursores de este tópico dentro de la Inteligencia Artificial es la Universidad de Stanford, que desde el año 2005 organiza y dirige la General Game Playing Annual Competition para incentivar los avances en el área. Los participantes compiten unos contra otros en una gran variedad de juegos. Se trata de juegos de  $n$  jugadores, finitos, determinísticos y de información perfecta, descriptos utilizando un lenguaje formal denominado Game Description Language [9]. Este lenguaje es una variante de la lógica de primer orden que incluye símbolos para la conceptualización de juegos. En esta descripción se explican, entre otras características, las movidas válidas y las condiciones de éxito. Muchos investigadores en Inteligencia Artificial han volcado recientemente parte de sus desarrollos científicos y tecnológicos en este tipo de juegos [8, 10–15], pues es un campo válido y concreto para evaluar y comparar avances. Sin embargo, todavía no se ha explorado el uso de técnicas de argumentación rebatible en agentes adaptativos de este tipo, aun cuando ésta es un área de aplicación razonable para la argumentación.

La argumentación es una forma de razonamiento en la cual, para una afirmación dada, se presta atención explícita a las justificaciones presentadas y a la resolución de los posibles conflictos entre ellas. En este tipo de razonamiento, una afirmación es aceptada o no según el análisis de los argumentos a su favor y en su contra. La forma en que los argumentos y las justificaciones para una afirmación son considerados, permite definir un tipo de razonamiento automático, en el cual puede haber información contradictoria, incompleta, e incierta. El estudio de la argumenta-

ción ha sido abordado desde diferentes enfoques en años recientes. A nivel lógico puede considerarse como una forma de modelar inferencia rebatible, y a nivel dialógico como una forma de interacción entre agentes. Actualmente es un atractivo paradigma para conceptualizar el razonamiento de sentido común (ver por ejemplo los surveys [16, 17]). Esto produjo como resultado la formalización de diferentes frameworks de argumentación abstracta como [18–20], entre otros; y de Sistemas de Argumentación Basados en Reglas (SABR) como [21–23]; lo que permitió el desarrollo de diversas aplicaciones del mundo real basadas en argumentación. En el último tiempo, el campo de aplicación de la argumentación se ha expandido velozmente, en gran parte debido a los avances teóricos, pero también gracias a la demostración exitosa de su uso práctico en un gran número de dominios de aplicación, tales como el razonamiento legal [23], el razonamiento temporal [24], los sistemas multiagentes [2, 25] y la representación de conocimiento [26, 27] entre muchos otros. En particular, la argumentación rebatible es adecuada para la implementación del razonamiento y representación de conocimiento en agentes inteligentes y, tal cual la hipótesis de esta línea de investigación, adecuada para modelar agentes que exhiban un alto grado de adaptabilidad en escenarios dinámicos.

### 3. Líneas de investigación y desarrollo

Esta línea de investigación está centrada en la integración de formalismos de argumentación rebatible en agentes inteligentes adaptativos en entornos dinámicos. En particular, resulta de especial interés el estudio de los Juegos Generales. El diseño de agentes capaces de jugar juegos generales de manera exitosa comprende la incorporación de conceptos como el aprendizaje automatizado, la representación de conocimiento y el razonamiento, entre otros. En este contexto, es posible identificar sub-líneas de trabajo.

En cuanto a los juegos generales, resulta esencial el estudio del lenguaje formal de descripción de dichos juegos, denominado Lenguaje de Descripción de Juegos (Game Description Language) [9]. Este lenguaje es una variante de la lógica de primer orden que permite describir formalmente los juegos detallando una serie de aspectos fundamentales. El análisis del lenguaje hace posible avanzar en tareas

importantes como la representación de la información conocida y la generación de conocimiento nuevo a partir de las descripciones mencionadas.

Otro de los ejes de esta línea de investigación consiste en el estudio de diferentes arquitecturas de agentes inteligentes. En particular, se consideran los formalismos de agentes adaptativos y de entornos dinámicos. La línea de investigación está enmarcada en el uso de formalismos argumentativos como medio para alcanzar grados aceptables de adaptabilidad en el comportamiento inteligente. En este sentido, se estudian comparativamente distintos sistemas de argumentación abstracta ([18–20, 28], entre otros) y los basados en reglas (principalmente Defeasible Logic Programming [21]).

### 4. Resultados y Objetivos

Siguiendo la línea de investigación propuesta, se analizaron en detalle las características del Lenguaje de Descripción de Juegos [9]. Se trata de un lenguaje puramente axiomático, en el cual ningún conocimiento previo es asumido. Está basado en un conjunto de palabras clave con significado predefinido que establecen elementos como los jugadores, las jugadas válidas y las condiciones de terminación y éxito, entre otros. En este contexto, el agente es el encargado de razonar sobre las acciones que le son permitidas, el próximo estado del juego teniendo en cuenta las movidas posibles, y si ha o no resultado ganador.

Se estudiaron formas alternativas de representación de las descripciones formales derivadas del lenguaje. Esto persigue el objetivo de lograr la mejor integración posible con los formalismos empleados para representar el conocimiento de los agentes y modelar su comportamiento. De la misma forma, se observaron las facilidades que provee el lenguaje para realizar nuevas inferencias sobre los juegos. Se evaluaron distintos mecanismos existentes [11, 28], y se diseñaron nuevos métodos que permiten alcanzar un mayor grado de conocimiento de los juegos a partir de sus formalizaciones. La generación de información a grados de abstracción más elevados, más allá de los aspectos básicos del juego explicitados por el lenguaje, resulta fundamental para el desarrollo de estrategias. Entre los resultados de esta línea de investigación, se encuentra el diseño de un mecanismo de razonamiento que dota al agente con la capacidad de discernir si un juego está basado en turnos o permite ac-

ciones simultaneas de distintos jugadores. Esto admite, en el primer caso, el empleo de algoritmos específicos para la resolución del juego.

Como objetivo general se tiene el diseño y desarrollo de técnicas y tecnologías argumentativas que permitan proveer a agentes la capacidad de enfrentar juegos generales como modelo de problema de razonamiento. Entre las metas particulares relacionadas se encuentra enfocar los estudios en dos tipos de formalismos particulares de argumentación: los sistemas argumentativos abstractos (SAA) y los sistemas argumentativos basados en reglas (SABR). Los primeros permitirán enfocarse en el significado general (es decir, la semántica) de la argumentación en el proceso de comprensión del entorno dinámico y en la deliberación de pasos hacia la meta, abstrayendo algunas características particulares. Los segundos son formalismos concretos de sistemas argumentativos que utilizan reglas de inferencia. Los SAA y los SABR son dos formalismos que difieren principalmente en el nivel de abstracción y en ambos casos no han sido explorados como mecanismo de razonamiento en entornos cambiantes y en especial en juegos formalmente especificados como ejemplo de problema de razonamiento. El uso de argumentación en agentes adaptativos para la asimilación de nuevas reglas, la elaboración y análisis de heurísticas y la planificación de movidas en el juego, representará un avance significativo dentro del área de sistemas argumentativos en Inteligencia Artificial y Ciencias de la Computación.

## 5. Formación de Recursos Humanos

Actualmente se encuentran trabajando dos becarios doctorales Tipo I de CONICET vinculados a esta línea de investigación. Participan además profesores y alumnos avanzados en el desarrollo de tesis de grado sobre juegos interactivos y entornos virtuales dinámicos. Existe además un Proyecto de Investigación y Desarrollo PIDDEF 2012-2014 en actividad, financiado por el Ministerio de Defensa, que guarda relación con la investigación en entornos dinámicos y agentes adaptativos y que contempla la participación de un becario para Maestría.

## Referencias

- [1] Mike Van Lent, John Laird, et al. Intelligent agents in computer games. In *In Proceedings of the National Conference on Artificial Intelligence*, pages 929–930, 1999.
- [2] Anand S. Rao and Michael P. Georgeff. Bdi agents: From theory to practice. In *Proceedings of the First International Conference on Multi-agent Systems (ICMA-95)*, pages 312–319, 1995.
- [3] Liang Xiao and Des Greer. Adaptive agent model: Software adaptivity using an agent-oriented model-driven architecture. *Inf. Softw. Technol.*, 51(1):109–137, January 2009.
- [4] Candelaria Sansores and Juan Pavón. An adaptive agent model for self-organizing mas. In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems - Volume 3, AAMAS '08*, pages 1639–1642, 2008.
- [5] Shawn Singh, Mubbasir Kapadia, et al. A modular framework for adaptive agent-based steering. In *Symposium on Interactive 3D Graphics and Games, I3D '11*, pages 141–150 PAGE@9, New York, NY, USA, 2011. ACM.
- [6] Jonathan Schaeffer, Yngvi Björnsson, et al. Solving checkers. In *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence (IJCAI)*, pages 292–297, 2005.
- [7] Michael R. Genesereth, Nathaniel Love, and Barney Pell. General game playing: Overview of the aaai competition. *AI Magazine*, 26(2):62–72, 2005.
- [8] J. Reisinger, E. Bahceci, et al. Coevolving Strategies for General Game Playing. In *Computational Intelligence and Games, 2007. CIG 2007. IEEE Symposium on*, pages 320–327, 2007.
- [9] Nathaniel Love, Timothy Hinrichs, et al. General game playing: Game description language specification. Technical report, Stanford Logic Group, March 2008.
- [10] Martin Günther, Stephan Schiffel, and Michael Thielscher. Factoring General Games. 2009.
- [11] James Clune. Heuristic evaluation functions for general game playing. In *AAAI'07: Proceedings of the 22nd national conference on Artificial Intelligence*, pages 1134–1139. AAAI Press, 2007.
- [12] Stephan Schiffel and Michael Thielscher. Fluxplayer: A successful general game player. In *Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence*, pages 1191–1196, 2007.

- [13] Bikramjit Banerjee and Peter Stone. General game learning using knowledge transfer. In *The 20th International Joint Conference on Artificial Intelligence*, pages 672–677, 2007.
- [14] Hilmar Finnsson and Yngvi Björnsson. Simulation-based approach to general game playing. In *Proceedings of the 23rd national conference on Artificial intelligence - Volume 1*, AAAI'08, pages 259–264. AAAI Press, 2008.
- [15] Ziad Kobti and Shiven Sharma. A multi-agent architecture for game playing. In *IEEE Symposium on Computational Intelligence and Games*, pages 276–281, 2007.
- [16] Carlos I. Chesñevar, Ana G. Maguitman, and Ronald P. Loui. Logical models of argument. 32:337–383, 2000.
- [17] Iyad Rahwan and Guillermo R. Simari. *Argumentation in Artificial Intelligence*. Springer Publishing Company, Incorporated, 1st edition, 2009.
- [18] Phan Minh Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77:321–357, 1995.
- [19] Diego C. Martínez, Ro J. García, and Guillermo R. Simari. G.: Modelling well-structured argumentation lines. In *In: Proc. of XX IJCAI-2007*, pages 465–470, 2007.
- [20] Diego C. Martínez, Alejandro García, and Guillermo R. Simari. Strong and weak forms of abstract argument defense. In *Proceedings of the 2008 conference on Computational Models of Argument: Proceedings of COMMA 2008*, pages 216–227. IOS Press, 2008.
- [21] A. Garcia and G. Simari. Defeasible logic programming: An argumentative approach. *Theory and Practice of Logic Programming (TPLP)*, 4:95–138, 2004.
- [22] Sebastian Gottifredi, Alejandro Garcia, and Guillermo Simari. Argumentation systems and agent programming languages, 2009.
- [23] Henry Prakken and Giovanni Sartor. The role of logic in computational models of legal argument - a critical survey, 2001.
- [24] Maria Laura Cobo, Diego C. Martínez, and Guillermo Ricardo Simari. On admissibility in timed abstract argumentation frameworks. In Helder Coelho, Rudi Studer, and Michael Wooldridge, editors, *ECAI*, volume 215 of *Frontiers in Artificial Intelligence and Applications*, pages 1007–1008. IOS Press, 2010.
- [25] Leila Amgoud, Nicolas Maudet, and Simon Parsons. An argumentation-based semantics for agent communication languages. In Frank van Harmelen, editor, *ECAI*, pages 38–42. IOS Press, 2002.
- [26] Edgardo Ferretti, Marcelo L. Errecalde, et al. Decision rules and arguments in defeasible decision making. In *Proceedings of the 2008 conference on Computational Models of Argument: Proceedings of COMMA 2008*, pages 171–182. IOS Press, 2008.
- [27] Leila Amgoud and Souhila Kaci. An argumentation framework for merging conflicting knowledge bases. *Int. J. Approx. Reasoning*, 45(2):321–340, July 2007.
- [28] David M. Kaiser. Automatic feature extraction for autonomous general game playing agents. In *Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems*, AAMAS '07, pages 93:1–93:7. ACM, 2007.



# Desarrollo de gestores de noticias incorporando técnicas basadas en la noción de confiabilidad

Cristian E. Briguez

Marcela Capobianco

Ana G. Maguitman

Laboratorio de Investigación y Desarrollo en Inteligencia Artificial  
Departamento de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur,  
Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)  
Av. Alem 1253, (B8000CPB) Bahía Blanca, Argentina  
e-mail: {ceb, mc, agm}@cs.uns.edu.ar

## Resumen

La búsqueda y lectura de noticias en los distintos medios de información se ha convertido en una de las actividades más importantes dentro de la web. Como existen numerosas opciones de fuentes de noticias disponibles, es de suma importancia proveer mecanismos inteligentes para seleccionarlas y presentarlas de manera conveniente al usuario. Un factor importante que puede considerarse al momento de seleccionar noticias es la confiabilidad. El objetivo de esta investigación consiste en seguir avanzando en el diseño e implementación de gestores de noticias, mediante el desarrollo de nuevas técnicas inteligentes de procesamiento de noticias capaces de combinar nociones de relevancia, preferencia y confiabilidad. Dichas nociones serán abordadas utilizando tanto métodos cuantitativos como cualitativos, siendo estos últimos los que permitirán realizar un seguimiento sistemático y fundamentado de las noticias que interesan a los distintos usuarios. Para esto, se combinarán técnicas clásicas de administración de noticias con mecanismos de razonamiento lógicos tales como la argumentación rebatible.

**Palabras clave:** Gestión de Noticias, Argumentación, Búsqueda Contextualizada

## Contexto

Esta línea de investigación se llevará a cabo dentro del ámbito del Laboratorio de Investigación y Desarrollo en Inteligencia Artificial (LIDIA), y está asociada a los siguientes proyectos de investigación:

- Diseño y Evaluación de Mecanismos de Búsqueda Contextualizada en Sistemas Centralizados y Distribuidos. PGI: 24/N029. Director: Ana G. Maguitman. Fuente de financiamiento: Universidad Nacional del Sur. 01/01/2011-31/12/2014.
- Soporte inteligente para el acceso a información Contextualizada en entornos centralizados y distribuidos. PIP: 11220090100863. Director: Ana G. Maguitman. Fuente de financiamiento: CONICET. 2010-2012.

## Introducción

Buscar y leer noticias en diarios y otros medios de información que se encuentran en la Web se ha vuelto una de las actividades más importantes en este ámbito. Tanto la cantidad de usuarios como el tráfico en los principales buscadores de noticias (tales como Yahoo!

News, Google News) se ha incrementado de manera notoria en los últimos años. A esto debemos sumarle que también podemos encontrar noticias en las redes sociales (tales como Facebook, Twitter), las cuales han ganado en importancia y se han vuelto masivas. La gran cantidad de noticias online y la diversidad de fuentes de información disponibles refleja la necesidad que tienen los usuarios de estar informados y de obtener información y opiniones pluralistas.

La cantidad de fuentes de noticias disponibles en la Web es muy numerosa, por esto es de gran importancia proveer mecanismos inteligentes para seleccionarlas, combinarlas y presentarlas de manera conveniente al usuario. Los mayores esfuerzos de investigación en esta área se han concentrado en la *clasificación de noticias* de acuerdo a contenido de las noticias y el perfil de los usuarios. Sin embargo, debido a que las noticias publicadas en los periódicos no pueden ser verificadas en su totalidad, un factor importante que puede considerarse al momento de seleccionar noticias es la *confiabilidad*.

Existen aspectos importantes que deben ser tenidos en cuenta al momento de desarrollar un modelo realista de la noción de confiabilidad:

- La confiabilidad debe ser **justificada**.
- La confiabilidad es **rebatible**.
- La confiabilidad es **subjetiva**.

Los meta-buscadores de noticias generalmente pasan por alto la dinámica de la confiabilidad de noticias, o si es tenida en cuenta, la misma ha sido estudiada mediante técnicas puramente cuantitativas (*e. g.* [24]). Estas técnicas son limitadas en varios sentidos ya que dificultan la construcción de justificaciones para sostener la credibilidad de ciertas noticias. Dado que las técnicas cuantitativas no están dotadas de capacidad inferencial, mucha de la información

implícita no es revelada. Estas técnicas, por otra parte, son incapaces de tratar de manera formal con la naturaleza rebatible de la noción de confiabilidad.

Como ha ocurrido con otros aspectos de la Web, la gestión de noticias se ha vuelto una tarea altamente colaborativa. Sin embargo, la noción de confiabilidad es subjetiva y por lo tanto los modelos que tratan con dicha noción de manera objetiva no son totalmente realistas. En este caso, opiniones individuales de lectores confiables son más valiosas que las opiniones que provienen de agregar opiniones de un pool de lectores desconocidos.

En base a lo expuesto arriba, un paso importante para avanzar en el área del diseño e implementación de gestores de noticias es el desarrollo de nuevas técnicas inteligentes capaces de combinar nociones de relevancia, preferencia y confiabilidad. El trabajo propuesto consiste en la creación de un servicio de gestión de noticias que integre dichas técnicas y nos permita ver su comportamiento en distintas comunidades del mundo real.

## Líneas de investigación y desarrollo

Esta línea de investigación se centrará sobre el área del diseño e implementación de gestores de noticias, mas concretamente en el desarrollo un sistema buscador y recomendador que integre nuevas técnicas de procesamiento de noticias creadas teniendo en cuenta la combinación de ciertas nociones, como relevancia, preferencia y confiabilidad.

En la actualidad, existe una gran cantidad de buscadores de noticias comerciales entre los que podemos destacar Google News [1], Yahoo! News [3], etc. Aunque ninguno de ellos da a conocer mediante publicaciones la manera en que ordenan las noticias, es evidente que toman en cuenta factores tales como la novedad, la fuen-

te a la que pertenece el artículo y la cantidad de veces que se repite en diferentes medios.

Una revisión de la literatura actual nos permite identificar una serie de desarrollos en el área de buscadores de noticias. NewsInEssence [27] es un sistema que busca y agrupa en clusters noticias relacionadas. QCS [12] es una herramienta que hace más eficiente la tarea de agrupar y categorizar los documentos en tópicos. En NewsJunkie [13] se propone y analiza un sistema que personaliza las noticias identificando las que son primicia en el contexto del usuario. Velthune [16] es un motor de búsqueda de noticias que extrae información tanto de la Web como de News feeds. Otro sistema, Compare&Contrast [21] es un sistema de razonamiento basado en casos que utiliza la Web como base de conocimiento para descubrir casos comparables de noticias.

Muchas técnicas han sido propuestas para organizar las noticias en tópicos. La mayoría tiene como objetivo reconocer, resumir y realizar un seguimiento de ciertos episodios (*e. g.* [25, 23, 20, 4]).

Otras propuestas han abordado los problemas de *credibilidad* y *sesgo* en los medios. En [24] se propone un método para evaluar la credibilidad de los artículos de noticias basado en la comparación de los contenidos de fuentes diversas. El servicio de noticias PolyNews [26] tiene como objetivo mitigar el efecto del sesgo introducido por diversos medios mediante la creación de múltiples puntos de vista. NewsTrust [2] es un servicio donde los usuarios pueden evaluar noticias, a sus redactores y a sus fuentes. Además, caben destacarse algunos modelos para el análisis de la propagación de la confiabilidad (*e. g.* [15, 17]) así como técnicas orientadas a analizar la diversidad y el sesgo de noticias de internet (*e. g.* [22]).

Otro área de especial interés se orienta a la detección de noticias de alto impacto (*e. g.* [18, 29]). Otro área de interés se enfoca en la representación de la reputación de los usuarios del

sistema como podemos ver en HARM [19].

En los últimos años la *argumentación rebatible* [11] evolucionó para convertirse en un mecanismo exitoso para modelar el razonamiento de sentido común cualitativo. En este sentido la *programación en lógica rebatible* (PLR) [14] y sus especializaciones la *programación en lógica rebatible posibilística* (PLRP) [5, 8] y la *programación en lógica rebatible con observaciones* (PLRO) [7]. Estos sistemas de razonamiento se han aplicado con éxito a diversas áreas, como *teorías de agencia* [6], *ingeniería de conocimiento* [9], *clustering* [10], y *negociación* [28]. En particular la PLRP fue diseñada para integrar el manejo de la incertidumbre en un sistema argumentativo. En trabajos posteriores [8] este sistema fue redefinido y optimizado para permitir incorporar percepciones y usar un mecanismo de optimización basado en el uso de conocimiento precompilado. Mediante este mecanismo el sistema podrá cumplir con los requerimientos de tiempo real necesarios para modelar el razonamiento en ambientes dinámicos.

La argumentación rebatible ha sido ya usada en ArgueNet [10] para diseñar un sistema de recomendaciones que clasifique los resultados de búsqueda de acuerdo a un conjunto de criterios de preferencia especificados por el usuario. Este acercamiento permite integrar en forma novedosa un motor de búsqueda tradicional con un framework argumentativo. El presente trabajo proyecta continuar esta línea de investigación, con especial énfasis en el tratamiento cualitativo de la noción de confiabilidad.

## Resultados y Objetivos

Esta línea de investigación tiene por objetivo el desarrollo de un gestor de noticias que incorpore técnicas inteligentes capaces de combinar las nociones de relevancia, preferencia y confiabilidad. El trabajo a realizar consiste en

diseñar técnicas de procesamiento de noticias capaces de realizar un manejo cualitativo de dichas nociones. El objetivo específico es proponer un servicio que facilite el manejo cualitativo de la información. La meta es agregar un nivel adicional a los gestores de noticias básicos, donde se aplicarán nuevas heurísticas, se integrarán noticias basadas en fuentes antagónicas y donde será posible fundamentar posiciones contrapuestas. Esto nos permitirá integrar un sistema de recomendación de noticias con sistemas dialécticos, lo cual facilitará un seguimiento más sostenido y amplio de las noticias de interés para el usuario.

## Formación de Recursos Humanos

Actualmente el equipo de trabajo de esta línea de investigación se encuentra compuesto por un becario de posgrado que accedió a una beca interna del conicet, su directora y su codirectora. Por otra parte se vincula con un grupo de trabajo sobre minería de datos compuesto por doctorandos e investigadores formados.

## Referencias

- [1] <http://news.google.com/>.
- [2] <http://newstrust.net/>.
- [3] <http://news.yahoo.com/>.
- [4] James Allan, Rahul Gupta, and Vikas Khandelwal. Temporal summaries of news topics. In *SIGIR*, pages 10–18, 2001.
- [5] Teresa Alsinet, Carlos I. Chesñevar Lluís Godo, and Guillermo R. Simari. A logic programming framework for possibilistic argumentation: Formalization and logical properties. *Fuzzy Sets and Systems*, 159(10):208–228, 2008.
- [6] Daniel Bryant and Paul J. Krause. An implementation of a lightweight argumentation engine for agent applications. In *JELIA*, pages 469–472, 2006.
- [7] Marcela Capobianco, Carlos I. Chesñevar, and Guillermo R. Simari. Argumentation and the dynamics of warranted beliefs in changing environments. *Autonomous Agents and Multi-Agent Systems*, 11(2):127–151, 2005.
- [8] Marcela Capobianco and Guillermo R. Simari. A proposal for making argumentation computationally capable of handling large repositories of uncertain data. In *SUM*, pages 95–110, 2009.
- [9] Daniela V. Carbogim, David Robertson, and John Lee. Argument-based applications to knowledge engineering. *The Knowledge Engineering Review*, 15(2):119–149, 2000.
- [10] Carlos I. Chesñevar and Ana G. Maguitman. Arguenet: An argument-based recommender system for solving web search queries. In *Proc. of the 2nd IEEE Intl. IS-2004 Conference*, pages 282–287, 2004.
- [11] Carlos I. Chesñevar, Ana G. Maguitman, and Ronald P. Loui. Logical models of argument. *ACM Computing Surveys*, 32:337–383, 2000.
- [12] Daniel M. Dunlavy, John M. Conroy, and Dianne P. O’Leary. Qcs: A tool for querying, clustering, and summarizing documents. In *HLT-NAACL*, pages 11–12, 2003.
- [13] Evgeniy Gabrilovich, Susan T. Dumais, and Eric Horvitz. Newsjunkie: providing personalized newsfeeds via analysis of information novelty. In *WWW*, pages 482–490, 2004.



- [14] Alejandro J. García and Guillermo R. Simari. Defeasible logic programming: An argumentative approach. *Theory and Practice of Logic Programming*, 4(1):95–138, 2004.
- [15] Ramanathan V. Guha, Ravi Kumar, Prabhakar Raghavan, and Andrew Tomkins. Propagation of trust and distrust. pages 403–412, 2004.
- [16] Antonio Gulli. The anatomy of a news search engine. In *WWW (Special interest tracks and posters)*, pages 880–881, 2005.
- [17] Claudia Hess. Trust-based recommendations for publications: A multi-layer network approach. *TCDL Bulletin*, 2(2), 2006.
- [18] Yunliang Jiang, Cindy Xide Lin, and Qiaozhu Mei. Context comparison of bursty events in web search and online media. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, EMNLP '10*, pages 1077–1087, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics.
- [19] Kalliopi Kravari and Nick Bassiliades. Harm: A hybrid rule-based agent reputation model based on temporal defeasible logic. In *RuleML*, pages 193–207, 2012.
- [20] Andrew J. Kurtz and Javed Mostafa. Topic detection and interest tracking in a dynamic online news source. In *JCDL*, pages 122–124, 2003.
- [21] Jiahui Liu, Earl Wagner, and Larry Birnbaum. Compare&contrast: using the web to discover comparable cases for news stories. In *WWW*, pages 541–550, 2007.
- [22] Qiang Ma and Masatoshi Yoshikawa. Topic and viewpoint extraction for diversity and bias analysis of news contents. In *Proceedings of the Joint International Conferences on Advances in Data and Web Management, APWeb/WAIM '09*, pages 150–161, Berlin, Heidelberg, 2009. Springer-Verlag.
- [23] Masaki Mori, Takao Miura, and Isamu Shioya. Topic detection and tracking for news web pages. In *Web Intelligence*, pages 338–342, 2006.
- [24] Ryosuke Nagura, Yohei Seki, Noriko Kando, and Masaki Aono. A method of rating the credibility of news documents on the web. In *SIGIR*, pages 683–684, 2006.
- [25] Ramesh Nallapati, Ao Feng, Fuchun Peng, and James Allan. Event threading within news topics. In *CIKM*, pages 446–453, 2004.
- [26] Souneil Park, Seungwoo Kang, and Junehwa Song. Polynews: Delivering multiple aspects of news to mitigate media bias. Technical Report KAIST, 2006.
- [27] Dragomir R. Radev, Jahna Otterbacher, Adam Winkel, and Sasha Blair-Goldensohn. Newsinsence: summarizing online news topics. *Commun. ACM*, 48(10):95–98, 2005.
- [28] Iyad Rahwan, Sarvapali D. Ramchurn, Nicholas R. Jennings, Peter McBurney, Simon Parsons, and Liz Sonenberg. Argumentation-based negotiation. *The Knowledge Engineering Review*, 18(3):343–375, 2003.
- [29] Aaron Sun, Daniel Dajun Zeng, and Hsin-chun Chen. Burst detection from multiple data streams: a network-based approach. *Trans. Sys. Man Cyber Part C*, 40:258–267, May 2010.

## Propuesta preliminar de Sistema Experto para determinar el método de Ensayo No Destructivo aplicado a una Central Hidroeléctrica

Javier H. Lafont<sup>1,2</sup>, Marisa Panizzi<sup>1</sup>, Iris Sattolo<sup>1</sup>, Carlos Desimone<sup>2</sup>, Claudio Ziobrowsky<sup>2</sup>,  
Alejandro García<sup>2</sup>, Alba Obrutsky<sup>2</sup>, Jorge Méndez<sup>2</sup>, Daniel Acosta<sup>2</sup>

<sup>1</sup> Facultad de Informática, Ciencias de la Comunicación y Técnicas Especiales, Universidad de Morón  
Cabildo 134 – CP(1708) – Morón – Prov. de Bs. As.  
Tel: 5627-2000

<sup>2</sup> Departamento de Ensayos No Destructivos y Estructurales, Comisión Nacional de Energía Atómica,  
Centro Atómico Constituyentes.  
Av. Gral. Paz 1499 – CP(1650) – San Martín – Prov. de Bs. As.  
Tel: 6772-7470

[lafontjavier@hotmail.com](mailto:lafontjavier@hotmail.com), [marisapanizzi@speedy.com.ar](mailto:marisapanizzi@speedy.com.ar), [iris.sattolo@gmail.com](mailto:iris.sattolo@gmail.com), [desimone@cnea.gov.ar](mailto:desimone@cnea.gov.ar),  
[ziobrowsky@cnea.gov.ar](mailto:ziobrowsky@cnea.gov.ar), [aldarcia@cnea.gov.ar](mailto:aldarcia@cnea.gov.ar), [obrutsky@cnea.gov.ar](mailto:obrutsky@cnea.gov.ar), [mendez@cnea.gov.ar](mailto:mendez@cnea.gov.ar),  
[dacosta@cnea.gov.ar](mailto:dacosta@cnea.gov.ar).

### Resumen.

El propósito de esta línea de investigación consiste en la conceptualización de un Sistema Experto que permite determinar que método de Ensayo No Destructivo es el indicado para ser aplicado en una Central Hidroeléctrica. Se empleará para ello la metodología I.D.E.A.L., de la Ingeniería del Conocimiento. En este estadio del trabajo de investigación se realizó la definición del dominio, el cálculo de viabilidad de la propuesta y la adquisición del conocimiento de los expertos. Luego se avanzará con la conceptualización, formalización, y con los casos de prueba. Este trabajo aportará a los usuarios una herramienta que determinara el método y el tipo de técnica de Ensayo No Destructivo a emplear. En el dominio del SE, se consideran factores como las características del material, las etapas del procesos de fabricación, de terminación, de transformación y de servicio; como también, las características de las discontinuidades que deseamos detectar en el material.

### Palabras clave:

*Sistema Experto/ Ingeniería del Conocimiento/ método de Ensayo No Destructivo/ técnica de ensayo.*

### Contexto.

Esta línea de investigación, tuvo su origen debido a la falta de un sistema informático que pudiera determinar el método de ensayo no destructivo. La solución propuesta se encuentra enmarcada en la Ingeniería del Conocimiento, más específicamente en los sistemas expertos. La propuesta del sistema experto podrá ser aplicada en diferentes campos, como por ejemplo, en una Central Hidroeléctrica véase *Figura Nro.1*, en una Central Nuclear, en una Central Térmica o en una fábrica industrial.

Los expertos con los que se trabajará se encuentran en la Comisión Nacional de Energía Atómica, Centro Atómico Constituyentes, en el Departamento Ensayos No Destructivos y Estructurales.

### Introducción.

Los ensayos no destructivos se han practicado por muchas décadas. Se tiene registro de las mismas desde 1868, cuando se comenzó a trabajar con campos magnéticos. El método de prueba no destructiva original, y más antiguo, es la *inspección visual (IV)*. (Báez, 1985).



**Figura Nro. 1.** Imagen de la represa (Central Hidráulica).

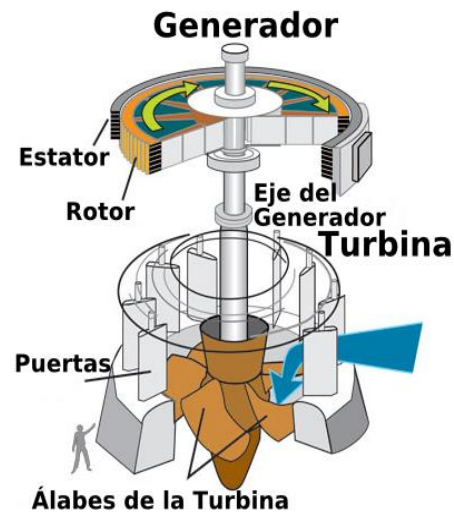
Los **Ensayos no Destructivos END** (castellano) Non Destructive Testing NDT (inglés), son herramientas fundamentales y esenciales para el control de calidad de materiales de ingeniería, procesos de manufactura, confiabilidad de productos en servicio y mantenimiento, cuya falla prematura puede ser muy costosa.

Se definen como aquellos métodos de ensayo utilizados para examinar o inspeccionar un material o un sistema sin impedir la utilidad futura del mismo. (Alba Obrutsky, 2006).

Se plantea un objeto que se va a ensayar, que en este caso es una unidad de turbina, véase *Figura Nro.2*, compuesta por cinco álabes identificada con una letra, y posee dos lados de inspección “aguas arriba” y “aguas abajo”.

Investigan específicamente la integridad material del objeto ensayado para permitir establecer el desempeño futuro del mismo.

Son de fundamental importancia para la detección, análisis y evaluación de discontinuidades, defectos y para la caracterización de materiales.



**Figura Nro. 2.** Unidad de turbina.

Están orientados al diagnóstico de componentes, sistemas y estructuras de instalaciones, con el propósito de garantizar su operación segura y confiable, sin alterar ni su integridad ni sus propiedades.

Los Métodos de Ensayos No Destructivos son:

1. Inspección Visual.
2. Líquidos Penetrantes.
3. Partículas Magnéticas.
4. Ultrasonido (*véase Figura Nro.3*).
5. Termografía Infrarroja.
6. Radiografía Industrial.
7. Corrientes Inducidas.
8. Pruebas de Fuga.
9. Emisión Acústica.



**Figura Nro. 3.** Inspección por Ultrasonido.

Para la construcción del dominio de la aplicación se revisaron los conceptos de los métodos de END, con sus respectivas técnicas (Asociación Española de END, Partículas Magnéticas – Ultrasonido - Líquidos Penetrantes, 2006) y la estructura de la unidad de turbina de una central hidroeléctrica (Luis Antonio Restrepo Arango, 2007).

En la actualidad esta tarea es realizada manualmente por una persona experta en END (Ensayos No Destructivos) con una calificación de nivel dos de cada uno de los métodos de END y puede llegar a demorarse entre 24 a 48 horas.

Para poder evaluar la conveniencia de resolver el problema con la construcción del prototipo, se llevó a cabo el estudio de viabilidad siguiendo los lineamientos propuesto en la metodología I.D.E.A.L. (García Martínez, 2004). Se analizaron las siguientes dimensiones: éxito, justificación, plausibilidad y adecuación. Los resultados de los cálculos, se presentan en la *Tabla Nro.1*.

Dimensión	Peso	Valores Intervalo				Peso*Valor			
Plausibilidad	8	8.78	9.15	7.83	8.291	70.232	73.18	62.672	66.328
Justificación	3	7.8	8.8	10	10	23.4	26.4	30	30
Adecuación	8	6.94	7.9	7.19	8.184	55.496	63.17	57.488	65.472
Éxito	5	6.56	7.62	8.93	9.899	32.78	38.12	44.655	49.495
	24					181.908	200.9	194.82	211.3

Intervalo Resultado Final: **7.5795 8.369 8.117 8.804**

**RESULTADO FINAL: 8.2**

**Tabla Nro. 1.** Estudio de viabilidad realizado para el SE para determinar el método de END aplicado a una Central Hidroeléctrica

El cálculo final del estudio dio como resultado final **8.2**, indicando de esta manera que era posible llevar a cabo el desarrollo del mismo.

El domino del conocimiento, se construye a través de la aplicación de entrevistas del tipo estructurado,

propuestas en el ciclo de educación detallado propuesto por la metodología I.D.E.A.L. (García Martínez, 2004).

Esta adquisición tuvo dos facetas de extracción, la de los conocimientos públicos en la cual se revisaron los conceptos de ensayos no destructivos (Obrutsky Alba, 2006), ensayos no destructivos de líquidos penetrantes (AEND - Asociación Española de Ensayos No Destructivos, 2006), ensayos no destructivos, partículas magnéticas (AEND - Asociación Española de Ensayos No Destructivos, 2006), y ensayos no destructivos ultrasonido (AEND - Asociación Española de Ensayos No Destructivos, 2006).

Se revisaron la propuesta de ensayos no destructivos de Báez (Báez, 1985), antecedentes de ultrasonido de Echeverría (Echeverría, 2002) y cuestiones asociadas a la energía hidroeléctrica (Restrepo Arango, 2003).

La otra faceta, la educación de los conocimientos privados de los expertos, que se alterna cíclicamente con la etapa de conceptualización para poder modelizar el comportamiento del experto.

Para la educación del conocimiento privado se trabajó con los siguientes expertos:

**Ing. Carlos Desimone**  
(Ingeniero Mecánico).

**Mg. Ing. Claudio Ziobrowsky**  
(Magister en Materiales,  
Ingeniero Metalúrgico).

**Ing. Alejandro García**  
(Ingeniero Electrónico).

**Ing. Alba Obrutsky**  
(Ingeniera Metalúrgica).



**Ing. Jorge Méndez**  
(Ingeniero Metalúrgico).

**Ing. Daniel Acosta**  
(Ingeniero Electromecánico).

Se está analizando que herramienta de desarrollo se utilizará para poder realizar el prototipo. Estas herramientas pueden ser: Kappa-PC, Lisp, Shells o APL y coinciden con lo propuesto en la metodología I.D.E.A.L. (García Martínez, 2004)

En esta propuesta de trabajo de los nueve métodos de END, se desarrollaron tres, de los cuales son:

1. *Líquidos Penetrantes.*
2. *Partículas Magnéticas.*
3. *Ultrasonido.*

A su vez, cada uno de estos métodos están formados por distintas tipos de técnicas de END, de los cuales son:

1. *Líquidos Penetrantes.*
  - *Líquidos Coloreados.*
  - *Líquidos Fluorescentes.*
2. *Partículas Magnéticas.*
  - *Vía Seca Coloreado.*
  - *Vía Seca Fluorescente.*
  - *Vía Húmeda Coloreado.*
  - *Vía Húmeda Fluorescentes*
3. *Ultrasonido.*
  - *Pulso Eco*

Para que el Sistema Experto devuelva como solución el Método y su tipo de técnica de END, el mismo tendrá que realizar los pasos que se encuentran en la *Tabla Nro. 2.*

1. Identificar el material	2. Identificar las etapas del proceso	3. Clasificar las discontinuidades a detectar
1.1. Identificar la composición química.	2.1. Identificar el proceso de fabricación.	3.1. Clasificar la forma.
1.2. Identificar la geometría.	2.2. Identificar el proceso de terminación.	3.2. Clasificar la posición.
1.3. Identificar material según tipo metálico.	2.3. Identificar el proceso de transformación.	3.3. Clasificar las dimensiones.
1.4. Identificar las limitaciones en el campo de observación.	2.4. Identificar el proceso de servicio.	
1.5. Identificar la calidad superficial.		

**Tabla Nro. 2.** Etapas de evaluación del material para la selección del método de END.

### Resultados Obtenidos/ Esperados.

Para avanzar con el objetivo propuesto en esta línea de investigación, se analizaron diferentes metodologías para la construcción de Sistemas Expertos, optando por trabajar con la metodología I.D.E.A.L.

Actualmente se está trabajando en la construcción del prototipo.

Se espera contar con el prototipo finalizado y los casos de prueba necesarios a modo de ser presentado en CACIC 2013.

De esta línea de investigación se desprende la posibilidad de completar los métodos de ensayo que quedaron fuera del alcance de esta solución.

### Formación de Recursos Humanos.

El equipo está integrado por un tesista que se encuentra desarrollando su trabajo de fin de carrera tutorado por una docente de la cátedra de Trabajo de Diploma (asignatura de fin de carrera) y una ayudante de cátedra.

Se pretende realizar la transferencia de esta línea de investigación a futuros tesistas para que completen el alcance del sistema experto.

**Bibliografía.**

AEND - Asociación Española de END, V.: Ensayos No Destructivos, Líquidos Penetrantes. Nivel II. (2006).

AEND - Asociación Española de END, V.: Ensayos No Destructivos, Partículas Magnéticas. Nivel II. (2006).

AEND - Asociación Española de END, V.: Ensayos No Destructivos, Ultrasonido. Nivel II. (2006).

Askeland, Donald R., V.: La ciencia e ingeniería de los materiales. Universidad de Missouri – Rolla – Estados Unidos. (1998).

Ing. Agüero, Alva Hugo L., V.: Metrología. Ensayos No Destructivos (2008).

Báez., V.: Introducción a los Ensayos No Destructivos. CNEA – CAC – UA ENDE – (1985).

Di Caprio, Gabriele., V.: Los aceros inoxidables. Grupinox – Barcelona (1999).

Ing. Echevarría, Ricardo., V.: Ultrasonido. Universidad Nacional del Comahue – Facultad de Ingeniería (2002).

García Martínez, R. y Britos, P., V.: Ingeniería de Sistemas Expertos. Nueva Librería, Buenos Aires – Argentina (2004).

Ing. Laufgang, Sergio G., V.: Aceros Inoxidables. Facultad de ingeniería de la Universidad de Buenos Aires. (2003).

Montero, Ricardo., V.: Replicas Metalográficas. CNEA – CAC – UA ENDE – Instituto Sábato – UNSAM (2005).

NIDI (Nickel Development Institute), V.: Manual del Acero Inoxidable. Serie No 1. Universidad del Zulia, R. C. (2002).

Restrepo Arango, Luis Antonio., V.: La Energía Hidroeléctrica. Universidad Nacional de Colombia – Medellín (2007).

Obrutsky, Alba., V.: Introducción a los Ensayos No Destructivos. CNEA – CAC – UA ENDE – Instituto Sábato – UNSAM (2006).

# Estudio de soluciones tecnológicas para el desarrollo de un modelo factible de participación ciudadana

Ing. COLOMBO, Humberto Luis - Ing. ANTONINI, Sergio Andrés  
Ing. CHONG ARIAS, Carlos Daniel - Ing. ISTVAN, Romina Mariel  
Ing. PATERNOSTER, Felix Miguel

Laboratorio de Ingeniería en Sistemas, LINES - FRLP- UTN

124 y 60 S/N La Plata – Prov. Buenos Aires

(0221) 421-7578 int. 2254 – int. 2255

[antonini@frlp.utn.edu.ar](mailto:antonini@frlp.utn.edu.ar)

[ristvan@frlp.utn.edu.ar](mailto:ristvan@frlp.utn.edu.ar)

## Resumen

Particularmente en Argentina, y a nivel social, el sistema tradicional de votación es concebido como un símbolo de transparencia y veracidad del escrutinio. En este sentido, los sistemas que no utilizan la boleta electoral eliminan una parte de la legitimidad del proceso y su fundamento social.

Por este motivo, se propone el diseño e implementación de un sistema de votación mediante boleta única, que permita automatizar una gran parte del proceso electoral como signo de modernidad y seguridad, respondiendo al reto tecnológico y manteniendo, a su vez, la boleta como único soporte formal y real para comprobación de resultados electorales.

**Palabras clave:** voto electrónico, votación electrónica, urna electrónica, boleta única.

## Contexto

Desde el año 2003 LINES ha analizado los métodos de votación existentes, estudiando en cada caso las ventajas y desventajas que ofrecen respecto a cinco puntos esenciales:

- autenticación del votante
- presentación de opciones electorales
- selección de candidatos por parte del votante
- proceso de votación (emisión del voto)
- recuento y totalización de resultados electorales.

Se tomaron casos de estudio, que fueron implementados a nivel nacional, en varias provincias del país, como así también a nivel internacional, en países como España y Estados Unidos.

De ellos se evaluó, además de las características nombradas la velocidad en el acto eleccionario, la eliminación de las boletas papel, la rapidez en el recuento de los votos y la posibilidad de transmisión de los resultados a un centro totalizador.

De este análisis surge la propuesta de mejora del sistema de votación por uno de boleta única, que permita automatizar una gran parte del proceso electoral como signo de modernidad y seguridad respondiendo al reto tecnológico, pero manteniendo la boleta, como único soporte formal y real para comprobar los resultados electorales.

Este análisis, además, se complementa con la experiencia de los integrantes como:

- auditores en pruebas pilotos que se desarrollaron en el año 2003 en las elecciones provinciales con urnas electrónicas brasileras
- veedores en el año 2004, en las elecciones del Centro de Estudiantes y Claustros de la Facultad de Derecho de la UNLP
- diseñadores y desarrolladores de un sistema de votación con urnas electrónicas; que participaron en las elecciones de la provincia de Buenos Aires, en el año 2007 y luego en el Presupuesto Participativo del Municipio de La Plata en los años 2008 y 2009, Presupuesto Participativo del Municipio de Morón 2009, Elecciones de Delegados Municipales y Consejeros Escolares del Municipio de Mar del Plata, 2009
- experiencias con la Junta Electoral de la provincia de Buenos Aires.

## Introducción

Existen básicamente tres tipos de modalidades de votación, que varían según las ventajas que ofrecen respecto a varios puntos esenciales: autenticación del votante, selección de candidatos por parte del votante, proceso de votación propiamente dicho y gestión y procesamiento del contenido de la urna electoral que incluye recuento de

resultados garantizando seguridad y transparencia [1].

Los sistemas de votación actuales a nivel mundial son [2]:

1. **Sistemas de voto tradicional:** implican la movilización del ciudadano al lugar de votación. La selección del candidato depende de la existencia de la totalidad de las boletas y del tipo de elección. Si la misma implica cortes de boleta, la emisión y el recuento de los votos se realiza meramente de manera manual. Como ventaja presenta la constancia de la emisión del voto en papel.

2. **Sistemas de voto electrónico:** en este tipo de sistemas se combinan los procedimientos tradicionales con el uso de urnas electrónicas que facilitan la identificación de los electores, selección de alternativas de voto y la transferencia de resultados a una base de datos centralizada para facilitar el recuento de votos totales de la elección.

Mantiene un aspecto fundamental del sistema tradicional que es la convergencia de los ciudadanos a los lugares de votación. Esta modalidad se fue implantando en varios países con variantes que van desde urnas lectoras de boletas como la urna electrónica Demotek en el País Vasco en España que se basa en el uso de boletas tradicionales e incorpora lectores ópticos o dispositivos especiales de selección de alternativas sin uso de boletas como el sistema e-slate desarrollado por la empresa Hart Inter Civic en Estados Unidos. También el uso de pantallas táctiles desarrollado por la empresa Diebold y utilizado en Brasil a partir de las elecciones municipales del año 2000 a nivel nacional.

Hasta ahora en los países en que se ha utilizado o puesto a prueba la modalidad de voto electrónico los resultados indican que es la alternativa tecnológica de más



fácil implantación ya que se ha comprobado en casi todos los casos su viabilidad técnica, política y social.

3. Sistemas de voto vía internet: tiene un grado de complejidad superior a las dos modalidades anteriores pero tiende relativamente a la facilidad en cuanto a su acceso y uso para el ciudadano común. Comprende el uso total de tecnología avanzada en computación y aplicaciones web, además de tecnología criptográfica avanzada. Posibilita el registro en una base de electores y el acceso al programa de votación desde cualquier terminal, protocolos de seguridad y conocimiento de resultados inmediatos.

Las ventajas fundamentales de este esquema es la reducción de costos a largo plazo, la rapidez con la que se procede el recuento de votos y la comodidad para el votante.

No todos los sistemas de votación electrónica utilizan boletas electorales, esto tiende a eliminar el mecanismo tradicional de verificación de los resultados en caso de una falla o intento de fraude; con lo cual no quedaría la posibilidad de comprobar la situación de la elección a posteriori.

Así vemos como los sistemas de voto electrónico presentan importantes retos en tres áreas principales: tecnológica, social y legislativa [3].

Particularmente en Argentina, y a nivel social, el sistema tradicional de votación es concebido como un símbolo de transparencia y veracidad del escrutinio. En este sentido, los sistemas que no utilizan la boleta electoral eliminan una parte de la legitimidad del proceso y su fundamento social. Por este motivo, se propone un nuevo sistema de votación que permita automatizar una gran parte del proceso electoral.

## Líneas de investigación y Desarrollo

La UTN-FRLP comienza la investigación sobre el voto electrónico en el año 2003. Momento en el cual, Gobierno Electrónico de la Provincia de Buenos Aires la convoca para realizar una auditoría en la prueba piloto que se desarrolló en las elecciones provinciales con urnas electrónicas brasileñas.

En el año 2004, participa como veedor en las elecciones del Centro de Estudiantes y Claustros de la Facultad de Derecho de la UNLP, las que se realizaron con urnas electrónicas españolas.

Al año siguiente, ingenieros viajan a España para profundizar en el estado de arte del voto electrónico y firma digital, mediante el programa de intercambio "Red Alfa" del Gobierno Electrónico.

En el mes de abril de 2007, da comienzo al armado del hardware y software de un sistema de votación con urnas electrónicas. Basado en la experiencia individual y grupal de los integrantes del proyecto, se diseña una urna 100% nacional, que se implanta con éxito en 52 mesas para las elecciones provinciales del 28 de octubre de 2007.

El hardware consiste de una urna de votación con teclado, donde el elector puede seleccionar el número de lista que desea, un teclado para el presidente de mesa, una UPS y una impresora.

El software visualiza las listas de candidatos por categorías presentando todas las opciones electorales, permitiendo al votante seleccionar la opción deseada, pudiendo realizar de manera rápida, cambios antes de cerrar su votación.

Los datos se encuentran encriptados y se almacenan por duplicado en dos memorias.

Desde esa fecha se han utilizado en varias elecciones, entre las que se destacan Presupuesto Participativo del Municipio de La Plata (2008 y 2009), Presupuesto Participativo del Municipio de Morón (2009), Elecciones de Delegados Municipales y Consejeros Escolares del Municipio de Mar del plata (2009).

En el año 2012 la UTN-FRLP, da comienzo a una nueva etapa en investigación de sistemas de votaciones electrónicas, con el fin de mejorar el sistema existente.

## Resultados y Objetivos

Como objetivo se espera la mejora del sistema de votación por boleta única, que permita automatizar una gran parte del proceso electoral como signo de modernidad y seguridad respondiendo al reto tecnológico, pero manteniendo la boleta, como soporte formal y real para comprobar los resultados electorales.

Además se espera que los beneficios principales que provea sean la transparencia en la obtención de los resultados y la rapidez en que se totalizarán los mismos.

Se proyecta su fácil adaptación a elecciones con o sin padrón, tanto para votaciones municipales, provinciales y/o nacionales, o cualquier otro tipo de elección incluyendo las encuestas.

## Formación de Recursos Humanos

El equipo de investigación se encuentra dirigido por el Ing. COLOMBO Humberto

Luis de la UTN FRM (director) y el Ing. ANTONINI Sergio Andrés (codirector).

Asimismo cuenta con un grupo de profesionales docentes que participan de las tareas de coordinación:

- Ing. CHONG ARIAS, Carlos Daniel
- Ing. ISTVAN, Romina Mariel
- Ing. PATERNOSTER, Felix Miguel

Y un número de becarios que desarrollan sus tareas de formación en investigación:

- BIFANO, Lautaro
- FALCO, Mariana
- FERNANDEZ, Pablo
- FRIELLO, María Antonella
- IRURETA, Mariano
- LUPI, Francisco Martin
- PEREZ, Ezequiel

## Referencias

[1] ALEUY Mahmud, La votación electrónica, 2007.

[2] BARRIENTOS DEL MONTE Fernando, El voto electrónico: contexto, experiencias y dilemas, 2006.

[3] MORALES ROCHA Victor Manuel, Seguridad en los Procesos de Voto Electrónico Remoto, 2009.

[4] TULA Maria Ines, Voto electronico entre votos y maquinas, 2005.

[5] GONZÁLEZ DE LA GARZA, Voto electronico por internet, constitucion y riesgos para la democracia, 2008.

[6] COLOMER Josep, Como Votamos Los sistemas electora del mundo: pasado presente y futuro, 2004.

[7] PRINCE Alejandro, Consideraciones, aportes y experiencias para el Voto electrónico en Argentina, 2005.

[8] RENIU I VILAMALA Josep,  
Oportunidades estratégicas para la  
implementación del voto electrónico  
remoto, 2006.

## Desarrollo de Sistema Experto Generador de Recomendaciones para la Implementación Sistemas Informáticos de Gestión

Estayno, Marcelo G., Panizzi, Marisa D. , Arenas, Diego O. , Giménez, Damián J.

Facultad de Informática, Ciencias de la Comunicación y Técnicas Especiales, Universidad de Morón  
Cabildo 134 – Morón (CP 1708)  
Tel: 5627-2000

[mestayno@gmail.com](mailto:mestayno@gmail.com); [marisapanizzi@speedy.com.ar](mailto:marisapanizzi@speedy.com.ar);  
[diegoarenas@gmail.com](mailto:diegoarenas@gmail.com) ; [damianjesusum@gmail.com](mailto:damianjesusum@gmail.com)

**Resumen:** El presente desarrollo del sistema experto se basa en el trabajo de investigación en el cual se formularon desarrollo un conjunto de recomendaciones para la implementación de sistemas informáticos de gestión. Se empleará para ello la metodología I.D.E.A.L<sup>1</sup> de la Ingeniería del Conocimiento. En este estadio de la investigación se trabajó con la definición del dominio, el cálculo de viabilidad de la propuesta y la adquisición del conocimiento de los expertos. Luego se avanzará con la conceptualización, formalización, implementación y validación de la herramienta. Este trabajo aportará a los profesionales de la Ingeniería de Software una herramienta que les dará soporte para la generación de recomendaciones para la implementación de sistemas informáticos (de gestión empresarial en base a la medición de satisfacción de las necesidades de los usuarios de un *sistema informático de gestión*).

### Palabras clave:

*Sistema Experto / Ingeniería del Conocimiento/ Implementación de sistemas informáticos / Satisfacción de los usuarios.*

<sup>1</sup> Acrónimo de las fases que comporta: Identificación de la tarea, Desarrollo de los prototipos, Ejecución de la construcción del sistema integrado, actuación para conseguir el mantenimiento perfecto y lograr una adecuada transferencia tecnológica.

### Contexto.

La línea de investigación tiene su origen a partir del conocimiento de los expertos (Ing. Estayno y Mag. Panizzi) para la generación de recomendaciones para la implementación de sistemas informáticos. Esta línea comenzó a gestarse bajo la dirección del Ing. Marcelo Estayno, la cual se encuentra enmarcada en el área sistemas de información-ingeniería de software, más específicamente en la problemática de la implementación de sistemas de informáticos.

La misma se desarrolla en la Facultad de Informática, Ciencias de la Comunicación y Técnicas Especiales de la Universidad de Morón y tiene como objetivo desarrollar un prototipo de sistema experto que permita generar las recomendaciones para la implementación de sistemas informáticos.

### Introducción.

Los usuarios directos del mencionado prototipo serán *los profesionales de Sistemas que trabajan fuertemente vinculados con los usuarios*. El mismo les brindará el soporte necesario para realizar las recomendaciones para la implementación de un sistema informático de gestión empresarial. Esta herramienta puede emplearse en dos momentos diferentes, un primer momento denominado "*pre implementación*" para conocer cómo abordar el trabajo con la comunidad



usuaria y un segundo momento denominado "post implementación" en el caso que se trabaje con un sistema informático implementado y se tengan que realizar mejoras en la implementación.

Para la construcción del dominio de la aplicación se revisaron los antecedentes propuestos en la tesis de Maestría de Marisa Panizzi (Panizzi, 2012) como así también una serie de papers publicados en WICC y CACIC por Estayno, M., & Panizzi, M.

Se toma como marco teórico para este trabajo, la herramienta de medición de la satisfacción de las necesidades de los usuarios, la cual es capaz de medir la satisfacción socio-técnica de las necesidades de los usuarios de sistemas informáticos de gestión empresarial (Estayno, M., & Panizzi, M., CACIC 2012).

Dicha herramienta se compone dos cuestionarios:

- **MES** es la abreviatura de *Medición Enfoque Socio*. Este cuestionario permite la medición socio de la satisfacción de las necesidades de los usuarios. La medición socio contempla las siguientes dimensiones:
  - Satisfacción en la empresa
  - Satisfacción en el puesto de trabajo
  - Satisfacción con el medio ambiente
  - Comunicación
  - Participación
- **MET** es la abreviatura de *Medición Enfoque Técnico*. Este cuestionario permite la medición técnica de la satisfacción de las necesidades de

los usuarios. La medición técnica contempla las siguientes dimensiones y sub dimensiones:

- Requerimientos Funcionales
- Requerimientos Funcionales No Funcionales
  - Interfaces de Usuario
  - Restricciones de Diseño
  - Requerimientos de Performance
  - Operación
  - Interfaces de Software
- Características del Usuario
  - Educación
  - Experiencia
  - Especialización Técnica

Se utilizarán los resultados de las mediciones obtenidas con los cuestionarios **MES** y **MET** como entradas del prototipo de SE. A partir de las reglas definidas con los expertos se forma la base de conocimiento. Del contenido de la misma se generará el conjunto de recomendaciones para la implementación de los sistemas informáticos, aquellos que den soporte a la gestión empresarial (pudiendo ser parametrizados o desarrollados a medida).

Para poder evaluar la conveniencia de resolver el problema con la construcción del prototipo, se llevó a cabo el estudio de viabilidad siguiendo los lineamientos propuesto en la metodología I.D.E.A.L. Se analizaron las siguientes dimensiones: éxito, justificación, plausibilidad y adecuación. Los resultados de los cálculos se presentan en la **Figura Nro. 1** que se presenta a continuación:

Dimensión	Peso	Valores Intervalo				Peso*Valor				
Plausibilidad	8	7.362	7.73	8.128	8.305	58.9	56.9	62.8	67.5	
Justificación	3	7.8	8.8	10	10	23.4	26.4	30	30	
Adecuación	8	4.259	4.648	7.495	7.831	34.1	19.8	34.8	58.7	
Éxito	5	4.18	4.694	5.312	5.694	20.9	19.6	24.9	30.2	
						24				
							137	123	153	186
Intervalo Resultado Final:							5.72	5.11	6.36	7.77
RESULTADO FINAL:							6.2			

**Figura Nro. 1.** Estudio de viabilidad realizado para el SE generador de recomendaciones para la implementación de sistemas informáticos.

El cálculo final del estudio dió como resultado final **6.2**, indicando de esta manera que era posible llevar a cabo el desarrollo del mismo.

El domino del conocimiento, se construye a través de la aplicación de entrevistas del tipo estructurado, propuestas en el ciclo de educación detallado propuesto por la metodología I.D.E.A.L..(Gómez A., 1997).

Para la construcción del prototipo se ha decidido utilizar la herramienta de desarrollo Kappa- PC de Intellicorp Inc. (Kappa-PC,1992). Esto se debe a que es ampliamente utilizada para la construcción de aplicaciones basadas en conocimiento además de brindar un entorno de desarrollo que permite realizar un prototipado rápido. Esta herramienta da como resultado aplicaciones que incrementan su conocimiento mediante su uso y con esto lograr un desarrollo basado en el prototipo incremental, coincidente con lo propuesto en la metodología I.D.E.A.L..(Gómez A., 1997).

#### Líneas de investigación y Desarrollo.

Esta línea de investigación se originó dentro de los sistemas de información,

más específicamente dentro de la organización, nos orientamos a la satisfacción laboral de los usuarios en sus puestos de trabajo respecto de la implementación de un nuevo sistema informático. De la misma, se desprendieron dos sub líneas, la primera orientada hacia la Ingeniería del conocimiento (la que se presenta) y la segunda orientada al reajuste y validación del instrumento de medición de la satisfacción socio técnica de los usuarios en la cual se viene trabajando de manera paralela.

#### Resultados Obtenidos/ Esperados.

Para avanzar con el objetivo propuesto de este trabajo, se analizaron diferentes metodologías para la construcción de Sistemas Expertos, optando por trabajar con la metodología I.D.E.A.L. Actualmente se está trabajando en la construcción del prototipo.

Se espera contar con el prototipo finalizado y los casos de prueba necesarios a modo de ser presentado en CACIC 2013.

#### Formación de Recursos Humanos.

En equipo de investigación está integrado por un Director, por un

investigador de apoyo y dos estudiantes de la carrera Licenciatura en Sistemas.

### **Bibliografía.**

Asunción Gómez, Natalia Juristo, César Montes, Juan Pazos. Ingeniería del Conocimiento, Editorial: Centro de Estudios Ramón Areces. 1997

Estayno, M., & Panizzi, M. "Prototipo de herramienta para la medición socio técnica de la satisfacción de las necesidades de los usuarios". IV Workshop Innovación en Sistemas de Software (WISS). XVIII Congreso Argentino de Ciencias de la Computación se realizará en la ciudad de Bahía Blanca, del 8 al 12 de Octubre 2012. CACIC 2012- Argentina. ISBN 978-987-1648-34-4.

Estayno, M., & Panizzi, M. "Aproximación a una Herramienta para la medición de la satisfacción de las necesidades de los usuarios". XIV Workshop de Investigadores en Ciencias de la Computación. - WICC 2012. Misiones. 26 y 27 de Abril de 2012. ISBN 978-950-766-082-5.

Estayno, M., & Panizzi, M. "Medición socio-técnica de las implementaciones de los Sistemas de Información Automatizados", Área de Ingeniería de Software y Base de Datos del XIII Workshop de Investigadores en Ciencias de la Computación. Rosario, Santa Fe. 5 y 6 de Mayo de 2011. ISBN 978-950-673-892-1.

Estayno, M., & Panizzi, M. "Implementación mejorada por Participación", Área de Ingeniería de Software y Base de Datos del XII Workshop de Investigadores en Ciencias de la Computación. Calafate, Santa Cruz. 5 y 6 de Mayo de 2010. ISBN 978-950-34-0652-6.

Estayno, M., & Panizzi, M. "Participación de los usuarios en el desarrollo de software", Área de Ingeniería de Software y Base de Datos del XI Workshop de Investigadores en Ciencias de la Computación. San Juan, 7 y 8 de Mayo de 2009. ISBN 978-950-605-570-7.

Kappa PC, Quick Start, Intellicorp, Inc. 1992.

Kappa PC, User's Guide, Intellicorp, Inc. 1992.

Kappa PC, Advanced Topics, Intellicorp, Inc. 1992.

Panizzi, Marisa Daniela (2012). Propuesta de Recomendaciones para la implementación de Sistemas Informáticos. MS. Tesis de Maestría en Informática. Universidad Nacional de La Matanza. 168 pp.

# Mashups en Ambientes Inteligentes

Diego A. Godoy<sup>(1,a)</sup>; Eduardo O. Sosa<sup>(1,b)</sup>;

<sup>(1)</sup>Centro de Investigación en Tecnologías de la Información y Comunicaciones (C.I.T.I.C.). Universidad Gastón Dachary (U.G.D.)

Campus Urbano, UGD. Posadas, Argentina  
Teléfono: +54-376-4438677

<sup>(a)</sup>diegodoy, <sup>(b)</sup>eduardo.sosa@citic.dachary.edu.ar

## Resumen

La aparición de los mashups conjuntamente con la Inteligencia Ambiental abre nuevas posibilidades para el desarrollo de aplicaciones que se adapten al contexto de sus usuarios. Esto da lugar a la necesidad de nuevas arquitecturas en las que se posibiliten la creación de mashups que sean sensibles al contexto y que puedan ser creados de forma práctica por el usuario. Es por ello que en este trabajo se analizarán dos arquitecturas existentes para la creación de esta nueva clase de mashups y se discutirán posibles mejoras. El trabajo presentado aquí se realiza en el marco del proyecto de investigación denominado “Diseño de arquitecturas de soporte a la Internet del Futuro y Ambientes Inteligentes” que se ha formalizado por la resolución (19/A/12) de la Universidad Gastón Dachary.

**Palabras clave:** Mashups, Sensibilidad al Contexto, Inteligencia Ambiental

## Contexto

El trabajo presentado aquí se realiza en el marco del proyecto de investigación denominado “Diseño de arquitecturas de soporte a la Internet

del Futuro y Ambientes Inteligentes” que se ha formalizado por la resolución (19/A/12) del la U.G.D. mediante el llamado al 5to Concurso de Proyectos de Investigación en 2012. EL objetivo de este proyecto es "Diseñar arquitecturas de soporte a Internet del futuro y Ambientes Inteligentes para su aplicación a Ciudades Inteligentes". Como antecedentes y resultados parciales del mencionado proyecto se pueden indicar 3 proyectos de trabajos de grado en curso, correspondientes a la Carrera de Ingeniería en Informática de la U.G.D. y un trabajo final especialización Ingeniería de Software de la Universidad Nacional de la Plata. Indicar el proyecto en que está inserta la línea de I/D presentada y la/s instituciones que coordinan el proyecto. En los casos que corresponda indicar la Institución que acredita el proyecto y los organismos/empresas que contribuyen a su financiamiento.

## Introducción

En los últimos años hemos sido testigos del gran desarrollo de dos tecnologías: los dispositivos móviles (basados en redes inalámbricas, con GPS y sensores) y la Web.



En cuanto a la primera de estas tecnologías se han desarrollado grandes avances en redes inalámbricas, como ser: protocolos más eficientes, mayores anchos de banda, mayores áreas de cobertura, nuevos tipos de sensores, etc. Además los dispositivos móviles cuentan cada vez más, con mayor poder de cómputo, pantallas más grandes y se consiguen a un menor precio. Estos avances facilitan el desarrollo de aplicaciones en donde la movilidad y ubicuidad son parte esencial.

Sin duda una de las características más importantes de los sistemas móviles, es que brindan la posibilidad de aprovechar la posición geográfica para asistir al usuario mientras este se desplaza. Ejemplos de ello son los sistemas que guían a los usuarios en las visitas a museos, sistemas de navegación basados en GPS, o proyectos como Friend Finder de A&T que notifica la presencia de personas geográficamente cercanas.

Hoy en día el desarrollo de sistemas para escenarios móviles ha aumentado considerablemente. Se ha pasado de aplicaciones casi experimentales o utilizadas solo en ambientes académicos, a algunas aplicaciones comerciales como las orientadas a negocios, el uso personal o el ocio.

De la misma manera que las redes inalámbricas, el aumento del poder de cómputo de los dispositivos y capacidad de censado de los dispositivos móviles supone una revolución, el crecimiento de la web también fue revolucionario en cuanto a la forma de presentar la información a los usuarios de sistemas informáticos. Esto dio origen a una evolución de la web, llamada web 2.0, en donde ya los usuarios no son únicamente consumidores, y donde la experiencia del usuario a usar estos sistemas se hizo más agradable, posibilitando personalizar las aplicaciones a las

necesidades concretas de los usuarios y mejorar la interfaz gráfica.

Con estas mejoras de la web, muchas veces ya no son suficiente sitios donde se puedan obtener datos de una sola fuente, es más, se hace sumamente necesario contar con datos de diversas fuentes integradas, y para ello se han desarrollado plataformas que permiten hacer mezclas o “mashups” tanto de información como de comportamiento de las aplicaciones. La particularidad de estas mezclas es que son los mismos usuarios los que tienen la posibilidad de crearlas para cumplir con sus requisitos y no necesitan de un programador o conocer lenguajes como Java o C++.

La Inteligencia Ambiental (AmI) o Computación Ubicua no está ajena a los avances de estas dos tecnologías y se ve beneficiada por ambas. Estos avances permiten que la AmI se acerque cada vez más al cumplimiento a la visión del Information Society Technologies Program Advisory Group (ISTAG) de personas rodeadas de interfaces inteligentes embebidas en objetos de la vida cotidiana que permitan la interacción de forma natural y sin esfuerzo con diferentes sistemas de información.

Este trabajo tiene como objetivo conocer cuáles son las plataformas existentes para la creación de mashups sensibles al contexto en entornos de AmI y proponer mejoras a estas.

La evolución y creciente miniaturización de los dispositivos de cómputo, que hace posible que pequeños procesadores y diminutos sensores se integren cada vez más en objetos cotidianos. Mark Weiser ya había previsto ésta evolución hace más de 10 años y la publicó en su artículo "The Computer for the 21st Century" [1]. En un trabajo posterior Weiser se planteaba los siguientes interrogantes “¿Cuál es la idea de la computadora del futuro? ¿El agente inteligente? , ¿La televisión (multimedia)?, ¿El mundo gráfico

3-D (realidad virtual)? ¿El equipo de voz ubicua de la película StarTrek?, ¿Las interfaces gráficas de usuarios –GUI-, elaboradas y seleccionadas?, ¿La máquina que por arte de magia cumpla y acceda a nuestros deseos?” [2]. La respuesta que él mismo ha dado es: ninguna de las anteriores. La sencilla razón es que todos estos conceptos comparten un defecto básico: se llevan a la práctica con algo tangible, visible.

Es así Weiser que definió el concepto de "computación ubicua", refiriéndose a las computadoras omnipresentes al servicio de las personas en su vida cotidiana, en el hogar y en el trabajo, y funcionando de manera invisible y no intrusita. Los principios enunciados originalmente por Weiser han sido validados por diferentes investigadores e ingenieros, convirtiéndose en un escenario real en ciertos ámbitos y extendiéndose a escala global [3].

En 1999, el Information Society Technologies Program Advisory Group (ISTAG) de la Unión Europea ha utilizado el término "inteligencia ambiental" del inglés Ambient Intelligence (AmI) de manera similar para describir una visión donde "las personas estarán rodeadas de interfaces inteligentes e intuitivas embebidas en objetos cotidianos de nuestro alrededor y un entorno que reconocen y responden a la presencia de individuos de manera invisible [4]. Es así que el objetivo de la AmI consiste en la creación de espacios donde los usuarios interaccionen de forma natural y sin esfuerzo con los diferentes sistemas.

Uno de los aspectos más importantes de la Inteligencia ambiental son las aplicaciones sensibles al contexto, es por ello que es necesario entender a que nos referimos con este concepto. Existen varias definiciones: La primera fue dada en [5], la cual restringía a estas a ser simplemente aplicaciones que eran informadas sobre el contexto para que se adaptaran a este. Otras

definiciones como la de [6] definían a la computación sensible al contexto como la habilidad que poseen los dispositivos de detectar, sensor, interpretar y responder a los aspectos locales al ambiente de un usuario. Por su parte, Dey [7] define la noción de context-awareness como el “uso del contexto para automatizar un sistema de software, modificar su interface y proveer la máxima flexibilidad en términos de servicios”.

Otra definición un poco más pasiva es la dada en [8], donde una aplicación sensible al contexto es aquella que pueda variar o adaptar dinámicamente su comportamiento en base al contexto. Salber [9] define que una aplicación sensible al contexto es aquella que tiene la capacidad de proporcionar la máxima flexibilidad de un servicio de cómputo basado en el censo del contexto en tiempo real.

Por último, en una forma más general pero a la vez concreta, en [10] Dey define a un sistema sensible al contexto como: “Un sistema (aplicación) es sensible al contexto si este hace uso de información del contexto para proveer información o servicios relevantes al usuario, donde la relevancia depende de actividad actual del usuario”. Esta definición resulta ser la mayormente aceptada.

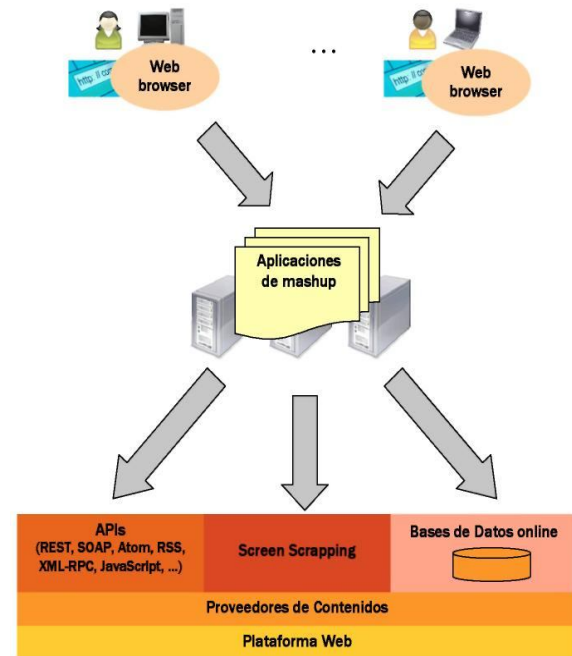
El auge de la Web 2.0 ha hecho que la industria incorpore las nuevas tecnologías que dieron lugar a su origen y se ajusten a los estándares emergentes. Mientras aun hoy no existe consenso sobre el alcance del término Web 2.0, O'Reilly ofrece una definición comúnmente aceptada, que incluye una gama de servicios mejorados, como ser los servicios web, wikis, blogs, bittorrents, y la sindicación de contenidos [11].

El crecimiento de la Web 2.0 también ha introducido un número de patrones de diseño y nuevos estilos arquitectónicos en el desarrollo

web. Una de las técnicas más destacadas consiste en la “mezcla” o mashup.

En [12] se presenta una definición precisa y completa la cual define a un mashup como una “Aplicación basada en Web que se crea mediante la combinación y procesamiento de los recursos en línea de terceros, que contribuyen con datos, forma de presentación o funcionalidad”. En esta definición, los recursos en línea de terceros se refieren a cualquier tipo de recursos disponibles en Internet, independientemente del formato. Como resultado, un mashup proporciona un nuevo recurso.

En términos generales, los mashup se crean usando el navegador web, “arrastrando y soltando” es decir, poniendo juntas aplicaciones de diferentes fuentes. Sin embargo, estas operaciones deben tener una infraestructura en el back-end para que de soporte al mashup, por ello en [13,14], se presenta una arquitectura de mashup que consta de tres componentes la cual se puede ver en la Figura 1. a) La API / proveedores de contenido. Estos son los proveedores de contenido (a veces heterogéneos) de los cuales se desea hacer el mashup. Para facilitar la recuperación de datos, los proveedores suelen dar acceso a sus contenidos a través de protocolos de Internet tales como REST, los Web Services o feeds RSS / Atom o utilizando screen scraping [15]. b) El hosting del mashup. Es donde el mashup se encuentra alojado. Los mashups pueden utilizarse de manera similar a las aplicaciones Web tradicionales utilizando un servidor de contenido dinámico (Java Servlets, CGI, PHP o ASP). Alternativamente, el contenido del mashup se puede generar directamente en el navegador del cliente utilizando por ejemplo JavaScript o applets. c) El navegador Web de los consumidores. Es donde la solicitud se representa gráficamente y la interacción del usuario con el mashup se lleva a cabo.



**Figura 1. Arquitectura de Mashups**

## Líneas de investigación y desarrollo

En esta sección se analizarán dos proyectos de plataformas existentes para construir mashups sensibles al contexto:

1) El Caso del Proyecto Deusto Sentient Graffiti (DSG): El proyecto Deusto Sentient Graffiti [16, 17] busca la convergencia de la computación móvil y la computación ubicua con la web 2.0 para configurar lo que hoy día se llama Web Ubicua (UW). Esta última se puede definir como una infraestructura de Internet omnipresente en la que todos los objetos físicos son recursos accesibles a través de una URI. Estas URIs pueden proporcionar información y servicios que enriquezcan la experiencia de los usuarios en su contexto físico.

Su funcionamiento se basa en las anotaciones espontáneas o “Graffitis” (en formato de documento XML) que realiza una comunidad de usuarios sobre objetos, lugares o incluso

personas, con contenido web multimedia y/o servicios que pueden ser descubiertos y utilizados a su vez por otros usuarios móviles cuyos atributos contextuales coincidan con las de las anotaciones. La funcionalidad de DGS consiste en dos procesos principales: 1) Anotación de Grafitis: Los usuarios de los clientes DGS anotan objetos o regiones espaciales las cuales consistiendo de Descripciones, palabras clave y atributos contextuales y 2) descubrimiento y consumo de Grafitis, cuando los usuarios, pueden moverse virtual o físicamente a través de un ambiente navegando y consumiendo las anotaciones disponibles

DGS puede verse como mashup sensible al contexto dado que los clientes de DGS consiguen combinar la información geográfica en forma de mapas (ejemplo Google Maps) y la información

suministrada por el servidor DSG en las descripciones.

La arquitectura de DGS presenta los siguientes componentes (Figura 2): a) Administrador de Grafiti y Web Server: administra todos los grafitis virtuales publicados b) Disparador y Notificador de Grafitis: Es un motor de inferencia basado en reglas que deduce los conjuntos de anotaciones activas que tienen notificarse a los usuarios c)

Motor de Consultas de Grafitis: Este componente es responsable del procesamiento de consultas y de la recuperación de la anotaciones sensibles al contexto d) Monitores de Cambio de Contexto: Se ejecutan en el dispositivo móvil o navegador web vigilando el contexto del usuario, y e) Reproductor de Grafitis: Muestra grafitis virtuales en el dispositivo del usuario.

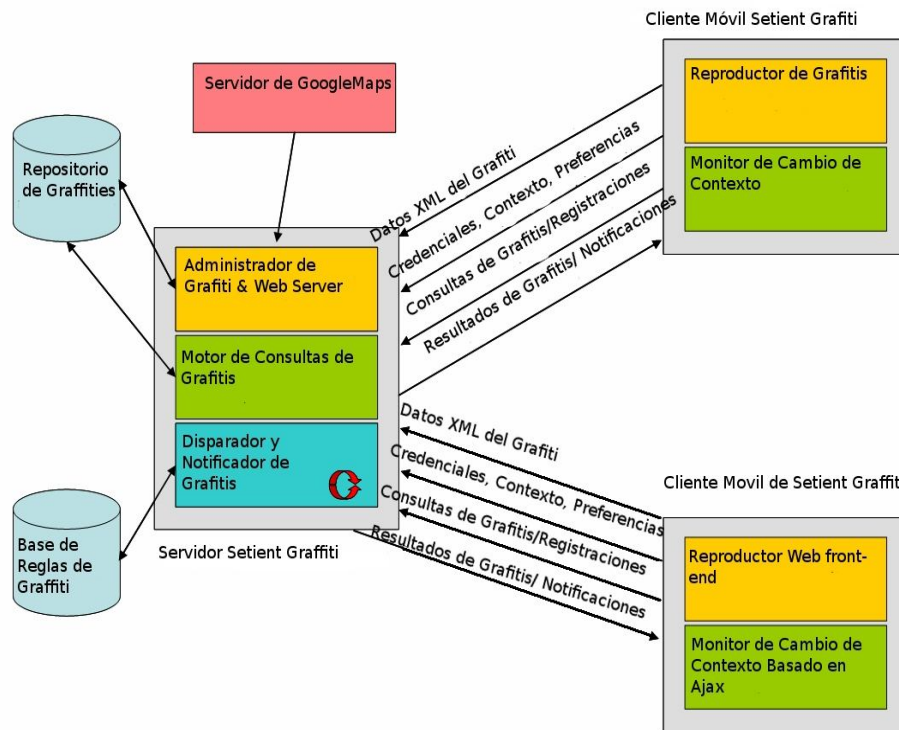


Figura 2. Componentes de DSG



Para su implementación DGS utiliza una sintaxis llamada GraffitiXML que se usa para especificar contenido. Como todo sistema basado en Web 2.0, como Google maps o Flickr, DGS provee una API basada en HTTP que permite a programadores de terceras partes desarrollar aplicaciones que se beneficien de su funcionalidad.

La implementación del servidor DGS está basada en Java y se ejecuta sobre el servidor de aplicaciones Tomcat utilizando JAVA EE. Además utiliza los marcos de trabajo, Tapestry, Hibernate, y Java Expert Systems Shell y como base de datos MySQL.

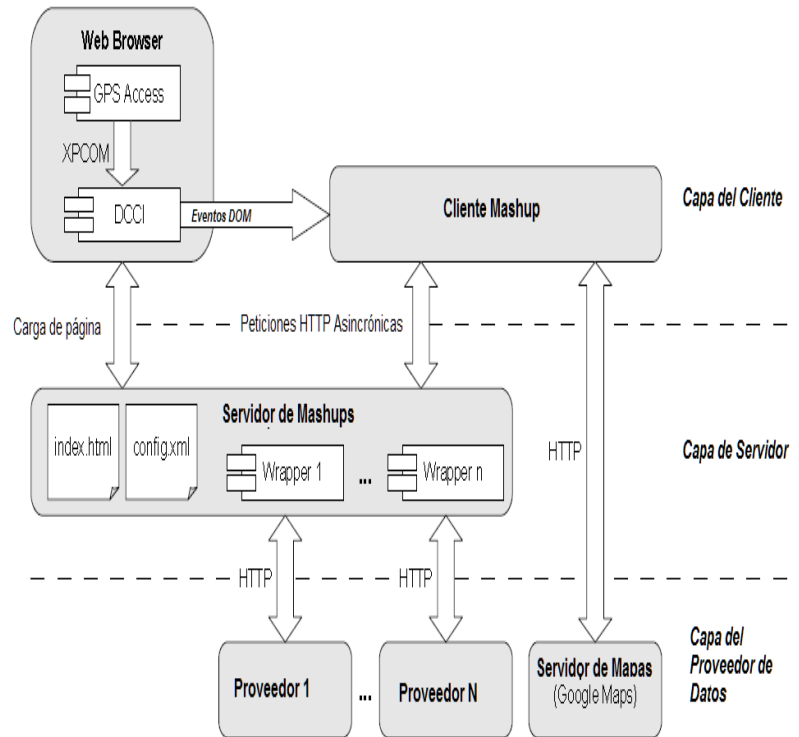
2) El Caso del Proyecto TELAR: Los autores de TELAR para la plataforma en [18] y que luego amplían en [19] se pueden resumir de la siguiente manera: a) La adaptación debe estar basada en sensores que se integran en el dispositivo móvil o conectado localmente. B) Los mashups deben ser centrados en el usuario, es decir, el usuario de un dispositivo móvil debe beneficiarse de las aplicaciones web híbridas, en lugar de otros usuarios distantes o los proveedores de servicios, c) Los mashups deben ser visibles con el navegador web del dispositivo móvil. Esto impide una solución nativa para el dispositivo móvil y tiene influencia potencial en el rendimiento, d)

Debe existir una versión no sensible al contexto del mashup en caso de que la información de contexto no esté disponible. Esto permite que se puedan ver los mashups en los dispositivos móviles equipados con sensores, así como en equipos que no los posean e) Debe existir la posibilidad de que múltiples fuentes de datos con formatos de datos distintos

e interfaces arbitrarias puedan integrarse en los mashups.

La base para la integración es un mapa provisto por un servicio de mapas en línea (Google Maps) y de la ubicación actual del usuario (GPS). Se integran diferentes proveedores de puntos de de interés (POIs). Los POIs se muestran en el mapa con diferentes símbolos. A medida que el usuario se mueve, el mapa se mantiene centrado en su posición y se muestra un rastro de su movimiento si el usuario así lo desea. Por el lado del cliente, la especificación Delivery Context Client Interfaces (DCCI) se utiliza para integrar la información de contexto de local obtenida a través de sensores con el navegador web móvil, que adapta el mashup a la localización actual del usuario.

En la Arquitectura presentada en la Figura 4 se puede observar la arquitectura del sistema TELAR, que de la misma manera que una aplicación mashup típica tiene tres niveles. El mashup se visualiza en la capa del cliente. El navegador carga la página web del mashup y ejecuta el código Javascript (AJAX) en el cliente. Esta página del mashup se carga desde el servidor de Mashup, que se encuentra en Internet en la capa servidor. Los datos ofrecidos por terceros proveedores que se utilizan se distribuyen por toda la web, estos se encuentran en la Capa de Proveedores de datos. Los mapas se cargan desde un servicio de mapas, que junto con los proveedores de datos, constituyen la capa de datos. Esta última capa se encuentra fuera de los límites de la organización del mashup.



**Figura 3. Componentes de TELAR**

El cliente mashup utiliza lo desplegado en el servidor de mashup y estos pueden ser configurados/ parametrizados pero no necesitan ser programados por los usuarios finales. El cliente mashup lee asincrónicamente la configuración cuando la página del mashup se carga. Luego el cliente mashup construye la interfaz de usuario, esta muestra el mapa y visualiza los datos POIs de los proveedores de datos.

Una capa de normalización se utiliza para lograr una interfaz consistente para la recuperación de datos utilizando wrappers, ya que este enfoque ofrece el más alto grado de flexibilidad. Para otros escenarios con menos diversidad, se pueden utilizar enfoques de normalización automática (por ejemplo, proveedores de datos que proporcionan canales RSS, como el usado en Yahoo Pipes [20]).

Tanto la información de contexto, como la ubicación del usuario, se integran al mashup, ampliando el navegador web. Para ello se utilizan dos componentes adicionales: el módulo DCCI y el módulo de acceso GPS. El módulo que implementa la especificación DCCI constituye la interfaz para proporcionar datos de contexto a las páginas web. El cliente mashup se registra como un detector de eventos en el módulo DCCI y este es notificado cada vez que se produce un cambio en la localización del usuario. El módulo de acceso GPS se conecta al dispositivo GPS y este pasa la información de ubicación necesaria al módulo DCCI.

Para implementar el cliente se utilizó el marco de trabajo Google Web Toolkit (GWT). Este dio la posibilidad de desarrollar un cliente mashup complejo en Java utilizando herramientas profesionales. Del lado del servidor, TELAR

requiere una aplicación servidor desde la cual el mashup se cargue y donde residan los wrappers de los proveedores de datos.

## Resultados

Si bien existen herramientas para construir mashups sensibles al contexto en entornos de AmI estas son incipientes y todavía están en fase de prototipo. Aún no se cuenta con herramientas con la potencia y facilidad de uso de Yahoo Pipes para escenarios AmI.

Las tecnologías utilizadas, la forma de interacción con los usuarios, y de adquisición de conocimiento de este nuevo tipo de mashup, no están estandarizadas y existen múltiples estrategias y plataformas (como las mencionadas en la sección 4) para su construcción por lo que la mayoría de estas propuestas planteen una solución práctica para el escenario específico en cual son aplicadas.

Esto hace difícil integrar datos obtenidos desde sensores y utilizarlos para adaptar el contenido o el comportamiento de los mashups (y de cualquier aplicación web) ya que no existe un estándar de cómo capturar, representar y utilizar la información de contexto en las aplicaciones web.

Lo comentado anteriormente da lugar a la necesidad de patrones, tanto para la adquisición de datos desde sensores (e información derivada) como en las estrategias para que estos datos sean aprovechados por los sistemas mashups de forma segura y que no comprometan la privacidad. Detallar sintéticamente los ejes del tema que se están investigando.

Como líneas futuras se proponen dos pilares, uno desde el punto de vista de las aplicaciones para la creación de mashups sensibles al contexto, se plantea la necesidad de contar con una herramienta del tipo de Pipes de Yahoo, pero que

además de fuentes de datos RSS, APIS, y cajas de texto de entrada de datos (user input) pueda incorporar información de contexto provista por los sensores (context input) del dispositivo móvil del usuario de manera automática.

Siguiendo con esta línea como el segundo pilar sería interesante contar con un navegador web sensible al contexto, que de soporte a la AmI. Este Navegador de Web Ubicua (navegador sensible al contexto), se lo podría plantear como una evolución de un navegador tradicional como Mozilla Firefox o una extensión para los ya existentes. De esta forma se podría realizar una navegación de Ambientes Inteligentes. Esta extensión o mejora permitiría recabar datos de los sensores de dispositivos móviles de forma transparente para el usuario y las aplicaciones mashups e independientemente del hardware o tecnología subyacente, En este sentido se está trabajando con HTML5 y formatos interoperables como ser REST y RSS y abstraer WSN (Redes de Sensores Inalámbricos por sus siglas en inglés) de la publicación de datos que ellas generan para lograr el objetivo.

## Formación de Recursos Humanos

El equipo de trabajo se encuentra formado por un Doctor en Ciencias Informáticas, Un Doctorando en Ingeniería Telemática, un auxiliar de investigación graduado, y ocho auxiliares (Resolución rectoral 21/I/12) de investigación en período de realización de trabajos finales de grado. El número de tesis de grado en curso con proyecto aprobado es tres y el número de trabajo de especialidad en curso con proyecto aprobado es uno. Los proyectos de grado se titulan "Plataforma para la publicación de datos de Redes de Sensores Inalámbricos, orientada a la visión de la Internet de las Cosas, Ambientes

Inteligentes y Mashups”, “Diseño de un prototipo para monitoreo eficiente de iluminación basado en WSN utilizando HTML5” y “Contribución a la Gestión de Residuos Domiciliarios como una Aplicación en Ciudades Inteligentes”. El Trabajo de Especialidad se titula “Plataformas para la creación de Mashups Sensibles al Contexto en Entornos de Inteligencia Ambiental”

## Referencias

- [1] Weiser M, The Computer for the 21st Century. Scientific American, 265(3):66–75, September 1991
- [2] Weiser, M.. Some computer science issues in Ubiquitous Computing. Communications of ACM, 36(7), 75-84 1993
- [3] Sosa, Eduardo O., Contribuciones al establecimiento de una red global de Sensores Inalámbricos. Tesis Doctoral. s.l.: Universidad Nacional de La Plata, Junio 17, 2011.
- [4] Ahola J, Ambient Intelligence, ERCIM News, No 47, October 2001.
- [5] Schilit, B., Theimer, M. Disseminating Active Map Information to Mobile Hosts. IEEE Network, 8(5) (1994) 22-32
- [6] Pascoe, J. Adding Generic Contextual Capabilities to Wearable Computers. 2nd International Symposium on Wearable Computers (1998) 92-99
- [7] Dey, A.K. Context-Aware Computing: The CyberDesk Project. AAAI 1998 Spring Symposium on Intelligent Environments, Technical Report SS-98-02 (1998) 51-54
- [8] Gerd Kortuem, Zary Segall, and Martin Bauer. Context-aware, adaptive wearable computers as remote interfaces to ‘intelligent’ environments. In ISWC, pages 58–65, 1998.
- [9] Salber, D., Dey, A.K., Abowd, G.D. Ubiquitous Computing: Defining an HCI Research Agenda for an Emerging Interaction Paradigm. Georgia Tech GVU Technical Report GIT-GVU-98-01 (1998)
- [10] Anind Kumar Dey. Providing architectural support for building context-aware applications. PhD thesis, Georgia Institute of Technology, 2000. Director-Gregory D. Abowd.
- [11] T. O’Reilly, “What Is Web 2.0, Design Patterns and Business Models for the Next Generation of Software”, O’Reilly Media Inc., September 2005.
- [12] A. Koschmider, V. Torres, V. Pelechano: Elucidating the Mashup Hype: Definition, Challenges, Methodical Guide and Tools for Mashups. 2nd Workshop on Mashups, Enterprise Mashups and Lightweight Composition on the Web (MEM 2009).
- [13] Duane Merrill. Mashups: The new breed of Web app--Anintroduction to mashups. <http://www.ibm.com/developerworks/xml/library/x-mashups.html>
- [14] Xuanzhe Liu; Yi Hui; Wei Sun; Haiqi Liang, "Towards Service Composition Based on Mashup," Services, 2007 IEEE Congress on , vol., no., pp.332-339, 9-13 July 2007
- [15] Sitio de Internet: Screen Scrapper, <http://www.screen-scraper.com/>
- [16] López de Ipiña D., Vázquez J.I. and Abaitua J., Context-Aware Mobile Mash-up for Ubiquitous Web, Puertollano, Spain, ISBN: 84-6901744-6, pp. 19-34, November 2006
- [17] D. Lopez-De-Ipina, J. I. Vazquez, J. Abaitua. "A context-aware mobile mashup platform for ubiquitous web". Intelligent Environments, 2007. IE 07. 3rd IET International Conference on (2007), pp. 116-123.
- [18] Andreas Brodt , Daniela Nicklas, The TELAR mobile mashup platform for Nokia internet tablets, Proceedings of the 11th international conference on Extending database technology: Advances in



database technology, March 25-29, 2008, Nantes, France

[19] A. Brodt, D. Nicklas, S. Sathish, and B. Mitschang. Contextaware mashups for mobile devices. In Web Engineering (WISE '08), pages 280–291. Springer-Verlag, 2008.

[20]Sitio de Internet: Yahoo Pipes, <http://pipes.yahoo.com/pipes/>

# Desarrollo de interface de detección de emociones para su utilización en redes sociales y entornos virtuales de aprendizaje

Eduardo Nicolás Campazzo, Marcelo Martinez,  
Alejandra Elena Guzmán, Andrea Leonor Agüero

Secretaría de Ciencia y Tecnología/Departamento de Ciencias Exactas Físicas y  
Naturales/Universidad Nacional de La Rioja

Av. Luis M. de la Fuente S/N

Tel: 0380-4457000

[ecampazzo@yahoo.com.ar](mailto:ecampazzo@yahoo.com.ar)/[mmartinez@estudioe.com.ar](mailto:mmartinez@estudioe.com.ar)/[aleguzman2002@hotmail.com](mailto:aleguzman2002@hotmail.com)/[aaguero1903@gmail.com](mailto:aaguero1903@gmail.com)

## Resumen

Las emociones condicionan la interpretación de los mensajes que recibimos y el proceso de enseñanza-aprendizaje y desarrollo, tanto en la comunicación presencial como a través de las redes de datos (entornos virtuales de aprendizaje o redes sociales). Uno de las deficiencias de las redes de virtuales, para las relaciones interpersonales, es que no existe la posibilidad de conocer el estado anímico de nuestros interlocutores, ya sea personas a las cuales enviamos el mensaje o alumnos que se encuentran conectados a través de un entorno virtual de aprendizaje.

La interfaz que se propone desarrollar en la ejecución de este proyecto realizará un monitoreo en tiempo real de las ondas cerebrales y mediante un algoritmo computarizado inferirá emociones en tiempo real para ser transmitidas por la red. La interfaz propuesta como solución al problema consta de electrodos estratégicamente ubicados en la parte exterior de la cabeza (sistema no intrusivo) a través de un casco flexible diseñado para tal fin. Los electrodos capturan las señales eléctricas producidas por el cerebro, las que son amplificadas por circuitos electrónicos, y posteriormente estas señales analógicas son convertidas en señales digitales con conversores analógico/digital. Una vez capturadas y almacenadas las señales digitales serán

analizadas a través de comparaciones con patrones previamente establecidos para inferir las emociones del usuario y posteriormente transmitir las a la red de datos.

**Palabras clave:** interfaz Hombre-Máquina, ondas cerebrales, emociones, redes sociales.

## Contexto

El presente trabajo se enmarca en la Universidad Nacional de La Rioja, por iniciativa del equipo de investigación conformado por docentes de las carreras de Licenciatura e Ingeniería en Sistemas, pertenecientes al Departamento de Ciencias Exactas, Físicas y Naturales.

El proyecto fue presentado, evaluado y aprobado en forma externa y financiado por la Secretaría de Ciencia y Tecnología de la Universidad Nacional de La Rioja.

El proyecto presentado a través del presente documento fue remitido por la Secretaría de Ciencia y Tecnología de la Universidad Nacional de La Rioja, para la gestión de evaluación interna y externa dentro de la comunidad de investigadores.

Actualmente se encuentra en la etapa de ejecución, presentando los primeros avances del mismo.

## Introducción

En la ejecución de este proyecto se desarrolló una interfaz hombre-máquina compuesta por hardware y software, integrando cada una de las partes componentes para lograr un dispositivo funcional.

Para la captura y conversión de las señales eléctricas se utiliza un dispositivo denominado Arduino, que posee un microcontrolador, conversores A/D y conexión a puerto USB. A través de la conexión USB se transmite en tiempo real la captura de las ondas cerebrales. Estas señales son procesadas en la computadora para luego ser interpretadas y transmitidas a la red de datos.

Existen interfaces hombre-máquina que permiten realizar el control de computadoras desde diferentes dispositivos. Desde el teclado hasta el control por detección de movimiento de los ojos, siempre valiéndose de movimientos del usuario introducir los datos al sistema computarizado. Actualmente se experimenta con el control de computadoras a través de Brain Control Interface (BCI), y con el mismo se pretende lograr el control computarizado aun cuando se carece de posibilidad de movimiento del usuario.

En la comunidad científica se experimenta diferentes maneras de detectar emociones en los usuarios de computadoras, ya sea a través de detección gestual del rostro o bien del modo de escritura. Lo cierto es que incorporar las emociones a los mensajes que transmitimos a través de las redes permite una mejor interpretación de esos mensajes, ya que las emociones condicionan la interpretación de los mensajes, tanto al enviarlos como al recibirlos.

En la actualidad existen interfaces que permiten el censado de parámetros biométricos. Entre los que encontramos aquellos que detectan actividad eléctrica del cerebro (electroencefalogramas), los que detectan la actividad eléctrica del sistema circulatorio (electrocardiogramas), los que detectan la conductividad de la piel para inferir ansiedad, entre otros.

Esta interfaz permitirá convertir la actividad eléctrica del cerebro, inducida por los cambios

emocionales del usuario, en señales digitales para ser analizadas y procesadas por un sistema computarizado.

La interfaz consta de electrodos ubicados en la parte exterior de la cabeza a través de un casco flexible diseñado para tal fin. Los electrodos capturan las señales eléctricas producidas por el cerebro, estas señales son amplificadas por amplificadores de instrumentación y filtradas para eliminar las señales de frecuencias no significativas en el análisis y el procesamiento. Una vez adaptadas las señales son muestreadas por conversores analógico/digital y transmitidas a una computadora a través de un dispositivo con un microcontrolador, denominado Arduino, conectado al puerto USB. En la computadora se instala y ejecuta un software desarrollado en lenguaje C#, el mismo es el encargado de capturar y procesar las señales digitales y a su vez ejecutar código java script instalado en un hosting. Este código java script es conectado a una aplicación Facebook que es la que realiza publicaciones automáticas en el muro del usuario de la red social.

El sistema desarrollado en C# permite automatizar la detección de las variaciones en las señales eléctricas del cerebro, tanto en amplitud como en frecuencia. Actualmente la interfaz desarrollada permite inferir estados de excitación y somnolencia en los usuarios, a través de la detección de variaciones de frecuencia en las ondas cerebrales, para posteriormente realizar la respectiva publicación en la red social. Se prevé en un futuro próximo poder inferir también estados emocionales básicos como alegría, tristeza e indiferencia en base al estudio de la respuesta de distintos usuarios a diferentes estímulos.

Con el uso de esta interface se podrán incorporar en un futuro próximo aplicaciones a las redes sociales y entornos virtuales de aprendizajes que incorporen el estado anímico de los usuarios y de esta forma agregar un parámetro más al perfil de quienes se encuentran registrados y así extender las funcionalidades de estas redes a campos relacionados con la neurociencia o inteligencia emocional.

Es necesario destacar que esta interface se

encuentra en estado experimental, desarrollado un prototipo capaz de detectar estados de somnolencia y excitación. Actualmente está en desarrollo la detección de patrones de señales de acuerdo al estado emocional.

En un futuro este tipo de interfaces permitirán poder detectar emociones y utilizarlas en el campo de la medicina pudiendo ayudar a quienes padecen enfermedades como la depresión, diagnosticando patologías neuronales a través de las redes sociales.

El presente proyecto estará constituido por cuatro (5) etapas:

- Se diseñará y desarrollará amplificadores de instrumentación para adecuar la señal eléctrica del organismo a la tarjeta de adquisición de datos.
- Se desarrollará un software que permita graficar las señales eléctricas censadas en el organismo del usuario. Conjuntamente se realizara una investigación de la bibliografía que hace referencia la relación que existe entre las zonas eléctricamente activas con los estados emocionales de la persona para de esta forma inferir en tiempo real el estado anímico de los usuarios.
- Se realizara una investigación exploratoria con respecto a las alteraciones de los parámetros bio-eléctricos del organismo humano que sufre cuando es sometido a cambios emocionales. De esta forma se determinará la factibilidad de medición a través del censado electrónico de estos parámetros.
- Se desarrollará una taxonomía que permitirá a través de tablas la relación directa entre actividad eléctrica del organismo y su correspondiente estado anímico.

Es importante destacar, que a febrero del 2013, el equipo de investigación se encuentra en la etapa dos (2) de las arriba mencionadas. Esto implica que ya se encuentra desarrollado un prototipo de interfaz capaz de capturar las ondas cerebrales y también conectar en forma automática las redes sociales.

### **Incorporación de emociones a las plataformas virtuales de aprendizajes (EMOEVA UNLaR)**

Una de las desventajas de los entornos virtuales de aprendizaje es que a través de estas plataformas no hay un contacto visual entre el educador y el educando que permita inferir las emociones que están sintiendo ambos en el transcurso de una clase. Con el desarrollo de una interfaz que detecta cambios emocionales en los usuarios de estas plataformas y transmitirlos a través de redes de datos y ser visualizados por los interlocutores que participan en las plataformas virtuales de aprendizajes, estaríamos potenciando el uso y la aplicabilidad de estas plataformas, aumentando las capacidades de desarrollo de los usuarios.

Por otro lado existen entornos virtuales de mundos 3D, en los cuales cada usuario tiene una representación gráfica inmersa en el mundo virtual llamado avatar. El avatar es una representación icónica del usuario, un modelo que carece de representatividad emocional del mismo, esto significa que el usuario no puede transmitir su estado emocional a través de su representación icónica, es por ello que se pretende desarrollar una interfaz (hardware-software) que permita adicionar al avatar una animación en representación del estado anímico del usuario en respuesta a la actividad eléctrica de su propio organismo.

En definitiva lo que se quiere lograr a través del diseño y desarrollo de interfaces emocionales para ser utilizadas en plataformas virtuales de aprendizajes, es subsanar los problemas que se generan al no tener un contacto presencial con los educandos y no poder percibir a través de ese contacto las emociones y sentimientos que los mismos experimentan en el transcurso de la clase.

Utilizar una herramienta tecnológica que nos permita detectar emociones o cambios emocionales en los educandos puede ser aplicada para la reestructuración de una planificación preestablecida con el objeto de optimizar recursos y mejorar el rendimiento académico.



## Líneas de investigación y desarrollo

El equipo de investigación se encuentra trabajando en dos ejes temáticos:

### **Desarrollo de interfaz de hardware para la detección de ondas cerebrales.**

En esta línea de investigación y desarrollo se desarrolla un hardware capaz de censar en tiempo real las ondas cerebrales. Para lo cual se diseñó y desarrollo un circuito electrónico para de capturar y adecuar las débiles señales eléctricas del cerebro y amplificarlas para posteriormente digitalizarlas y procesarlas.

### **Desarrollo de interfaz de software para el análisis y procesamiento de ondas cerebrales digitalizadas.**

En esta línea de investigación se analiza y procesa las señales digitales provistas por la interfaz de hardware, con el objetivo de inferir estados anímicos a través del procesamiento de dichas señales. Y de esta manera poder conectar y publicar en redes sociales o entornos virtuales de aprendizajes un estado emocional del usuario.

## Resultados y Objetivos

Objetivos generales:

- Desarrollar interfaz de hardware para censar actividad eléctrica del organismo.
- Desarrollar una interfaz de software para interconectar la interfaz de hardware las redes sociales y los entornos virtuales de aprendizaje.

Objetivos específicos:

- Determinar los parámetros biométricos más relevantes en los cambios emocionales.
- Desarrollar un amplificador de instrumentación para adecuar las señales eléctricas del organismo e introducir las a la tarjeta de adquisición de datos.
- Desarrollar un software para visualizar las señales bio-eléctricas censadas.

- Determinar que representa la alteración de señales bio-eléctricas contrastándolas y determinando cambios emocionales.
- Desarrollar un software que permita generar código para ser utilizado en tiempo real en el mundo virtual 3D utilizado en el proyecto SLEVA-UNLaR (entorno virtual de aprendizaje).
- Desarrollar una aplicación que permita conectar las interfaces desarrolladas con las redes sociales.

Con la ejecución del presente proyecto se pretende diseñar y desarrollar interfaces hombre-máquina capaz de incorporar parámetros biométricos, relacionados con los estados anímicos de los usuarios, a las redes sociales y a los entornos virtuales de aprendizajes basados en SLEVA.

## Formación de Recursos Humanos

El equipo de investigación se encuentra formado por:

En proceso de desarrollo de Tesis Doctoral, Universidad de Vigo, España:  
Campazzo, Eduardo Nicolás

En proceso de desarrollo de tesis de maestría Universidad Nacional de San Luis:

Campazzo, Eduardo Nicolás

Martinez, Marcelo

Guzmán, Alejandra

Profesores de Universidad Nacional de La Rioja:

Agüero, Andrea Leonor

Gramajo, Susana Cristina

Roldán, Marcelo

También en el equipo de investigación participan alumnos avanzados de la carrera de Ingeniería en Sistemas y Licenciatura en Sistemas que se encuentran realizando el trabajo final de carrera.

## Referencias

1. Campazzo E., Martínez M., Guzmán A. y Agüero L. "De la Presencialidad a la

- Interacción Virtual 3D” 2010. ISBN: 978-987-661-047-6
- 2 Campazzo E., Martínez M., Guzmán A. y Agüero L. "El tutor y su rol en los mundos tridimensionales en la WEB 3.0" 2010. ISBN: 978-950-579-168-2
  - 3 Campazzo E., Martínez M., Guzmán A. y Agüero L. "La diversidad de aprendizajes con la evolución de la WWW. Del modelo estático plano WEB 1.0 al tridimensional interactivo, corpóreo y persistente de la WEB 3.0" - 2010. ISSN:1682-2749 .
  - 4 Campazzo E., Martínez M., Guzmán A. y Agüero L. "Mundos Virtuales 3D como nuevo paradigma en E-learning. Caso: SLEVA en la Universidad Nacional de La Rioja – Argentina" 2010- ISBN:978-950-9474-49-9 .
  5. Martínez Marcelo-Campazzo Eduardo-Guzmán Alejandra- Agüero Leonor "Aplicación de mundo virtuales 3D en e-learning. Caso: SLEVAUNLAR (Second Life + Moodle) en la Universidad Nacional de La Rioja.
  6. D Livingstone, M. Crowe, and P. Bloomfield, "HTML on a Prim: Uses and Abuses," presented at Second Life Education Community Conference, Tampa, Florida, 2008.
  7. López García, P; Sein, M; MOODLE: Difusión y funcionalidades - Dpto. Informática e Ingeniería de Sistemas 1 - C.P.S. Universidad de Zaragoza - plopezg@unizar.es Dpto. Matemática Aplicada 2 - C.P.S. Universidad de Zaragoza mlsein@unizar.es  
[http://www.unizar.es/ees/innovacion06/CO MUNIC\\_PUBLI/BLOQUE\\_III/CAP\\_III\\_10.pdf](http://www.unizar.es/ees/innovacion06/CO MUNIC_PUBLI/BLOQUE_III/CAP_III_10.pdf)
  8. M. Rymaszewski, W. J. Au, M. Wallace, C. Winters, C. Ondrejka, B. Batstone-Cunningham, and S. L. residents from around the world. Second Life: the office guide. Wiley Press, 2007.
  9. Puy, M; Larrainzar,A; Escudero Herrera, C; Santamaria Gonzalez, ,F. "El mundo virtual: Second life y su aplicación a la enseñanza del derecho". Universidad a distancia de Madrid. 2008.
  - 10 Qing Zhu, Tao Wang, Yufu Jia. "Second Life: A New Platform for Education". IEEE. 2007.
  - 11 Ruan Jianhai, Deng Xiaozhao. "On the Second Life-based Education in Virtual World". IEEE. 2009.
  - 12 Schnook M, Sullivan A. How To Get a Second Life. 1st. Ed. Fusion Press. 2007.
  - 13 Second Life. <http://www.secondlife.com>
  - 14 eLearning review Informe Especial Número 1 Virtual 3D Asesora editorial: Ruth Martínez
  - 15 Roldan, Marcelo F. Neuro Aprendizaje Sometido a Estímulos de Riesgos. Trabajo final de la carrera licenciatura en análisis de sistemas. UNLaR. 2005
  - 16 Belústegui, Gustavo D. "Los aspectos afectivos-emocionales en las teorías implícitas. Condiciones, procesos y resultados en la enseñanza aprendizaje". UCA – ISSN 16815653
  - 17 Lozares Colina, Carlos. Interacción, redes sociales y ciencias cognitivas. 2007.

## REPOSITORIO ONTOLOGICO INSTITUCIONAL

**Elena Durán<sup>1,2</sup>, Patricia Paola Zachman<sup>1</sup>, Margarita Alvarez<sup>1,2</sup>, Jonatan Aguilera<sup>1</sup>,  
Juan Nuñez<sup>1</sup>, Marcos Umaño<sup>1</sup>**

(1) Departamento de Tecnologías Aplicadas.-

Universidad Nacional del Chaco Austral.- Pcia. Roque Sáenz Peña.- Chaco

ppzsp@arnet.com.ar, jonemmanu@gmail.com.ar, sadamssh@esdebian.org, marcosrene@live.com.ar

(2) Departamento de Informática, Facultad de Ciencias Exactas y Tecnología,

Universidad Nacional de Santiago del Estero.

eduran@unse.com.ar, alvarez@unse.edu.ar

### Resumen

*En muchas instituciones universitarias la información resultante de actividades académicas, científicas e investigativas, se encuentra desvinculada; dificultando su disponibilidad para los nuevos procesos de investigación que realizan estudiantes y docentes. Esto se agudiza aún más cuando se requiere conocer detalles significativos de la información compilada. Antes esta problemática, las ontologías y los metadatos juegan un papel fundamental, y, constituyen la mejor vía (conceptual e instrumental) para poder investigar y definir, deducir y razonar, conocimiento institucional, en el dominio de un almacenamiento web sobre la base de las propias producciones científicas.*

*El objetivo de esta línea de investigación es iniciar los estudios en la UNCAus sobre repositorios institucionales para modelar, desarrollar e implementar un repositorio ontológico, sobre la plataforma virtual educativa de la Universidad, dirigido a los estudiantes y docentes que se inician en el campo de la investigación científica. Este repositorio incluirá trabajos de investigación, publicaciones, proyectos de grado y posgrado que se desarrollan a diario en la Universidad. Se pretende mostrar que estos elementos pueden expresarse como servicios innovadores para el usuario académico, otorgando valor agregado a la proliferación de material científico y su dispersión – desarticulación en el contexto de todas las áreas institucionales.*

**Palabras clave:** *Ontologías, Repositorio Institucional, Metadatos*

### Contexto

*Los repositorios institucionales en Argentina están en un momento de crecimiento y desarrollo (Volder, 2008). La creación de estos servicios de información científica en ambiente digital e interoperabilidad dedicado a la producción y difusión científica y/o académica es una experiencia bastante reciente. En general, han surgido como proyectos aislados, con criterios propios y normalizaciones heterogéneas respecto de un intercambio flexible de datos. Hasta octubre de 2011, es posible contar 18 repositorios pertenecientes a universidades, con una gran disparidad en cuanto a niveles de descripción en la terminología utilizada.*

*La línea de investigación que aquí se presenta se enmarca en el proyecto denominado “Repositorio Ontológico Institucional”, fue presentado a la convocatoria 2012-2013 de la Secretaría de Ciencia y Tecnología de la Universidad Nacional del Chaco Austral (UNCAus). Este proyecto, evaluado externamente y financiado por dicha institución, abre una nueva línea de investigación institucional. En este proyecto se busca crear modelo de organización de un repositorio científico universitario, sobre la base de metadatos y ontologías, con el objeto de mejorar la interoperabilidad y la usabilidad de los recursos de conocimiento generados dentro de la institución.*

## Introducción

Un fenómeno bastante común en las universidades, y del cual la Universidad Nacional del Chaco Austral (UNCAus) no es una excepción, es la desarticulación de los sujetos (estudiantes avanzados, tesis o becarios) que se inician en la investigación, respecto a otras producciones, posiblemente contiguas, que se llevan adelante dentro de la misma universidad. Frente a un inventario de las investigaciones heterogéneas producidas en una universidad durante un cierto lapso, resulta prácticamente imposible definir algún parentesco programático entre dos o más de los trabajos de investigación allí reseñados. Es como si cada trabajo constituyera un mundo aparte, totalmente independiente de los demás trabajos de investigación.

Por otra parte, si nos preguntamos cuál es el destino de los productos investigativos, podríamos responder que van a las revistas indexadas, a las comunidades científicas, pero no a los mismos ámbitos universitarios

Los usuarios pasan gran cantidad de tiempo para vincular documentos diseminados en la Web sin encontrar relaciones semánticas que involucren el dominio completo del problema (Koper, 2004). Recién al localizar las semejanzas y las diferencias entre fragmentos de información, éstas se trasladan a su trabajo y sirven para la creación de nuevo conocimiento. Una forma de aprovechar este conocimiento, es contar con una memoria organizacional que lo administre y almacene en forma eficaz y eficiente.

El surgimiento de la Web Semántica ha permitido obtener soluciones a problemas habituales en la búsqueda de información gracias a la utilización de una infraestructura común, mediante la cual, es posible compartir, procesar y transferir información de forma sencilla. Una de las herramientas que ofrece la web semántica para describir un dominio y facilitar las búsquedas son las ontologías. Estas poseen un enorme potencial para proporcionar homogeneidad entre recursos heterogéneos de acceso común (Castells, 2003).

De acuerdo con Gruber (1993) una ontología es una especificación formal y explícita de una conceptualización compartida, en donde propiedades, relaciones y funciones categorizan objetos de conocimiento y especifican acciones dentro de un dominio específico. No se puede describir una ontología sin mencionar a los metadatos que constituyen definiciones que proveen de un vocabulario común dentro del contexto de estudio, proporcionando estrategias y mecanismos formales para la búsqueda de información, interoperabilidad y reutilización. (Pedraza-Jiménez, et al., 2007).

A partir de ello, se presenta en este trabajo un modelo basado en metadatos y ontologías para describir una colección centralizada de información institucional. Esto es un esquema organizado de proyectos de investigación institucionales, concluidos y en ejecución, resultados parciales y finales de los mismos, tesis de grado y posgrado, para compartir y reutilizar.

## Descripción de la línea de investigación

En esta línea de investigación se propone abordar la construcción de un modelo de repositorio institucional. Su finalidad se resume en: 1) Compartir el entendimiento común de la estructura de información referida a proyectos de investigación, sus resultados, así como tesis de grado y posgrado que se desarrollan en el marco universitario, 2) Facilitar la búsqueda y recuperación de información científica local a docentes, investigadores y estudiantes, 3) Inferir información acerca de preguntas complejas relacionadas con sujetos que están trabajando sobre un problema o campo común, publicaciones que sustentan esta información, datos y variables para para sustentar una idea, entre otros.

A partir de ello, se plantean los siguientes objetivos específicos:

1. Modelar, desarrollar e implementar un repositorio ontológico, para la plataforma virtual educativa de la UNCAus.



2. Favorecer el desarrollo de conocimiento científico-tecnológico en el campo de las ontologías y repositorios web.
3. Contribuir a la formación y capacitación de investigadores
4. Alcanzar un mayor entendimiento de los modelos ontológico y los paradigmas de la semántica web.

Para lograr estos objetivos, se realizó en primer lugar una búsqueda, selección, traducción y lectura comprensiva de bibliografía y material de referencia sobre modelos ontológicos, objetos de aprendizaje y paradigmas de la semántica web, que aporten un marco de referencia a la investigación. Paralelamente, se relevaron los contenidos producidos por los usuarios existentes en la institución; identificando: formatos de trabajo, áreas temáticas, nivel de importancia, formas de búsqueda, tipos de archivos, finalidad de los recursos, software y hardware utilizado, tamaños en bytes de contenidos. Luego estos recursos fueron normalización y clasificados, y se procedió a la modelización inicial de los metadatos.

Actualmente se está trabajando en la identificación de requerimientos del sistema de información web, con el fin de caracteriza los servicios de búsqueda, recuperación, navegación, anotación, colaboración y visualización de la información contenida en el repositorio, que se podrá ofrecer a los usuarios. Con posterioridad se prevé el análisis y diseño del sistema de información web de gestión de servicios sobre el repositorio, para luego concretar la construcción del repositorio y el desarrollo del sistema de información web que permita la gestión sobre el repositorio. Finalmente se prevé el diseño de políticas institucionales para el uso del repositorio: acceso, estructuración jerárquica de contenidos, manejo y transferencia de contenidos, metabúsquedas, preservación, interoperabilidad.

El modelo de repositorio ontológico institucional, permitirá deducir y razonar nuevo conocimiento e investigación, realizar clasificaciones apoyadas por procesos de inferencia y generar un vocabulario común

para facilitar la comunicación entre sistemas y sujetos utilizando como partida los productos científicos que se realizan en el dominio universitario.

## Resultados alcanzados y esperados

Los resultados esperados a partir de la ejecución del proyecto son:

- El desarrollo de un prototipo de repositorio ontológico, con conceptos asociados a significados y lógicas de inferencia basadas en la automatización y acceso a colecciones estructuradas de información.
- La transferencia y ofrecimiento de servicios de asesoramiento y capacitación de estudiantes de grado, la realización de trabajos finales, publicaciones en revistas y comunicaciones en eventos científicos.

Respecto de los resultados alcanzados, se destaca la recuperación, organización y clasificación del caudal de conocimiento académico y de investigación en el contexto de la Universidad.

Por otra parte, se ha avanzado en el diseño de la arquitectura básica del sistema de gestión del repositorio ontológico (Zachman et al., 2012) y en el modelo de organización del repositorio científico universitario, sobre la base de metadatos y ontologías (Durán et al., 2012).

Actualmente se está trabajando en la categorización de recursos, clasificación taxonómica, refinamiento del modelo inicial de metadatos y ontologías; como así también en la definición de las especificaciones de diseño para el sistema web, a partir de un acercamiento a las funciones que debería contener el sistema y su viabilidad a la plataforma e-learning de la Universidad.

## Formación de Recursos Humanos

El equipo de trabajo está integrado por el Director, un docente investigador y tres alumnos de los últimos años de la carrera Ingeniería en Sistemas de Información, más un asesor en el tema de ontologías.

El proyecto ha permitido que un docente de la carrera Ingeniería en Sistemas de Información

se inicie en actividades de investigación bajo la dirección de un investigador formado y categorizado en el sistema de incentivos. Por otra parte, se espera que en el transcurso de este año uno de los estudiantes, integrante del proyecto, inicie su trabajo final de graduación en el marco del proyecto de investigación.

investigadores noveles, *Anales VIII Jornadas de Ciencia y Tecnología de Facultades de Ingeniería del NOA, Septiembre de 2012, San Miguel de Tucumán. Libro de Resúmenes pag. 138.*

## Referencias

- Castells, P., *La Web Semántica. Escuela Politécnica Superior, Universidad Autónoma de Madrid. Conferencia impartida en el Curso de Verano sobre Interacción en la Web, Universidad de Castilla - La Mancha, Puertollano, Septiembre 2003.*
- Durán, E., Zachman, P., Aguilera, J., Nuñez, J. Umaño, M., "Repositorio Científico Institucional basado en Metadatos y Ontologías", *Proceeding Primer Congreso Argentino de la Interacción-Persona Computador@, Telecomunicaciones, Informática e Información Científica (IPCTIIC 2012), Francisco V. C. Ficarra y Miguel C. Ficarra eds.. Publicado por Blue Herons editions, Cordoba, Argentina, Diciembre 2012.*
- Gruber, T., *Toward Principles for the Design of Ontologies Used for Knowledge Sharing, 1993*
- Koper, R., *Use of the Semantic Web to Solve Some Basic Problems in Education: Increase Flexible, Distributed Lifelong Learning. Decrease Teachers' Workload. Journal of Interactive Media in Education. Special Issue on the Educational Semantic Web, 2004*
- Pedraza-Jiménez, R., Codina, L. & Rovira, C. *Web semántica y ontologías en el procesamiento de la información documental. El profesional del la Información 16, 10, 2007.*
- Volder, C. D, *Los Repositorios De Acceso Abierto En La Argentina Situación Actual[Open Access Repositories In Argentina: An Update] Universidad de Buenos Aires. Facultad de Filosofía y Letras. Instituto de Investigaciones Bibliotecológicas (INIBI), No. 19, 5. 2008*
- Zachman, P., Durán, E. y Aguilera, J., *Arquitectura de un sistema de gestión de repositorio ontológico institucional de apoyo a*

## Conceptualización, modelado e implementación experimental de un framework sensible al contexto

de Jesús Sebastián; Guasti Hernán; Panizzi Marisa.

Facultad de Informática, Ciencias de la Comunicación y Técnicas Especiales, Universidad de Morón  
Cabildo 134 – Morón (CP 1708)  
Tel: 5627-2000

[sadejesus@yahoo.com.ar](mailto:sadejesus@yahoo.com.ar); [hguasti@gmail.com](mailto:hguasti@gmail.com); [marisapanizzi@speedy.com.ar](mailto:marisapanizzi@speedy.com.ar)

### Resumen:

El presente trabajo de desarrollo pretende conceptualizar, modelar e implementar un prototipo de marco de trabajo y un prototipo de aplicación que interactúe con él. Para su construcción se considerarán los principios e hipótesis de las teorías de la logística de la información y los sistemas ubicuos, en términos de los postulados de M. Weiser y posteriores aportes de A. Dey, S. Busse, P. Brown y H. Chen, entre otros. Su aplicación será en un contexto de bajo coste y reusabilidad, conforme a la apropiación de tecnologías y paradigmas de aplicación inmediata.

En una primera etapa del proyecto, se realizará el análisis, diseño y desarrollo de un prototipo de marco de trabajo experimental para validar los principios enunciados anteriormente. Una vez realizada la validación, en una segunda etapa del proyecto, se desarrollará una aplicación para interactuar con el marco de trabajo resultante de la primera etapa. Por último, se planificarán y desarrollarán los casos de prueba correspondientes para la validación de la solución propuesta.

### Palabras clave:

*Framework/ Computación Ubicua/ Sensibilidad al Contexto/ Logística de la Información/ Dispositivos Móviles/ Detección Posicional/ Redes Inalámbricas.*

### Contexto.

El tema de estudio y su proyección como objeto de investigación surge como propuesta orientativa de la Mg. Marisa Panizzi de la Facultad de Informática, Ciencias de la Comunicación y Técnicas Especiales de la Universidad de Morón en el marco del desarrollo del trabajo de fin de carrera.

Dentro del enfoque del proyecto podemos enunciar la apropiación de tecnologías, la exploración de paradigmas noveles y su aplicación en el ámbito pedagógico y académico mediante la producción de un artefacto de software esencial reutilizable que sirva de base para el desarrollo eficiente y ágil de soluciones puntuales en el área de los sistemas ubicuos y sensibles al contexto.

### Introducción.

Luego del extracto planteado en el resumen, se desprende que los objetivos del presente trabajo están centrados esencialmente en identificar un conjunto de tecnologías que permitan ensayar la funcionalidad del paradigma en estudio. Este conjunto de tecnologías debe de ser asequible en cuanto no requiera recursos extraordinarios, tanto tecnológicos como económicos, sino más bien, de los presentes y disponibles en la vida cotidiana de los usuarios.

Ello implica, si se quiere, una extensión del requerimiento funcional de mantenerse por debajo del “nivel de conciencia del usuario”, no sólo liberándolo de la exigencia de desarrollar una iniciativa de interacción con el contexto, sino también de la obligación de incorporar para tal fin, dispositivos “ad hoc” adicionales.

De la revisión de antecedentes sobre sensibilidad al contexto, computación ubicua y logística de la información, las líneas de investigación se basaban usualmente la interacción con el ambiente en el empleo de dispositivos complementarios adicionales, tales como etiquetas de identificación por radiofrecuencia “RFiD” (Ierache Jorge, 2007), tarjetas de banda magnética (magnetic stripe cards), códigos de barras unidimensionales (bar code) y multidimensionales (Datamatrix, Quick Response, etc.), memorias de contacto (contact memory), tarjetas inteligentes (integrated circuit cards), tokens, etc.

Cualquiera de estas estrategias termina convergiendo naturalmente a asimilarse con las estrategias de autenticación de un sistema de control convencional, por cuanto que resume los factores tradicionales de conocimiento (“algo que el usuario sabe”), de posesión (“algo que el usuario tiene”) o de inherencia (“algo que el usuario es”), y con ello, se dificulta sensiblemente exponer con claridad el potencial que expone el paradigma: desarrollar suficientemente, más que un simple entorno reactivo, el despliegue de una auténtica capacidad de inteligencia ambiental.

Existen enfoques alternativos orientados a la aplicación de técnicas y algoritmos de posicionamiento global satelital, que se asimilan con mayor naturalidad a los requerimientos conceptuales del *Everyware*, “*Procesado de la*

*información integrada en los objetos y las superficies de la vida cotidiana*” (Greenfield Adam, 2006).

Sin embargo, los dispositivos móviles con capacidad de gestionar dicha tecnología son específicos para tal fin, y por ende adicionales, o bien integrados a otras funcionalidades (teléfonos celulares, asistentes digitales personales, etc.) de gama superior y coste elevado, con lo cual, se contradicen inmediatamente las consignas planteadas.

Como puede advertirse, la complejidad del objeto de estudio radica en que no se trata tan sólo de determinar, si en las condiciones enunciadas, es factible la implementación de un marco de validación del paradigma, sino también, de intentar poner de manifiesto en el presente documento la significatividad del cambio conceptual que todo esto implica.

Brindar respuesta al planteo conceptual de un marco de validación del paradigma (ubicuo y sensible al contexto), desde la resolución del condicionante operativo (hacer uso de tecnologías asequibles y disponibles), puede considerarse de especial relevancia y significatividad, por cuanto viabilizaría derivar (sin la necesidad de recurrir al concurso de recursos extraordinarios), un sinnúmero de líneas de investigación y desarrollo. Desde la aplicación en ámbitos afines a la integración y estímulo de personas con capacidades especiales, marketing ubicuo (u Holístico), aprendizaje omnipresente (Ubiquitous Learning), realidad aumentada, etc.

En respuesta a las expectativas planteadas, desarrollaremos una solución basada en detección posicional, aplicable a entornos contextuales en escala de ambientes,



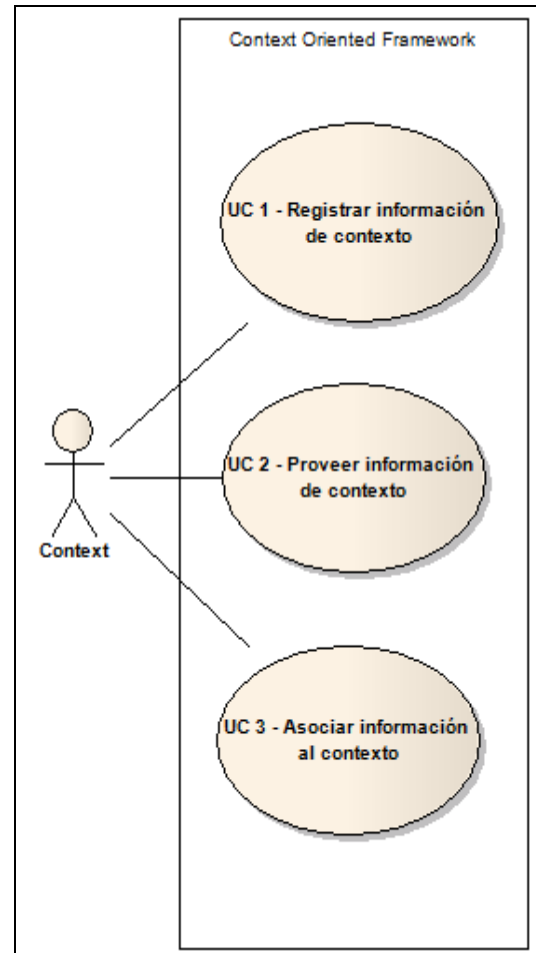
tanto abiertos como cubiertos (aquellos donde la detección geoposicional satelital – GPS - , además de onerosa, es inviable), basándose en el reconocimiento de patrones de presencia de señales correspondientes a áreas de cobertura de redes Wi-Fi.

La solución propuesta consta de un framework o marco operativo, implementado como un servicio en ejecución en un dispositivo personal portable estándar, tal como un teléfono celular inteligente (smartphone) o una tableta (tablet), que provee sensibilidad al contexto y abstracción de su implementación para, en nuestro caso de prueba, proveer a una aplicación el etiquetado virtual del espacio físico, operando sobre la misma plataforma.

La estrategia a ensayar y puesta de manifiesta en la solución, opera relevando los datos intrínsecos y la intensidad de señal de las redes presentes en el espacio eventual al momento de una captura, conformando con ellas, una serie-patrón unívoca para la posición dada, luego, a partir de una consulta sobre un repositorio de datos remoto alojado en la nube de recursos de internet y el correcto aparejamiento con alguna de las series correspondientes a los puntos (a los que llamamos “nodos”) en dicho repositorio registrados y recibidos, mediante un conjunto de algoritmos específicamente estructurados, el framework reporta a la aplicación solicitante la información contextual relevante al sitio de referencia.

Como consecuencia del análisis de lo expresado anteriormente, detectamos básicamente tres requerimientos funcionales sumamente importantes para la gestión del contexto, de los cuales surgen los siguientes casos de uso:

- ✓ UC1 – Registrar información de contexto
- ✓ UC2 – Proveer información de contexto.
- ✓ UC3 – Asociar información al contexto.



**Figura Nro. 1** Caso de uso preliminar para la gestión del contexto.

Para la especificación de los requerimientos de software (ERS) nos basamos en el estándar descrito en el documento de la ERS-IEEE ANSI 830-1998 (IEEE ANSI 830-1998, 1998).

Para el proceso de análisis y desarrollo de la solución propuesta, utilizaremos como marco de referencia el Proceso Unificado (UP), el cual fomenta buenas prácticas pero por sobre todo el desarrollo iterativo (Booch, G., 1998) y el modelado lo haremos bajo los estándares de UML (Larman, C., 2005).

**Resultados Obtenidos/ Esperados.**

Como criterio de éxito del presente trabajo, podemos enunciar la obtención y provisión concreta de un marco de trabajo operativo-funcional estable (artefacto de software modelado e implementado conforme proveer abstracción de la gestión del contexto). El cual en un ámbito de ejecución de un dispositivo naturalmente portable e incorporado al empleo cotidiano, ofrezca el servicio de capacidades y herramientas, para la gestión de un conjunto de parámetros singulares y significativos del entorno. Con los fines de proveer a cualquier aplicación, la plataforma de una estrategia y conjunto de recursos para la incorporación de las potencialidades que el paradigma plantea.

**Formación de Recursos Humanos.**

El equipo está integrado por dos tesis que se encuentran desarrollando su trabajo de fin de carrera tutorado por una docente de la cátedra de Trabajo de Diploma (asignatura de fin de carrera). El presente trabajo se enfoca en un dominio tecnológico incipiente, es posible extender líneas de investigación y desarrollo para ampliar los alcances de nuestra propuesta a diversos escenarios, como por ejemplo:

- La ampliación a otros entornos operativos y plataformas.
- La definición de estándares abiertos para el intercambio de información tanto entre un framework proveedor de contexto y aplicaciones ubicuas y para la serialización y persistencia de la información.
- La escalabilidad de la plataforma para la provisión de información contextual densa: imágenes, videos, sonidos, etc., extendiendo las funcionalidades conforme posibilite la

implementación de las aplicaciones del paradigma ubicuo: realidad aumentada, aprendizaje móvil, etc.

**Bibliografía.**

Booch Grady, M. R. (1998). Object Oriented Analysis and Design With Applications (2a edición ed.). Addison-Wesley.

Busse, Susanne; Deiters, Wolfgang; Fuchs-Kittowski, Frank; Lienemann, Carsten; Neuhaus, Jan; Pfennigschmidt, Stefan; Wojciechowski, Manfred (1999). Information Logistic - Internal Report. Berlin.

Dey, Anind K.; Salber, Daniel; Futakawa, Masayasu; Abowd, Gregory D. (1999). An Architecture To Support Context-Aware Applications.

Dey, Anind K. (2000). Providing Architectural Support for Building Context-Aware Applications - PhD Thesis. Georgia Institute of Technology.

Dey, Anind K. (2001). Understanding and Using Context - Personal and Ubiquitous Computing, Special Issue on Situated Interaction and Ubiquitous Computing. London.

Chen, H.; Finin, T.; Joshi, A. (2003). An Intelligent Broker for Context-Aware Systems. Adjunct Proceedings of Ubicomp 2003.

Weiser, Mark (1991). The Computer for the 21st Century. USA.

Greenfield, Adam (2006). The Dawning Age of Ubiquitous Computing. Berkeley, CA.

Ierache, Jorge.; Campos, Luis.; Bruno, Marcela.; Padovani, Hugo. "Modelo para el entrenamiento de usuarios con déficit auditivo a través integración de sonidos con el contexto". IV Workshop de Ingeniería de Software y Bases de Datos - XIII Congreso Argentino de Ciencias de la Computación. (CACIC 2007). Corrientes-Resistencia. ISBN 950-656-109-3.

IEEE ANSI 830-1998. (1998). IEEE Recommended Practice for Software Requirements Specifications. USA: Institute of Electrical and Electronics Engineers.

Larman Craig. (2005). Applying UML and Patterns: An Introduction to Object-Oriented Analysis and Design and Iterative Development. Pearson Education.

# AVANCES EN EL PROYECTO ARQUITECTURAS DE CONTROL PARA ROBOTS AUTÓNOMOS MÓVILES DIDÁCTICOS BASADAS EN SISTEMAS EMBEBIDOS

Azcurra, D., Rojo, S., Rodríguez, D.

Laboratorio de Sistemas Embebidos y Robótica

Departamento de Ingeniería  
Universidad Nacional de Tres de Febrero  
Valentín Gómez 4752 (1678) Caseros, Tres de Febrero  
Buenos Aires, Argentina

Grupo Investigación en Sistemas de Información

Departamento Desarrollo Productivo y Tecnológico  
Universidad Nacional de Lanús.  
29 de Septiembre 3901 (1826) Remedios de Escalada, Lanús  
Buenos Aires, Argentina.

## Contexto

Este proyecto de investigación integra la línea de trabajo en aplicaciones de sistemas embebidos y robótica en el marco de la carrera de Ingeniería en Computación de la Universidad Nacional de Tres de Febrero y de la Carrera de Licenciatura en Sistemas de la Universidad Nacional de Lanús.

## Resumen

En este proyecto se busca desarrollar y sistematizar el cuerpo de conocimiento de las arquitecturas de control de robots autónomos móviles (RAM) didácticos basadas en sistemas embebidos.

Un sistema embebido es un sistema informático diseñado para realizar un grupo de funciones dedicadas y específicas, empleando para ello una combinación de recursos de hardware y de software. Poseen características diferenciales (entre otras: procesamiento concurrente, paralelo y distribuido, robustez, fiabilidad, bajo consumo y bajo costo) que los hace altamente recomendables en la administración y control de robots autónomos móviles (RAM).

En este trabajo se presentan actividades, logros y objetivos alcanzados a la fecha con relación esta investigación.

**Palabras clave:** robótica, sistemas embebidos, automatización, arquitecturas de control.

## Líneas de investigación y desarrollo

Este proyecto se inscribe en una línea de investigación que busca desarrollar y sistematizar el cuerpo de conocimiento de las arquitecturas de control de robots autónomos móviles (RAM) didácticos basadas en sistemas embebidos.

Entre los supuestos que guían el proyecto se encuentran:

- I. Existe disponibilidad en el mercado (tanto local como internacional) de desarrollos de hardware de robótica que permiten construir prototipos de robots autónomos móviles didácticos. La posibilidad de contar con una amplia gama de sensores, unidades de procesamiento programables, actuadores versátiles e interfaces de usuario que permiten la programación de los prototipos, así como el acceso y administración de sus recursos; allana los aspectos de integración de hardware y controladores en la construcción de la mecánica del robot y permite focalizar el proyecto de investigación en los sistemas informáticos que los controlaran.
- II. Existen arquitecturas de sistemas informáticos que destacan en importancia en su utilización en robótica autónoma. Dentro de las mismas, las arquitecturas de sistemas embebidos poseen características específicas y diferenciales (entre otras: procesamiento concurrente, paralelo y distribuido, robustez, fiabilidad, bajo consumo y bajo costo) que las hacen



sobresalir para ser empleadas en la administración y control de robots autónomos móviles.

**III.** Es factible la caracterización de arquitecturas de control flexibles, sencillas y tolerantes a fallos basadas en sistemas embebidos para que un RAM pueda desplazarse de forma autónoma evitando colisiones.

## Resultados y Objetivos

El estado actual del proyecto, luego del primer año de trabajo, puede sintetizarse en:

- Se avanzó en la actividad de relevamiento y análisis de alternativas de arquitecturas de sistemas embebidos disponibles en mercado. El resultado de esta actividad encuentra resumido en una matriz de comparación de las diferentes posibilidades.
- Se adquirieron kits didácticos basados en microcontroladores de 8 bits sobre los que se están haciendo ensayos, y se han seleccionado otros que serán adquiridos y evaluados en las futuras etapas del proyecto. Se está a la espera de las futuras cuotas de subsidio para realizar la adquisición.
- Se avanzó en la actividad de relevamiento y análisis de las alternativas disponibles en el mercado local de prototipos de robot autónomo móvil para ser administrado y controlado por arquitecturas de sistemas embebidos.
- Se adquirió un prototipo de robot autónomo móvil (y un conjunto de sensores adicionales para ampliar posibilidades de adquisición de datos del

ambiente) y se han seleccionado los otros kits que se prevé adquirir.

- Atendiendo a lo previsto en cuanto a transferencia a la actividad docente, se comenzó a dictar la asignatura Diseño de Sistemas Embebidos, materia optativa de quinto año en la carrera de Licenciatura en Sistemas de la UNLa. Asimismo, en el marco de la asignatura Arquitectura de Computadoras de la carrera de Ingeniería en Computación del a UNTREF se realizaron trabajos prácticos sobre kits de microcontroladores de 8 bits, uno de los cuales resultó publicado.
- En el marco del proyecto, un alumno de la carrera Licenciatura en Sistemas de la UNLa está desarrollando su Trabajo Final. El mismo se titula: *Control de Robots Autónomos Móviles Basado Arquitecturas de Sistemas Embebidos. Propuesta de Modelo de Aplicación y Uso Potencial.*
- Se colaboró con el stand de la carrera de Licenciatura en Sistemas de la UNLa que participó de la Feria del Libro y la ExpoCarreras 2012 – de la misma Universidad- con un prototipo robot que realizaba distintas actividades en una pista diseñada al efecto.

## Formación de Recursos Humanos

Al grupo de trabajo, formado por un investigador formado y un estudiante avanzado, se le ha incorporado un becario alumno con beca de universidad.

Asimismo, en el marco del proyecto, un alumno de la carrera Licenciatura en Sistemas de la UNLa está desarrollando su trabajo de fin de carrera, titulado: *Control de Robots Autónomos Móviles Basado Arquitecturas de Sistemas*

*Embebidos. Propuesta de Modelo de Aplicación y Uso Potencial..*

## Referencias

J. González E.; B. Jovani A. Jiménez ; (2009); “LA ROBÓTICA COMO HERRAMIENTA PARA LA EDUCACIÓN EN CIENCIAS E INGENIERÍA”; Revista Iberoamericana de Informática Educativa, 10: 31-36; ISSN: 1699-4574

J. Ierache, M. Bruno, M Dittler, N. Mazza ; (2008); “ROBOTS Y JUGUETES AUTÓNOMOS UNA OPORTUNIDAD EN EL CONTEXTO DE LAS NUEVAS TECNOLOGÍAS EN EDUCACIÓN”; Proceedings VII Ibero-American Symposium on Software Engineering; 371-379

G. Zabala; “ROBOLIGA – ROBÓTICA EDUCATIVA EN LA ARGENTINA”; [http://caeti.uai.edu.ar/archivos/211\\_ROBOTICA\\_EDUCATIVA\\_EN\\_LA\\_ARGENTINA\\_-\\_ROBOLIGA.PDF](http://caeti.uai.edu.ar/archivos/211_ROBOTICA_EDUCATIVA_EN_LA_ARGENTINA_-_ROBOLIGA.PDF)

J. M. Molina López, V. Matellán Olivera; (1996) “ROBOTS AUTONOMOS: ARQUITECTURAS y CONTROL” Buran, 7: 19-24; ISSN 1698-7047

S. Geninatti, G. Gennai, G. Minnucci, S. Roatta y L. Hendryk; (2008) “EL ABORDAJE DE LOS SISTEMAS EMBEBIDOS MEDIANTE UN HARDWARE DIDÁCTICO”; 1ª Jornadas de Experiencias Innovadoras en Educación en la FCEIA; <http://www.fceia.unr.edu.ar/fceia/1jexpinnov/trabajos%20pdf/Geninatti-Gennai-Roatta-Hendryk.pdf>

J. Santa Lozano, M. A. Zamora Izquierdo, B. Úbeda Miñarro; (2008) “EL APRENDIZAJE BASADO EN PROYECTOS EN MATERIAS DE INGENIERÍA INFORMÁTICA Y SUS IMPLICACIONES”; I Jornadas sobre nuevas tendencias en la enseñanza de las ciencias y las ingenierías; [http://www.murciencia.com/UPLOAD/COMUNICACIONES/el\\_aprendizaje\\_basado\\_proyectos.pdf](http://www.murciencia.com/UPLOAD/COMUNICACIONES/el_aprendizaje_basado_proyectos.pdf)

Azcurra, D., Rojo, S., Rodríguez, D. (2012). "Arquitecturas de Control para Robots Autónomos Móviles Didácticos Basadas en Sistemas Embebidos". Proceedings del XIV Workshop de Investigadores en Ciencias de la Computación. Pág. 669-672. ISBN 978-950-766-082-5.

## SISTEMAS INTELIGENTES EN ARQUITECTURAS DE MOTORES PARA VIDEOJUEGOS

Merlino, H., Pytel, P., Rodríguez, D., Cartanilica, A., Caracciolo, B., García-Martínez, R.

Grupo Investigación en Sistemas de Información

Departamento de Desarrollo Productivo y Tecnológico. Universidad Nacional de Lanús  
29 de Septiembre 3901 (1826) Remedios de Escalada, Lanús. Argentina. Tel +54 11 6322-9200 Ext. 194  
rgarcia@unla.edu.ar

### RESUMEN

La industria de productos lúdicos informatizados (más conocidos como videojuegos) es una de las actividades económicas de mayor crecimiento en los últimos años. Durante el 2006, en los Estados Unidos los ingresos por videojuegos excedieron por primera vez en la historia a los del cine. Sin embargo, a pesar del auge en este mercado, todavía existen más ofertas de empleo que personas preparadas para ocuparlos.

Para lograr el desarrollo de un videojuego se requiere de diversos conocimientos, como ser, diseño multimedial, manejo de lenguajes de programación específica, uso de plataformas de actividades lúdicas, entre otros; sumado a estas actividades es necesario dotar al videojuego con un grado de inteligencia que lo haga no determinista; logrando que los jugadores mantengan durante una mayor cantidad de tiempo el interés por el mismo, pues de no ser así, los jugadores solo lo utilizarían hasta llegar hasta comprender la lógica de funcionamiento y perderían el interés por el videojuego.

Es por esto que es de interés para la industria del video juego el desarrollo de motores basados en sistemas inteligentes tomando la experiencia adquirida en otros dominios, como ser, robótica, minería de datos y control de procesos, haciendo las adaptaciones necesarias al dominio en cuestión.

**Palabras clave:** software lúdico, motores para videojuegos, sistemas inteligentes.

### CONTEXTO

El el año 2004, el Estado Nacional promulgó la Ley 25.922 (Ley de Promoción de la Industria del Software - Definición, Ámbito de Aplicación y Alcances) y el año pasado (CESSI, 2012) fue prorrogada la citada norma hasta el año 2019. En este contexto, el Fondo Fiduciario de Promoción de la Industria del Software del Ministerio de Ciencia y Tecnología, en su accionar para la consolidación de empresas existentes y contribuir a la generación de

nuevas empresas comerciales dentro del sector, ha señalado como una de las áreas estratégicas la de desarrollo de videojuegos (FONSOFT, 2012).

La industria del video juego en la Argentina es un área creciente de la industria de software que exporta el 90% de su producción (AEN, 2012), lo que la señala como un área promisoría de inversiones. La incorporación de tecnología de sistemas inteligentes en los motores de videos juegos hace que tengan un nivel de “jugabilidad” mayor (Millington y Funge, 2009) al que se podría obtener si no se utilizara.

La UNLa dispone de docentes-investigadores que acreditan experiencia de más de una década en la aplicación de Sistemas Inteligentes a problemas concretos, lo que la posiciona como una Institución con capacidades de generar conocimiento que de mayor valor agregado a este sector de la Industria.

### INTRODUCCIÓN

En los últimos años se ha producido un incremento en el interés de la comunidad científica por el estudio e investigación de los sistemas inteligentes aplicados al desarrollo de videojuegos, esto se debe en parte al auge que ha tenido el desarrollo de videojuegos [González. 2011]. Esto ha producido que el desarrollo de arquitectura para la construcción de videojuegos sea uno de los focos de interés para la industria informática, una lista de los diferentes motores categorizados por sus características se puede encontrar en [Wikipedia, 2012]. En este sentido no se puede dejar de lado a un libro que ha sido paradigmático en el estudio de los diferentes aspectos a tener en cuenta en la construcción de arquitecturas de videojuegos como ser Game Engine Architecture [Gregory, 2004], aquí se tratan temas de estilos de programación a el arte necesario para la construcción de videojuegos pasando por el diseño del sonido del mismo.

En el área del desarrollo de motores de sistemas inteligentes para videojuegos, es de apreciar la proliferación de bibliografía relacionada a la problemática, como ser [Millington y Funge, 2009][Schwab, 2004][Fung, 2004][Thompson,

2008][Buckland, 2004][Mark, 2009][González-Calero y Gómez-Martín, 2011], que abarcan aspectos teóricos y prácticos relacionados con los sistemas inteligentes dando también lineamientos de las características que debería tener un motor de sistemas inteligentes.

Entre las tecnologías de Sistemas Inteligentes aplicables a la industria de videojuegos se pueden citar:

**Sistemas Expertos:** Se pueden definir los Sistemas Expertos [SE] como una clase de programas que son capaces de [García-Martínez y Britos, 2004]: aconsejar, categorizar, analizar, comunicar, consultar, diseñar, diagnosticar, explicar, explorar, formar conceptos, interpretar, justificar, planificar; son en suma, programas capaces de manejar problemas que normalmente requieren para su resolución la intervención humana especializada. Las siguientes características guían el diseño [aunque no siempre obtenibles] de los sistemas expertos: [a] aplican su experiencia de una manera eficiente para solucionar problemas, pudiendo realizar inferencias a partir de datos incompletos o inciertos, [b] explican y justifican lo que están haciendo, [c] se comunican y adquieren nuevos conocimientos, [d] reestructuran y reorganizan el conocimiento, [e] pueden quebrantar reglas, es decir, interpretan simultáneamente el espíritu y la letra de de las mismas, [f] determinan cuando un problema está en el dominio de su experiencia, conocido como determinación de la relevancia del problema.

**Algoritmos TDIDT:** Estos algoritmos [TDIDT - Top Down Induction Decision Trees] pertenecen a los métodos inductivos del Aprendizaje Automático que aprenden a partir de ejemplos preclasificados [Quinlan, 1986]. A esta familia pertenecen los algoritmos el ID3, C4.5 y C5. Estos algoritmos generan árboles y reglas de decisión a partir de ejemplos preclasificados. Para construir los árboles se utilizan el método de aprendizaje automático basado en la estrategia propuesta por Hunt en [Hunt et al., 1966], que particiona el conjunto de ejemplos en subconjuntos a medida que avanza; trabajar sobre cada subconjunto es más sencillo que trabajar sobre el total de los datos.

**Redes Neuronales BP:** Son redes formadas por múltiples capas lo que les permite resolver problemas que no son linealmente separables. Pueden ser totalmente o localmente conectadas. En el primer caso cada salida de una neurona de la capa "i" es entrada de todas las neuronas de la capa "i+1", mientras que en el segundo caso, cada neurona de la capa "i" es entrada de una serie de neuronas [región] de la capa "i+1". Utilizan un algoritmo de

aprendizaje llamado regla delta generalizada [ó regla de retropropagación del error], que consiste en minimizar el error [comúnmente cuadrático] por medio del método del gradiente descendente en los parámetros de entrenamiento de la red neuronal [Freeman y Skapura, 1991; Hilera y Martínez, 1995]. Estas redes son conocidas como redes de retropropagación [Redes BP].

**Redes Neuronales SOM:** Los mapas auto organizados o SOM [Self-Organizing Map], también llamados redes de Kohonen [1995] son un tipo de red neuronal no supervisada competitiva, con capacidad para formar mapas bidimensionales de características a partir del principio de formación de mapas topológicos. Se orientan a descubrir la estructura subyacente de los datos ingresados a partir de establecer características comunes entre los vectores de información de entrada a la red. A lo largo del entrenamiento de la red; los vectores de datos son introducidos en cada neurona y se comparan con el vector de peso característico de la misma. La neurona que presenta menor diferencia entre su vector de peso y el vector de datos es la neurona ganadora [o BMU] y ella y sus vecinas verán modificados sus vectores de pesos.

**Redes Bayesianas:** Las redes bayesianas o probabilísticas se fundamentan en la teoría de la probabilidad y combinan la potencia del teorema de Bayes con la expresividad semántica de los grafos dirigidos; las mismas permiten representar un modelo causal por medio de una representación gráfica de las independencias / dependencias entre las variables que forman parte del dominio de aplicación [Pearl, 1988; Lauría y Duchéis, 2006]. Se puede interpretar a una red bayesiana de dos formas: [a] distribución de probabilidad en la que representa la distribución de la probabilidad conjunta de las variables representadas en la red, ó [b] base de reglas en la que cada arco representa un conjunto de reglas que asocian a las variables involucradas y están cuantificadas por las probabilidades respectivas.

**Algoritmos Genéticos:** Son métodos de optimización, en los que aquella variable o variables que se pretenden optimizar junto con las variables de estudio constituyen un segmento de información [Goldberg, 1989; Sivanandam y Deepa, 2008]. Aquellas configuraciones de las variables de análisis que obtengan mejores valores para la variable de respuesta, corresponderán a segmentos con mayor capacidad reproductiva. A través de la reproducción, los mejores segmentos perduran y su proporción crece de generación en generación. Se puede además introducir elementos aleatorios para la modificación de las variables [mutaciones]. Al cabo



de cierto número de iteraciones, la población estará constituida por buenas soluciones al problema de optimización, pues las malas soluciones han ido descartándose, iteración tras iteración.

Este interés suscitado sobre los aspectos específicos de los sistemas inteligentes en los videojuegos ha permitido la creación de una conferencia anual sobre este área de dominio, denominada GameAI Conference [GameAI, 2012], en la que se exponen sobre el uso aplicaciones originales de sistemas inteligentes en videojuegos. También se ha desarrollado un sitio de internet dedicado a la problemática antes mencionada AIGameDev [AIGameDev, 2012], en él se tratan y publican artículos de interés para la comunidad como también guías de buenas practicas y foros de discusión. También se han desarrollado librerías específicas para diferentes aspectos de sistemas expertos como ser el manejo de movimientos tales como en Libbehavior [2012]. En relación a motores de sistemas inteligentes para videojuegos, los avances han sido muy recientes con lo cual solo es posible referir AISandBox [AISandBox, 2012] un entorno de trabajo que se encuentra, al momento de presentación de este proyecto, en sus últimas etapas de pruebas.

### OBJETIVOS E HIPOTESIS DE INVESTIGACION

La pregunta problema que anima la investigación es: ¿Cuál es la mejor aproximación para la integración de un motor basado en tecnología de sistemas inteligentes dentro de la arquitectura de un videojuego?

*Hipótesis I:* La tecnología de sistemas inteligentes tiene la madurez suficiente para explorar aplicaciones en la industria de videojuegos.

*Hipótesis II:* La actual tecnología utilizada en motores de videojuegos no está basada en tecnología de sistemas inteligentes.

*Objetivo General:* El objetivo de este proyecto es sistematizar el cuerpo de conocimiento existente, y desarrollar el faltante, sobre el uso de tecnología de sistemas inteligentes en motores de videojuegos, con el propósito de establecer las bases que permitan a la industria de videojuegos incorporar el comportamiento no determinista en sus productos.

Objetivos Específicos:

1.- Sistematizar y desarrollar el conocimiento de sistemas inteligentes que sea de aplicación en la industria de desarrollo de videojuegos.

2.- Especificar, diseñar y desarrollar un motor de videojuegos de propósito general, basado en tecnología de sistemas inteligentes.

### METODOLOGÍA DE TRABAJO

Para el Objetivo Específico 1 se propone: (i) realizar investigación documental sobre sistemas inteligentes y su aplicación a la industria de videojuegos; (ii) identificar en la literatura casos de estudio de arquitecturas de motores de videojuegos; (iii) identificar componentes de la arquitectura que sean potenciadores con tecnología de sistemas inteligentes, (iv) desarrollar versiones basadas en tecnología de sistemas inteligentes de los componentes identificados, (v) hacer pruebas de performance de los componentes desarrollados.

Para el Objetivo Específico 2 se propone: con base en los componentes de la arquitectura potenciadores con tecnología de sistemas inteligentes identificados en el Objetivo Específico 1: (i) desarrollar mediante la metodología de prototipado evolutivo un arquetipo de motor de videojuego basado en las tecnologías de sistemas inteligentes aplicables; y (ii) realizar pruebas de concepto que validen la arquitectura propuesta.

### RESULTADOS OBTENIDOS/ESPERADOS

Como resultado de este proyecto, se esperan contar al fin de su desarrollo con arquetipos de motores de videojuego basados en las distintas tecnologías de sistemas inteligentes aplicables.

### FORMACIÓN DE RECURSOS HUMANOS

El grupo de trabajo se encuentra formado por un investigador formado, dos investigadores en formación dos becarios alumnos de la carrera Licenciatura en Sistemas de la UNLa y un supervisor científico. En su marco se desarrollan dos Trabajos Finales de Licenciatura en Sistemas.

### REFERENCIAS

- AEN, 2012. Crece la Industria Nacional de Videojuegos. Argentina en Noticias. [http://www.argentina.ar/\\_es/economia-y-negocios/C7990-crece-la-industria-nacional-de-videojuegos.php](http://www.argentina.ar/_es/economia-y-negocios/C7990-crece-la-industria-nacional-de-videojuegos.php). Página válida al 01/08/2012.
- AIGameDev. <http://aigamedev.com>. Página válida al 01/08/2012.
- AISandBox Framework de AI para videojuegos. <http://aisandbox.com/>. Página válida al 01/08/2012.
- Buckland, Mat (2004). Programming Game AI by Example. ISBN-13: 978-1556220784. Publisher: Jones & Bartlett.
- CESSI. 2012. <http://www.cessi.org.ar/ver-noticias-telam-la-presidenta-promulgo-la-ley-de-promocion-de-la-industria-del-software-297>. Página válida al 01/08/2012.
- Freeman, J., Skapura, D. (1991). Neural Networks: Algorithms, Applications, and Programming Techniques. Addison Wesley.
- FONSOFT, 2012. Charla sobre subsidios para emprendedores del software. Fondo Fiduciario de Promoción de la Industria del Software (blogspot). <http://fonsoft.blogspot.com.ar/>. Página válida al 01/08/2012.
- Fung, John (2004). Artificial Intelligence for Computer Games. ISBN-13: 978-1568812083. Publisher: A K Peters/CRC Press.
- GameAI Conferencia. <http://gameaiconf.com/>. Página válida al 01/08/2012.
- García Martínez, R. y Britos, P. 2004. Ingeniería de Sistemas Expertos. Editorial Nueva Librería. ISBN 987-1104-15-4.

- García Martínez, R., Servente, M. y Pasquini, D. 2003. *Sistemas Inteligentes*. Editorial Nueva Librería. Buenos Aires. ISBN 987-1104-05-7.
- Goldberg, D. 1989. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison Wesley Publishing Company.
- Gonzalez, Daniel (2011). *Diseño de Videojuegos*. ISBN-13: 978-8499640785. Publisher: Alfaomega Ra-Ma.
- González-Calero, Pedro (Editor), Gómez-Martín, Marco (Editor) (2011). *Artificial Intelligence for Computer Games*. ISBN-13: 978-1441981875. Publisher: Springer.
- Gregory, Jason (2009). *Game Engine Architecture*. Publisher: A K Peters, Ltd.
- Hilera, J., Martínez, J. (1995). *Redes Neuronales Artificiales. Fundamentos, Modelos y Aplicaciones*. Editorial RA-MA.
- Hunt, E., Marin, J., Stone, P. 1966. *Experiments in Induction*. Academic Press.
- Kohonen, T. 1995. *Self-Organizing Maps*. Springer Verlag Publishers.
- Lauria, E., Duchéis, P. 2006. A Bayesian Belief Network for IT Implementation Decision Support. *Decision Support Systems*, 42: 1573-1588.
- Libbehavior. 2012. librería de código abierto. <http://code.google.com/p/libbehavior/>.  
Página válida al 01/08/2012.
- Mark, Dave (2009). *Behavioral Mathematics for Game AI*. ISBN-13: 978-1584506843. Publisher: Course Technology PTR.
- Millington, Ian; Funge, John (2009). *Artificial Intelligence for Games, Second Edition*. ISBN-13: 978-012374731. Publisher: Morgan Kaufmann.
- Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann.
- Quinlan, J. 1986. Induction of decision trees. *Machine Learning*, 1(1): 81-106.
- Schwab, Brain (2004). *AI Game Engine Programming*. ISBN-13: 978-1584503446. Publisher: Charles River Media.
- Sivanandam, S., Deepa, S. 2008. *Introduction to Genetic Algorithms*. Springer-Business Media.
- Thompson, Guy (2008). *AI and Artificial Life for Video Games*. ISBN: 978-1584505587. Charles River Media.
- Wikipedia. List of Game Engines. [http://en.wikipedia.org/wiki/List\\_of\\_game\\_engines](http://en.wikipedia.org/wiki/List_of_game_engines).  
Página válida al 01/08/2012.

# Automatización en la Captura de Datos para el Modelado de Flujo Vehicular en Zonas Urbanas

Julio Monetti, Oscar Leon, Mariana Brachetta, Paola Caymes Scutari,  
Paulo Celeste Sanchez, Alesis Manzano

Universidad Tecnológica Nacional – FRMendoza

Rodriguez 273 – Mendoza

Teléfono: 5244502

jmonetti,oleon,mbrachetta,pcaymesscutari{frm.utn.edu.ar}  
caincelest08@hotmail.com,alesismanzano@gmail.com

## Resumen

El trabajo tiene por objeto establecer metodologías de análisis del flujo vehicular a través de técnicas no convencionales de captura y procesamiento de datos. Se han adoptado tecnologías basadas en GPS para complementar observaciones *in-situ* de las áreas de estudio. Esto permite la recolección automática de datos sobre la circulación de un vehículo en las áreas de estudio, obteniendo así una base de datos con información estadística sobre trayectorias obtenidas por vehículos de prueba. Esta información luego es utilizada por algoritmos que revelan condiciones especiales en el flujo vehicular.

Esta información dinámica es dispuesta sobre mapas previamente digitalizados a partir de algoritmos que establecen la georeferenciación de puntos de interés (como resultan el cruce de calles) o sectorización de los mismos.

**Palabras Claves:** Tránsito Vehicular, GPS, Georeferenciación, Android.

## Contexto

Esta línea de Investigación surge como interés de docentes y alumnos de la carrera Ingeniería en Sistemas de Información de la Universidad Tecnológica Nacional. Oportunamente se presentó un proyecto PID financiado por la UTN que abarca actividades de investigación y desarrollo relacionadas con la ingeniería de tránsito.

## Introducción

En la actualidad, la planificación y diseño de vías de circulación (calles, rutas, autopistas) representa una tarea imprescindible en aquellas áreas urbanas, donde el crecimiento del parque automotor amenaza la dinámica en la circulación de sus habitantes.

La ingeniería de tránsito, como una disciplina que estudia la solución de problemas de circulación [1,2], provee el conocimiento y las herramientas para realizar modelos de tránsito, donde pueden ser simuladas situaciones particulares. Luego, la observación de los componentes del sistema de transporte: vías de tránsito, vehículos, etc, permite describir tales si-

tuaciones, modelar escenarios y proponer soluciones basadas en obras civiles o modificación de políticas de circulación (por ejemplo establecer nuevos tiempos de semáforización).

Este proyecto tiene como objeto de estudio la confección de una metodología estándar que permita automatizar la captura, almacenamiento y procesamiento de datos referidos a la dinámica vehicular en zonas urbanas. Con ello se pretende determinar aquellas variables (endógenas o exógenas) que determinan situaciones de congestión vehicular.

El objetivo principal de la investigación está enfocado principalmente en afianzar conocimientos sobre la dinámica vehicular en ambientes urbanos, y en particular en la provincia de Mendoza, su modelado matemático y posterior presentación de escenarios para la propuesta de soluciones como las mencionadas anteriormente.

Se establece un marco teórico a partir de la observación y el uso de software destinado al modelado vehicular a través de micro y macro modelos de tránsito. Para la aplicación del software resulta necesaria una recolección de datos adecuada que permita calibrar variables, almacenar información estadística, etc. Para ello, el grupo ha indagado previamente el mercado de dispositivos GPS con el objeto de determinar cuál es la mejor forma para obtener datos de un vehículo en movimiento. Como resultado de esta investigación, se han adquirido dispositivos de recolección de datos, para generar información sobre la circulación de vehículos de prueba y posteriormente crear modelos sobre el escenario circundante.

Se han comenzado las actividades con el análisis teórico sobre distintas metodologías de medición y estudio de la dinámica vehicular. Para ello los integrantes del grupo se relacionaron con docentes de la

cátedra "Tránsito y Transporte" de la UTN, quienes a través de ejercicios prácticos (de campo y laboratorio) sobre conteo vehicular, análisis de flujo, dispositivos de medición, etc. exponen cuales son las variables susceptibles de estudio para conformar un modelo de tránsito útil. Las mediciones se basan en técnicas estándares de muestreo y conteo, las cuales son posteriormente informatizadas con el objeto de contar con información fácilmente accesible y suficiente para el modelado de variados escenarios sobre zonas urbanas. Las muestras corresponden a un conjunto de datos en bruto, provenientes de los dispositivos de georeferenciación, que procesados convenientemente, generan información pertinente y útil para describir la dinámica vehicular.

El procesamiento de la información propuesta requiere algoritmos y arquitecturas computacionales que permitan el procesamiento masivo, con el objeto de brindar tiempos de respuesta cualitativamente aceptables. Para ello, el grupo analiza la utilización de técnicas de procesamiento basadas en el cálculo paralelo/distribuido.

De la misma forma, se estudia la posibilidad de almacenamiento masivo, con el objeto de mantener la totalidad de las muestras obtenidas del conteo vehicular. Esto requiere de un adecuado análisis del motor de base de datos que contendrá los mismos.

### **Método del Auto Flotante**

El flujo vehicular puede ser analizado a través de la observación *in-situ* o por cámaras estratégicamente ubicadas. Este tipo de observación permite estudiar condiciones de tránsito en el punto de observación; y aunque es la metodología más utilizada, presenta un gran grado de error en el conteo automático de vehículos, y



resulta inadecuada si se intenta describir las características de circulación a través de una trayectoria.

El método del *auto flotante*, consiste en capturar datos desde un vehículo en movimiento, el cual debe circular junto con el flujo total de vehículos a una velocidad promedio, y en lo posible intentando imitar el comportamiento del resto de los vehículos. La principal variable de medición es la velocidad del vehículo en puntos de estudio particulares. Esta forma de captura de datos es útil cuando se estudia la circulación entre diferentes puntos de la ciudad. En el presente proyecto, todos los estudios están basados en el método del auto flotante, el cual establece una trayectoria  $x_\alpha(t)$  para cada vehículo de prueba  $\alpha$  en un tiempo  $t$ . Luego, es posible representar la trayectoria como un “collar de perlas” [4] o traza (ver figura 1), donde cada elemento (o perla) representa la información puntual de la circulación de un vehículo.



Figura 1. Conjunto de muestras obtenidas por el dispositivo GPS a través de una trayectoria  $x_\alpha(t)$  (collar de perlas).

Cada muestra (ver tabla 1), representada por un punto negro en la figura 1 contiene la información necesaria para realizar un análisis cruzado para diferentes tiempos y puntos de estudio. Para un tramo se cuenta con una determinada cantidad de “per-

las”, obtenidas a través de diferentes vehículos de prueba  $\alpha$  y diferentes tiempos  $t$ .

Id del dato	1
Fecha	12/03/2012
Hora	21:12:05
Velocidad	12
Latitud	S32.89775
Longitud	W68.85394

Tabla 1. Ejemplo de información contenida en una muestra puntual tras la circulación de un vehículo. (Corresponde a un punto particular en la figura 1)

El procesamiento conjunto de todas las muestras permite caracterizar la información almacenada en dos tipos de información principales:

1. *Basada en el Punto*: Para establecer las características de circulación en un punto específico demarcado por cada dato (*latitud, longitud*). Luego, y a partir de numerosas muestras sobre el mismo punto, es posible identificar situaciones anómalas en dicho punto.

2. *Basada en la Sección*: Permite establecer las características de circulación en un tramo que va de esquina a esquina. Este tipo de información permite establecer zonas de máxima aceleración, congestión, semaforización inadecuada, etc. en tramos particulares de una calle.

## Base de Datos

En la estructura de la base de datos se pueden diferenciar dos tipos principales de información:

1. *Información Estática*: Se cuenta con información sobre puntos georeferenciados a lo largo de las zonas de estu-

dio. Estos puntos luego conforman una malla interconectada que puede ser interpretada como un grafo dirigido.

2. *Información Dinámica*: Básicamente compuesta por información sobre la circulación de vehículos de prueba sobre diferentes trayectorias. Los datos corresponden a velocidades promedio, máximas, mínimas por tramo, conjunto de tramos etc. La contrastación de esta información con la malla de información estática permite generar estadísticas sobre la dinámica vehicular por tramos, observar zonas de congestión, etc.

Se debe considerar que cada traza puede contener miles de lecturas; y que para contar con una muestra estadística adecuada se debe contar con gran cantidad de trazas sobre la misma zona de estudio. Esto obliga a almacenar en la base de datos una enorme cantidad de registros (conteniendo entre otros datos, los descritos en tabla 1), los cuales deben ser consultados y procesados en forma conjunta. Luego se torna necesario una ingeniería pormenorizada de la base de datos: sus tablas, índices relaciones, etc. Se utiliza la base de datos *Postgresql* para contener los datos y el lenguaje *SQL (Structured Query Language)* para realizar la consulta de la misma.

### Procesamiento de los Datos

El correcto procesamiento de los datos provenientes de las trazas genera información útil para describir modelos urbanos, sus zonas de congestión, etc. El procesamiento de datos supone entre otras acciones:

1. La clasificación de cada traza de acuerdo al área de estudio.

2. La extracción de datos a partir del archivo de texto plano generado por los dispositivos GPS (*parser*).
3. El filtrado de datos “sucios” o inútiles dentro de cada traza.
4. La obtención de velocidades agrupadas por tramos: máximas, mínimas, promedio, etc.

Cada una de estas acciones se encuentran materializadas como métodos dentro de clases Java. La utilización de dichas clases permite el modelado de un sistema de información integral que permite un análisis pormenorizado de los datos previamente almacenados; permitiendo luego la confección de modelos matemáticos/estadísticos que describan el comportamiento vehicular, y en la medida de lo posible realizar predicciones sobre el estado futuro.

### Líneas de Investigación y Desarrollo

El alcance del presente proyecto consiste en realizar trabajos de investigación del funcionamiento de dispositivos GPS, la aplicación del un sistema de información adecuado para la recolección automática de datos, y la utilización de los sistemas de computación distribuidos para implementar un prototipo de software que informe sobre las características de cada tramo bajo la zona de estudio, de acuerdo las condiciones de tráfico de la ciudad.

Se prevé como línea de investigación futura la investigación tecnológica y posterior desarrollo de un sistema para la transmisión de datos en tiempo real. El funcionamiento del sistema se basa en la transmisión y recepción de datos de la localización de cada un vehículo que integra el sistema. La localización de un vehículo será determinada por tres parámetros, a saber (*longitud, latitud, hora*).

Además, el sistema tendrá en cuenta en todo momento el destino final del conductor. Estos datos, sumado al estudio y aplicación de diversos algoritmos, permitirán trazar una ruta óptima de desplazamiento que permita evitar puntos de mayor densidad vehicular entre el origen y el destino, minimizando el tiempo de desplazamiento y reduciendo los niveles de contaminación ambiental, entre otras cosas.

### Resultados y Objetivos

Actualmente se cuenta con un prototipo del sistema de información anteriormente descrito que permite el reconocimiento de ficheros de texto generados por diferentes dispositivos GPS, el almacenamiento y posterior generación de información.

Por otro lado, se cuenta también con módulos *Java Micro Edition*, los cuales han sido instalados en teléfonos celulares bajo el sistema operativo *Android*. Se han obtenido buenos resultados en la adquisición de datos, lo que permite fijar un nuevo objetivo en la investigación: la transmisión en tiempo real de los datos obtenidos por el auto flotante.

El grupo se plantea también como objetivo a mediano plazo contar con algoritmos que permitan un análisis de congestión a través del establecimiento de centroides o puntos de convergencia.

### Formación de Recursos Humanos

Los nuevos conocimientos adquiridos, se consideran de gran utilidad tanto en el ámbito laboral del Ingeniero en Sistemas de Información e Ingeniero Civil, como así también para aquellos integrantes del grupo de investigación que han mostrado un marcado interés por las actividades de investigación y desarrollo.

Cabe destacar que en el presente proyecto se ponen en práctica todos los conocimientos adquiridos durante el cursado de la carrera de Ing. en Sistemas de Información y Civil, articulando los mismos con los nuevos conocimientos adquiridos.

Esta línea de investigación prevé la formación de estudiantes (futuros ingenieros) en el área de Ingeniería de Tránsito, en una relación estrecha con docentes investigadores que puedan brindar su asistencia y transmitir su experiencia en el área de la investigación y desarrollo de sistemas de información.

### Referencias

1. XIE, Feng & LEVINSON, David. (2011) *Evolving Transportation Networks*. USA. Springer. ISBN: 978-1-4419-9803-3.
2. CAL Y MAYOR, Rafael & CARDENAS, James. (2007): *Ingeniería de Tránsito. Fundamentos y Aplicaciones*. 8° Edición. Alfaomega Grupo Editor. Mexico, DF. ISBN: 970-15-1238-3.
3. KITAMURA, Ryuichi & KUWAHARA, Masao. *Simulation Approaches in Transportation Analysis. Recent Advances and Challenges*. 2005. Springer. USA. ISBN: 0-387-24108-6.
4. WERNER, Huber; LÄEDKE, Michael & OGGER Rainer. *Extended Floating-Car Data for the Acquisition of Traffic Information*. Departamento de Ingeniería BMG AG. Munich, Alemania.

## La Gestión del Conocimientos en las Organizaciones

Marta C. FENEMA<sup>1</sup>, Carola V. FLORES<sup>1</sup>, María V. POLICHE<sup>1</sup>, Juan P. MORENO<sup>1</sup>, Marcela S. MOLINA<sup>1</sup>, Erika E. LOBO<sup>1</sup>

1) Departamento de Sistemas, Facultad de Tecnología y Ciencias Aplicadas, Universidad Nacional de Catamarca, Maximio Victoria 35, (4700), Catamarca, Argentina.

Tel: +54 (383) 4435112

e-mails: cfennema@unsta.edu.ar, {carolaflores, vpoliche, jpmoreno, marcela\_molina, erikalobo}@tecno.unca.edu.ar

### Resumen

Actualmente se habla con frecuencia sobre la necesidad de prestar mayor atención a los activos intangibles de la organización, sobre todo de aquellos capaces de aportar valor económico a la empresa.

En este contexto, el conocimiento se ha convertido en uno de los activos más importantes para las organizaciones, a causa de que su gestión crea riqueza o valores añadidos, que facilitan alcanzar una posición ventajosa en el mercado.

Teniendo en cuenta esta realidad que también se encuentra presente en nuestra región se propone este proyecto que tiene por finalidad diseñar desarrollar, modelar estratégicamente propuestas de cambio que permitan a las organizaciones actuales gestionar el conocimiento.

Para lograr este objetivo es necesario poder vincular las herramientas de TIC con la gestión del conocimiento (GC) y de esta manera fortalecer la gestión.

**Palabras Claves:** Conocimiento, TIC, Gestión del Conocimiento, Gestión de Proyectos.

### Contexto

Este proyecto de investigación aplicada pretende abordar la GC en las organizaciones, ya que se considera al conocimiento como una de los más importantes recursos que posee una organización y debe ser

administrado correctamente para tener éxito en este medio tan cambiante y competitivo.

Este proyecto de investigación fue aprobado por la Secretaria de Ciencia y Tecnología (SCyT) de la Universidad Nacional de Catamarca (UNCa), Argentina. Este proyecto se inicia en Enero de 2013 y su fecha de finalización es en Diciembre de 2014 y será financiado por la Secretaria de Ciencia y Tecnología (SCyT) de la UNCa

### Introducción

Desde la década del 90 las organizaciones se dieron cuenta de que sus activos físicos y financieros no tienen la capacidad de generar ventajas competitivas sostenibles en el tiempo, y descubren que los activos intangibles son los que aportan verdadero valor a las organizaciones. La GC es, en definitiva, la gestión de los activos intangibles que generan valor para las organizaciones. La mayoría de estos intangibles tienen que ver con procesos relacionados de una u otra forma con la captación, estructuración y transmisión de conocimiento. Por lo tanto, la GC tiene en el aprendizaje organizacional su principal herramienta.

En este sentido, se entiende a la GC como “El proceso sistemático de buscar, organizar, filtrar y presentar la información con el objetivo de mejorar la comprensión de las personas en una específica área de interés” (Davenport y Prusak, 1998).

Los objetivos de la gestión del conocimiento son (NIEVES LAHABA, 2001):

- Incrementar las oportunidades de negocio.



- Aumentar la comunicación.
- Aumentar la competitividad presente y futura.
- Elevar el liderazgo de las empresas en su mercado.
- Elevar el rendimiento.

Las bases de conocimiento son el instrumento para la GC, para el razonamiento, la representación de procesos y el procesamiento de ideas.

Las bases de conocimiento tienen asociadas las funciones para el análisis, la simulación y las búsquedas de diverso tipo: búsqueda semántica, textual, por categoría, por similitud, en estructuras de datos de diverso tipo. La base de conocimiento, por su interfaz gráfica (redes semánticas o conceptuales), es la plataforma consistente para el ejercicio de la creatividad y de la innovación, y para el ejercicio de la inteligencia de negocios (la business intelligence).

A menudo para el analista, gestor del conocimiento o dirigente es más conveniente asociar sus propias estructuras de datos a la base de conocimiento para realizar los propios cálculos, análisis, búsquedas y simulaciones. Esto le permite trabajar con datos de diverso tipo sin abandonar el contexto cognitivo.

El uso del conocimiento para una mejora de las estructuras organizativas y sociales ha dado lugar a un gran abanico de herramientas tecnológicas cuya finalidad es soportar estas estructuras y facilitar los flujos de conocimiento entre los agentes que las componen. Las organizaciones no solo deben disponer de medios tecnológicos para la generación, síntesis y transmisión del conocimiento, sino que deben existir otros sistemas que faciliten el flujo del conocimiento.

Como consecuencia las organizaciones que deciden implementar tecnologías relacionadas con la GC deben realizar cambios organizativos y, en muchos casos, cambios de cultura para conseguir que el uso de estas herramientas tecnológicas acompañados de otros sistemas no tecnológicos lleve a una

mejora de los procesos de la organización (Grau, 2002)

La GC mejora con la aplicación de tecnologías, donde la ventaja competitiva no depende solamente del dominio de las tecnologías esenciales sino de su uso correcto. De esta manera se intenta mantener y mejorar la posición competitiva de la empresa mediante la utilización de las tecnologías (Dankbaar, 1993).

Tiwana (2000) propone las tecnologías más apropiadas que han dado soporte a las distintas fases del Modelo Espiral del Conocimiento de Nonaka y Takeuchi (1995), Socialización, Externalización, Combinación e Internalización. En los avances más recientes las tendencias futuras de las tecnologías, describen las funciones que pueden cumplir algunos productos tecnológicos, desde la perspectiva de la GC, las que tienen que ponerse al servicio de las organizaciones (Joyanes, 2003). De ésta forma se pueden considerar como tendencias actuales a:

- Datamining, la utilización de algoritmos para analizar la agregación de datos y metadatos, para revelar relaciones, modelos y que a primera vista no fueron captados por el usuario
- Data Warehouse: como deposito separado, centralizado e integrado de los datos necesarios en una organización
- Gestión de contenidos: que consisten en identificación y definición de tipos de contenidos, su mantenimiento, publicación, distribución y protección
- Gestión documental: que incluye el almacenamiento, categorización, búsqueda y recuperación.
- Workflow: gestión del flujo de trabajo o procesos productivos por medio de una comunidad virtual on-line

De todas estas tendencias tecnológicas en este proyecto se pone especial atención en el Data Mining o minería de datos, considerando que la misma permite abordar problemas relacionados con la inducción de conocimientos a partir de datos almacenados.

La Minería de Datos es un término, que apareció hace más de 10 años y engloba resultados de investigación, técnicas y herramientas usadas para extraer información útil de grandes bases de datos. La Minería de datos es el descubrimiento semi-automático de patrones, asociaciones, cambios, anomalías y estructuras estadísticamente significantes y eventos en los datos, (Grossman, 1998).

La gestión de proyectos, entendida como una herramienta importante en la gestión moderna, la cual puede definirse como la asignación, vigilancia y utilización de recursos para alcanzar un objetivo particular dentro de un periodo concreto (Harvard Business Essentials, 2004).

El proyecto de investigación que se presenta, tiene como propósito proponer técnicas, metodologías, métodos y modelos que permitan que las organizaciones de nuestro medio (públicas o privadas) puedan gestionar sus recursos intangibles a través de estrategias innovadoras de cambio. En particular se pone interés en herramientas de TIC que puedan ser implementadas en organizaciones de nuestra región, en nuestra realidad, de manera tal que las mismas puedan ir adaptándose a modernos enfoques de cambio.

### **Líneas de investigación y desarrollo**

Este proyecto es una continuación del proyecto "Propuesta de cambio en organizaciones actuales desde la estrategia de los recursos intangibles" que se llevo a cabo en el periodo 2010-2012.

Las líneas de investigación que se abordarán son:

- *Data Mining*, es una técnica de descubrimiento de conocimientos en almacenes y bases de datos. Ante los volúmenes de información que se recogen diariamente de la organización, la minería de datos es el proceso de explotación de estos mediante el cual se extrae, desde los almacenes conocimientos provechosos.
- *Data Warehouse*: es un como deposito separado, centralizado e integrado de los

datos que son necesarios en una organización (Inmon, 1995).

- *Gestión de contenidos*: que consisten en identificación y definición de tipos de contenidos, su mantenimiento, publicación, distribución y protección.
- *Gestión documental*: que incluye el almacenamiento, categorización, búsqueda y recuperación.
- *Workflow*: gestión del flujo de trabajo o procesos productivos por medio de una comunidad virtual on-line.
- *Gestión de proyectos*: que abarca lo referente a buenas prácticas para la gestión de proyectos, brindando herramientas que facilitan o soportan las actividades inherentes.

### **Objetivos del Proyecto**

El proyecto procura apoyar a que las organizaciones de nuestro medio (públicas o privadas) puedan gestionar el conocimiento que las componen a través de estrategias y herramientas innovadoras de cambio.

Para alcanzar este fin se han planteado los siguientes objetivos:

- Comprender las distintas estrategias de GC para Identificar, recolectar y organizar el conocimiento existente en las organizaciones y así facilitar la creación de nuevo conocimiento.
- Contribuir a que las organizaciones puedan alcanzar ventajas competitivas a través de la GC.
- Promover el uso de herramientas de Gestión de Proyectos aplicándolas a la GC.
- Lograr un mayor entendimiento de la GC y de las TIC que den soporte a esta.
- Proponer modelos que permitan implementar estas estrategias de cambio en organizaciones de nuestro medio.
- Contribuir a la formación y capacitación de docentes/investigadores en las distintas estrategias de cambio y herramientas TIC.
- Transferir y ofrecer servicios al medio, a través de:

- Asesoramiento y capacitación a estudiantes de grado y posgrado por medio de cursos y seminarios.
- Realización de publicaciones en revistas especializadas y contribuciones.
- Asesoramiento y asistencia técnica a organizaciones del medio en los temas relacionados con la investigación.

## Metodología

A partir de los objetivos formulados y basados en un enfoque sistémico se detalla la metodología que se utilizara para la realización de este proyecto.

### 1. Universo de Análisis

**Descripción:** El presente proyecto se aplicará en organizaciones públicas o privadas de la Provincia de Catamarca. Las organizaciones del medio por sus características, objetivos y funciones, resultan ideales para llevar a cabo el proyecto de investigación, debido a que aún no son muy conocidas las estrategias de la GC.

### 2. Estudio y Análisis Exploratorio

**Objetivos:** Determinar el estado actual de las distintas áreas de conocimiento que abarca el proyecto de investigación.

**Descripción:** Esta etapa consiste en la búsqueda, recolección, lectura comprensiva y análisis de las fuentes de información referidas al tema que trata el proyecto, de manera de poder sustentar las bases teóricas y conceptuales en las cuales se apoya la investigación.

Para ello se llevaran a cabo las siguientes actividades:

- Búsqueda, selección, traducción y lectura comprensiva de la bibliografía y material de referencia sobre los temas en estudio.
- Análisis del material bibliográfico.
- Producción de informes sobre el estado del arte de los distintos temas en el ámbito del proyecto y publicaciones

En esta etapa se realizará el análisis y estudio de los siguientes temas:

- GC, modelos, metodologías
- Capital intelectual
- Gestión de la tecnología.
- Infraestructura tecnológica actual para la GC.
- Gestión de proyectos aplicada a la GC basada en las normas Project Management Body Of Knowledge (PMBOK)

### 3. Propuesta de Modelos, Técnicas y Tecnologías de Información

**Objetivo:** Diseñar estrategias, modelos, metodologías, técnicas y tecnologías de información adecuadas para la GC.

**Descripción:** Se seleccionará, modelos, metodologías y técnicas de GC existentes estudiados en la fase exploratoria. Asimismo, se tomará el análisis referido a la infraestructura tecnológica apropiada y la caracterización de la organización del entorno.

Se propondrá el diseño de estrategias, modelo y metodologías de GC que podría ser aplicado a la realidad de las organizaciones del medio.

### 4. Evaluación y Rediseño de la Propuesta Estratégica

**Objetivo:** Evaluar la propuesta estratégica obtenida en la fase de diseño.

**Descripción:** Se seleccionará una organización representativa de las organizaciones del medio (estudio de caso) o un conjunto de muestras seleccionadas en forma aleatoria. Se implementará en ésta/s las propuestas diseñadas. Se registrarán y estudiarán los resultados. En virtud de ello, se realizarán las modificaciones (si fuera necesario) en el diseño de la/s propuesta/s.

## Formación de Recursos Humanos

El equipo de trabajo del proyecto se encuentra conformado por: un director, un co-director, y cinco docentes investigadores. En este

proyecto también participan estudiantes de la Carrera Ingeniería en Informática de la Facultad de Tecnología y Ciencias Aplicadas. Se tiene previsto incorporar más alumnos de grado en los próximos años, a efectos de brindarles un marco apropiado para sus trabajos finales.

Se pretende que los integrantes de este proyecto desarrollen investigaciones bajo un paradigma científico común atendiendo la problemática que serán abordados en este proyecto de investigación. También se propone que trabajen en forma colaborativa, uniendo esfuerzos para promover un mayor desarrollo científico tecnológico de la región.

## Referencias

DANKBAAR B. (1993). Overall strategic review, Projecte SAST núm. 8 (Research and Technology Management in Enterprises: Issues for Community Policy), EUR-15426, Bruselas/Luxemburgo, Comisión Europea

DAVENPORT, Thomas H. and Prusak Laurence (1998), *Working Knowledge*. [en línea] Harvard Business School Press. Disponible en: <http://www.bus.utexas.edu/kman/glossary.htm>

GRAU, América (2002). *Herramientas de Gestión del Conocimiento* [en línea]. Fundación Iberoamericana del conocimiento. Disponible en: <http://www.gestiondelconocimiento.com/articulos.php>.

GROSSMAN, R., S. Kasif; R. Moore, D. Rocke; J. Ullman. (1998) *Data mining research: opportunities and challenges*. A report of three NSF workshops on mining large, massive, and distributed data, September 18, 1998.

Harvard Business Essentials (2004). *Gestión de Proyectos, enfoques y conceptos para avanzar*. Ediciones Deusto

INMON W.H. (1995) ."What is a Data Warehouse?" Prism, Volume 1, Number 1, 1995.

JOYANES L. (2003). *Innovaciones Tecnológicas y Sociedad del Conocimiento: Reflexiones para la Nueva Alfabetización* Actas del Congreso Internacional La Nueva Alfabetización: un reto para la educación del siglo XXI, pp 191-229. Madrid, 6 al 8 de diciembre de 2003

NIEVES LAHABA, Yadira y LEON SANTOS, Magda. (2001). *La gestión del conocimiento: una nueva perspectiva en la gerencia de las organizaciones*. ACIMED [online]. 2001, vol.9, n.2, pp. 121-126. ISSN 1561-2880. Mayo-ago. 2001

NONAKA, Ikujiro; TAKEUCHI, Hirotaka. (1995). *The Knowledge-Creating Company: how Japanese Companies create the dynamics of innovation*. Oxford University Press.

Project Management Body Of Knowledge (PMBOK) – Cuarta edición – Project Management Institute.

TIWANA, Amrit (2000). *The Knowledge Management Toolkit*. Prentice Hall



# Integración de Computación Heterogénea con Hadoop para Cloud Computing

Nelson Rodríguez<sup>1</sup>, María Murazzo<sup>2</sup>, Daniela Villafañe<sup>3</sup>, Maximiliano Alves<sup>4</sup>, Diego Medel<sup>5</sup>

Departamento e Instituto de Informática - F.C.E.F. y N. - U.N.S.J.

Complejo Islas Malvinas. Cereceto y Meglioli. 5400. Rivadavia. San Juan

0264 – 4234129 Fax: 0264-4234980

<sup>1</sup>nelson@iinfo.unsj.edu.ar <sup>2</sup>marite@unsj-cuim.edu.ar <sup>3</sup>villafane.unsj@hotmail.com

<sup>4</sup>maximilianoalvespinheiro@gmail.com <sup>5</sup>mdiego88@gmail.com

## Resumen

La potencia de procesamiento de las supercomputadoras de ayer está ahora disponible en el desktop, aunque la necesidad de mayor poder computacional para resolver los problemas más grandes continúa creciendo. Los sistemas de cluster escalables hoy tienen la promesa de ejecución ilimitada a un costo muy conveniente.

El paralelismo constituye una alternativa real para reducir el tiempo de ejecución de las aplicaciones. Arquitecturas paralelas homogéneas con elementos de proceso de características similares están siendo utilizadas con éxito en distintos ámbitos de la ciencia y de la industria. De forma natural, la ampliación y la potenciación de un sistema homogéneo con nuevos elementos, deriva en un sistema de naturaleza heterogénea en el que los nuevos componentes presentan características diferenciales con los anteriores.

Este modelo de computación presenta nuevos desafíos fundamentalmente para integrarse con tecnologías diversas como la computación distribuida, Internet, HPC, data centers y bases de datos entre otras.

**Palabras clave:** *Computación heterogénea, Cluster, Hadoop, GPGPU, Cloud Computing*

## Contexto

El presente trabajo se encuadra dentro del área de I/D Innovación en Sistemas de Software, y se enmarca dentro del proyecto de investigación Implantación de un ambiente de Cloud Computing para integración de recursos, el cual tiene como unidades ejecutoras al Departamento e Instituto de Informática de la FCEFyN de la UNSJ.

Durante 2011 y 2012 se llevó a cabo el mencionado proyecto que continúa por un año más. Como consecuencia del mismo, se vislumbró la necesidad de contar con la arquitectura de soporte necesarias para montar las distintas capas de la arquitectura IaaS, PaaS y SaaS. En el caso de IaaS, aparece HPC como la alternativa más atractiva para trabajar en ámbitos académicos y de investigación, por todo lo que puede ofrecer, y por supuesto sobre ella montar las otras capas de la arquitectura. Para ello se analizaron diversas posibilidades, de las cuales la computación heterogénea y Hadoop han resultado las opciones seleccionadas.

## Introducción

Con el uso masivo de Internet, el surgimiento de tecnologías asociadas, accesos desde dispositivos variados y nuevos servicios en tiempo real, ha crecido notablemente la demanda de aplicaciones de computadoras. Cloud Computing ha resultado un multiproveedor de servicios, que comparte información, software y recursos abiertos dentro de un ambiente basado en Internet.

El fenómeno comúnmente conocido como Cloud Computing representar un cambio fundamental en el sentido en que los servicios de tecnología de la información (IT) son desarrollados, desplegados, actualizados, escalados, mantenidos, y pagados [1].

En los últimos años, con la popularización del concepto y aplicación de Cloud Computing, surgieron diversas tecnologías para darle soporte. En particular Hadoop, es un modelo de computación open-source para Cloud Computing, y es tenido en cuenta cada vez más por la academia y los círculos industriales [2].

La computación heterogénea se refiere a sistemas que utilizan una variedad de diferentes tipos de unidades de computación. Una unidad de cálculo puede ser un procesador de propósito general (GPP) y un procesador de propósito especial (como DSP, GPU o FPGA). En general, una plataforma de computación heterogénea consiste en procesadores con diferentes arquitecturas de conjuntos de instrucciones.

Las arquitecturas heterogéneas a nivel de nodo han resultado atractivas durante la última década por varias razones comparadas con la CPU tradicional, estas ofrecen alto niveles de performance y son eficientes en consumo de energía y costos [3]. Así pues, es muy frecuente encontrar arquitecturas paralelas donde las características de los elementos que componen el sistema (los procesadores, la memoria, la red,...) pueden ser diferentes.

A finales del siglo XX, en la década de los 90, con la consolidación de estándares para la programación en arquitecturas de paso de mensaje como PVM y MPI, la programación

de sistemas heterogéneos se convierte en un problema tan sencillo o complicado como pueda ser la programación de arquitecturas homogéneas. Sin embargo, aunque la portabilidad de los códigos está garantizada, no ocurre lo mismo con el rendimiento observado en las aplicaciones.

Aparecen nuevas situaciones y cuellos de botella que no pueden ser resueltos mediante la aplicación directa de los modelos y las técnicas conocidas para el caso homogéneo. Se hace necesario adaptar los métodos conocidos y en muchos casos diseñar nuevas estrategias comenzando desde cero. Aparece un conjunto importante de problemas abiertos que están siendo intensamente estudiados y analizados [4].

Dentro del ámbito de la investigación científica, siempre ha existido la necesidad de tener computadoras con grandes capacidades de procesamiento (del orden de los Teraflops) y grandes capacidades de almacenamiento (del orden de los Petabytes). El aporte que ofrecen estos tipos de computadoras es de crucial importancia debido a que permiten realizar números de cálculos imponentes que se desarrollan a nivel de investigación como son las simulaciones en producción industrial, computología examinando bugs en grandes programas, explorando propiedades ferromagnéticas en la física, analizar dinámica molecular a nivel químico, etc. El área informática que lleva a cabo este tipo de procesamiento se llama HPC (High Performance Computing).

La dificultad de la aplicación de esta área a los ambientes de investigación radica en el costo del equipamiento que se necesita para obtener esas capacidades de procesamiento y almacenamiento, antes mencionadas, sobre todo en la formación de recursos humanos en el ambiente académico. Para poder solucionar este problema, se han desarrollado algunas áreas dentro de HPC que ofrecen una solución con costos accesibles y de excelentes resultados tanto para almacenamiento como procesamiento, estos son la Computación Heterogénea y Hadoop.

El concepto de Computación Heterogénea se aplica, en ocasiones, a sistemas compuestos por diferentes tipos de PCs y

máquinas con múltiples procesadores conectados mediante redes. Debido a las diferencias entre las máquinas que forman el sistema, es probable que las velocidades de cómputo de los procesadores sean distintas y los tiempos de transferencia de datos también pueden ser diferentes en las comunicaciones entre cada par de procesadores. La naturaleza de esta red es inherentemente dinámica y depende de qué máquinas se utilicen en cada momento para resolver un problema y cuales sean sus características (capacidades de cómputo, memoria, comunicaciones, etc.). La programación dependiente de la arquitectura de la máquina supone una dificultad adicional en este tipo de sistemas. Otra de las desventajas que aparecen en los sistemas heterogéneos se debe al comportamiento de los tiempos de ejecución de las aplicaciones.

Algunos autores definen La informática híbrida, la cual representa en la actualidad la intersección de tres paradigmas ampliamente usados para infraestructura de computación: (1) HPC (tradicional) centrada en el propietario; (2) la computación Grid (compartición de recursos); (3) Cloud Computing (provisión de recursos y servicios bajo demanda) [5]. A esto, para la propuesta presentada, se le debe agregar la capacidad computacional de GPU.

Cada paradigma está caracterizado por un conjunto de atributos de la generación de recursos hasta la infraestructura y de las aplicaciones ejecutando en esa infraestructura.

Los objetivos de este trabajo se centran en el desarrollo de una plataforma de HPC en Cluster que integre varios recursos de forma de poder analizar, evaluar y estudiar diversos aspectos y medidas como performance e interoperabilidad entre otros.

## Líneas de investigación y desarrollo

El rendimiento conseguido con un multiprocesador o multicomputador paralelo suele ser más elevado que el obtenido en un

entorno de máquinas heterogéneas; los usuarios pueden verse obligados a aceptar en algunos casos una reducción del rendimiento de sus aplicaciones a favor de una gran reducción en el coste del sistema. Una de las razones para que esta situación se produzca, es que la mayoría de los programas paralelos han sido desarrollados bajo la hipótesis de trabajar sobre una arquitectura homogénea [6]. La computación Heterogénea se encuentra dentro del contexto de HPC y es una solución potente, debido a que la computadora que lidera el TOP 500 de las maquinas con mayor capacidad de procesamiento del mundo utiliza esta tecnología [6]. Lo que se pretende es combinar la escalabilidad de los clúster con MPI y la gran capacidad de procesamiento de las placas GPGPU[7] con CUDA[8] que son dos modelos prácticamente ortogonales pero que combinados pueden alcanzar niveles colosales de procesamiento. Este modelo sirve para procesar grandes cantidades de cómputo, los cuales en la mayoría de las ocasiones va acompañado de una gran cantidad de datos, para los cuales la computación heterogénea no brinda un soporte muy estable, sino que su mayor preocupación es la cantidad de instrucciones por segundo que se pueden ejecutar.

HADOOP [9] segunda solución que se planteó ante la problemática citada anteriormente, es un framework que da soporte a aplicaciones paralelas del tipo MAP-REDUCE [10]. La misma ofrece no solo un motor de procesamiento, sino un sistema de archivos distribuido llamado HDFS (Hadoop Distributed File System), que fue diseñado para brindar un soporte de almacenamiento fiable sobre un gran número de máquinas en un gran clúster. Este FS (File System) maneja tolerancia a fallos y balanceo de carga, lo que lo convierte en una poderosa herramienta para el manejo de grandes volúmenes de datos.

La adopción de procesamiento de las GPUs heterogéneas en algunos de los sitios más importantes del mundo HPC indica que este paradigma se está moviendo más allá de la fase experimental, y las GPU son cada vez más confiables. Tanto como el hardware de la

GPU y las tecnologías de software avancen, a medida que más estudiantes universitarios y otras personas puedan aprender cómo aprovechar las GPUs y tarjetas gráficas, IDC cree que las GPU tendrá un papel cada vez más importante en el mercado mundial de HPC, como complemento de los procesadores x86 dentro del ecosistema de alto rendimiento. En cuanto a la combinación de las tecnologías computación Heterogénea y HADOOP puede ser un aporte de gran aceptación en el ambiente científico y académico, ya que ofrece un soporte a problemáticas investigativas de áreas científicas.

La arquitectura resultante permitirá realizar análisis de escalabilidad y rendimiento. A pesar de que ambos conceptos se confunden son muy diferentes. Caso testigo como el ocurrido con la migración de servidores del sitio web Amazon.com minorista web, que pasó de 300 transacciones por segundo (TPS) a tan sólo 3 TPS cada uno después de mudarse a una arquitectura más escalable. La ventaja es que, si bien todos los servidores web pueden tener un menor rendimiento individual, el sistema en su conjunto se convirtió en mucho más escalable y nuevos servidores web se podría añadir el infinito.

En un futuro Aspectos de eficiencia, extensibilidad, disponibilidad, reusabilidad y personalización son sin duda otros problemas a resolver. Además por la naturaleza distribuida de los datos, resulta dificultoso mantener la consistencia y coherencia de datos, contar con una representación homogénea de los mismos (en una plataforma heterogénea), eficiencia en la sincronización entre las plataformas (distribuida y paralela).

La programación y la paralelización de algoritmos para procesamiento heterogéneo, es un problema muy difícil de resolver, con lo cual la verificación y validación de programas de este tipo, parece por ahora imposible de lograr.

## Resultados y Objetivos

### Resultados Obtenidos

Los trabajos publicados más relevantes de mencionan en [12 a 20], los cuales se enmarcan fundamentalmente en el área de Cloud Computing. También se llevaron a cabo trabajos de divulgación.

Además se ha aprobado cuatro (4) tesinas sobre soluciones a distintos problemas que presenta Cloud Computing.

### Resultados Esperados (Objetivos)

El objetivo del grupo de investigación es la construcción de una plataforma para Cloud Computing basada en Cluster que integre la computación de alta performance con servicios como Hadoop.

Se espera que una vez realizada dicha plataforma, se puedan realizar estudios sobre aspectos como la eficiencia, balance de carga e interoperabilidad, y proponer las soluciones necesarias para resolver las fallas que ocurran, tratando de favorecer el uso de los estándares abiertos.

## Formación de Recursos Humanos

El proyecto marco sobre el que se realizan las investigaciones ha tenido como objetivo la interoperabilidad en Cloud Computing, a partir del mismo se han realizado publicaciones y trabajos de divulgación donde participan alumnos, becarios e investigadores, además cuatro tesina de licenciatura, una de tecnicatura en programación y otra de programador universitario. Por otro lado hay 2 (dos) tesinas de tecnicatura en desarrollo y 3 (tres) tesinas de licenciatura, y se espera realizar alguna tesis de maestría y aumentar el número de publicaciones. Se encuentran desarrollando sus proyecto dos becarios de iniciación a la investigación, que se enfocan en las líneas de investigación presentadas. Por otro lado también se prevé la divulgación de varios temas investigados por medio de cursos de postgrado y actualización.



## Referencias

- [1] Marston, Li, Bandyopadhyay, Zhang, Ghalsasi. "Cloud computing — The business perspective". *Decision Support Systems* 51 (2011) 176-180. Elsevier. 2011.
- [2] Lu, Hai-shan, Ting-ting."Research on Hadoop Cloud Computing Model and its Applications".2012. Third International Conference on Networking and Distributed Computing
- [3] Brodtkorba,, Dykena, Hagen, Hjelmervika, Storaasl.. "State-of-the-art in heterogeneous computing". *Scientific Programming* 18 (2010) 1–33.
- [4][6] Moreno de Antonio. "Computación paralela y entornos heterogéneos". *Soportes Audiovisuales e informáticos. Serie Tesis Doctorales. Servicio de Publicaciones. Universidad de la laguna. Curso 2004/05. Ciencias y tecnologías/23. ISBN.: 84-7756-662-3.*
- [5] Mateescua, Gentzsch, Ribbens. "Hybrid Computing—Where HPC meets grid and Cloud Computing". *Future Generation Computer Systems* 27 (2011) 440–453
- [7] Olexandr Isayev. "Computación Heterogénea: Nuevo Paradigma Para La Era Exaescala". *Paralelizados.com. Comunidad de usuarios de HPC. GPUScience. IDC-Exascale-Executive-Brief\_Nov2011. Noviembre 23, 2011 .*
- [8] [9] Programación paralela facilitada. NVIDIA Corporation.[http://la.nvidia.com/object/cuda\\_home\\_new\\_la.html](http://la.nvidia.com/object/cuda_home_new_la.html). 2013.
- [10] Hadoop. Welcome to Apache Hadoop. <http://hadoop.apache.org/>
- [11] J. D.Sanjay Ghemawat. "Google. MapReduce: Simplified Data Processing on Large Clusters".OSDI'04: Sixth Symposium on Operating System Design and Implementation,San Francisco, CA, December, 2004.<http://research.google.com/archive/mapreduce.html>
- [12] Murazzo, Rodríguez. "Mobile Cloud Computing". WICC 2010. Calafate. Mayo 2010.
- [13] Murazzo, Millán, Rodríguez, Segura, Villafañe. Desarrollo de aplicaciones para Cloud Computing. CACIC 2010. Morón. Oct. 2010.
- [14] Murazzo, Rodríguez, Millán, Segura y Villafañe."Plataformas Educativas Implementadas Con Cloud Computing". XVI CACIC 2010, Workshop de Tecnologías Informáticas Aplicadas a la Educación. Morón. Oct. 2010.
- [15] Murazzo, Rodríguez. "Una propuesta para el desarrollo de aplicaciones para Mobile Cloud Computing". Congreso Internacional de Computación y Telecomunicaciones – COMTEL 2010, Lima, Perú. Oct. 2010.
- [16] Rodríguez, Murazzo, Ene. "Cloud Computing". WICC 2009. San Juan. Nov 2009.
- [17] Rodríguez, Chávez, Martín, Murazzo, Valenzuela. "Interoperabilidad en Cloud Computing". WICC 2011. Rosario 2011.
- [18] Chávez, Martín, Rodríguez, Murazzo, Valenzuela Metodología AGIL para el desarrollo SaaS. WICC 2012. Posadas. 2012.
- [19] Rodríguez, Valenzuela, Chávez, Martín, Murazzo, Villafañe. "Ambiente de desarrollo para lengua de señas basado en Cloud". WICC 2012. Posadas. 2012.
- [20] Rodríguez, Villafañe, Murazzo, Gallardo, Tarrachano. "GAE, una estrategia para complementar SaaS y PaaS a través de la Web". 2do SABTIC. Tres de Maio, Brasil. Agosto 2012.

# Integración de aplicaciones web a través de sindicación de contenidos

María Soledad Romero, Mariano García Mattío, Pablo Andrés Vaca

Maestría en Ingeniería en Sistemas de Información / Universidad Tecnológica Nacional / Facultad Regional Córdoba

Maestro M. López esq. Cruz Roja Argentina – Ciudad Universitaria – Córdoba – Argentina C.P. (X5016ZAA)

Tel. (54 0351 5986034)

{romeroma.soledad, magm3333, vacapablo72}@gmail.com

## Resumen

RSS (Really Simple Syndication) o sindicación de contenidos es una característica de muchos sitios Web. Su propósito es organizar y difundir la información disponible en las páginas Web. Usa el principio de suscripción, el usuario que desea recibir determinados contenidos efectúa una suscripción de los mismos. Una de las ventajas principales de la suscripción consiste en que el usuario Web no invierte tiempo en buscar novedades ya que le llegan aquellas a las que está suscripto.

Con el advenimiento de internet, las aplicaciones tendieron a adaptarse por lo que el mercado de software invirtió fuertemente en desarrollo de entornos de programación que permitieran acompañar esta tendencia. Por consiguiente se desarrollaron productos para generar aplicaciones web. Desde el mercado de los dispositivos móviles también se incorporó a las aplicaciones web como lo habían hecho los navegadores. Ha crecido la integración de servicios entre aplicaciones. El usuario final de una aplicación web puede ingresar a la misma y a partir de estar autenticado acceder a otros servicios.

El problema radica entonces en brindar al usuario final de una aplicación Web (accedida a través de un navegador o de un teléfono celular) que requiere autenticación, la posibilidad de enviarle información personalizada y privada a través de RSS a fin de combinar las ventajas de RSS con las funcionalidades de la aplicación web, incluso aquellas que dan soporte a

workflow, sujeto a selección de categorías y/o combinación de categorías de información.

**Palabras clave:** RSS, autenticación, aplicaciones web, feed, middleware, publicación, REST, Rome Project, suscripción, token, workflow.

## Contexto

El proyecto forma parte de la propuesta de Tesis para la Maestría en Ingeniería de Sistemas de Información y es avalado por la empresa One Click e Business S.R.L. para el desarrollo de un prototipo que permita integrar aplicaciones web.

## Introducción

El paradigma de la comunicación está cambiando y podría decirse que “el usuario ya no va en busca de la información, sino que es la información la que va en busca de él” [1].

Actualmente los formatos de intercambio de información entre computadoras han evolucionado y tienden a continuar en esa dirección facilitando la comunicación y la integración. Contamos con formatos como JSON [2] y XML, open software y librerías que permiten agregar valor específico a los proyectos como Rome [3], arquitecturas que permiten desarrollos independientes a los ya funcionando como los servicios web o web services [4] y estándares de transformación

como XLST [5] que facilitan la ornamentación del contenido que llega al usuario final.

Las distintas versiones de RSS han ido surgiendo para adaptarse a los cambios en la tecnología. Sólo son estándares las versiones RSS 1.0 y RSS 2.0 y Atom. Cabe destacar que Atom y RSS se diferencian por el formato de fecha utilizado para transferir información. El vocabulario empleado en RSS en general no es reusable en otros vocabularios de XML, mientras que el vocabulario de Atom fue diseñado para permitir elementos de vocabulario que sean reutilizados fuera del contexto del documento que se envía al syndicar contenidos con Atom. Si bien RSS y Atom tienen distinta sintaxis por ser desarrollos de proveedores independientes, conservan un núcleo de elementos de código que les son comunes.

Acompañando esta evolución de RSS han surgido trabajos de investigación. En "Developing Categorized News Items Using Feeds and Web Services" [6], Saha, Sajjanhar, Gao, Dew y Zhao describen cómo en la última década el masivo crecimiento de Internet ha provocado cambios masivos en la vida cotidiana de muchos hombres. Proponen el uso de sindicación de contenidos basado en servicios web o web services generando un prototipo basado en el estudio de algoritmos a fin de reducir los tiempos de distribución de contenido a partir de varios sitios Web.

En "Survey on application tools of Really Simple Syndication (RSS): A Case Study at Klang Valley" [7], Lee, Ghani y Huang analizan programas agregadores de RSS de tipo escritorio y Web, concluyen con los diez más usados de cada tipo. Dada la popularidad de los últimos tiempos exploran sobre las ventajas y desventajas. También ofrecen una perspectiva sobre la necesidad de trabajar con menos versiones de RSS.

En "A Novel Approach: Secure Information Notifying System using RSS Technology" [8], Preechaveerakul y Wichuta Kaewnopparat exploran sobre las posibilidades de syndicar contenido diferenciándolo entre contenidos

públicos y contenidos privados. Desarrollaron un modelo con la premisa de agregar y distribuir información pública y privada utilizando mecanismos de encriptación de datos para dar soporte a la seguridad de la información. Consideran casos de usuarios móviles y autenticación de usuario por lo que el modelo plantea una base de datos para la administración de los perfiles de usuarios. Además hacen referencia al uso de desarrollos producidos por el proyecto.

El proyecto Rome [3] se basa en un modelo abstracto de "suscripción o distribución de fuentes de noticias". Permite analizar cualquier formato de suscripción de noticias, incluyendo las variantes de RSS y Atom en este modelo. Puede convertir la representación de cualquier modelo de formato de suscripción de noticia a otro formato de suscripción de noticia. A nivel interno, define los modelos objeto intermedios de los formatos específicos de suscripción de noticias, incluyendo todas las variantes Atom y RSS. Para cada formato, hay una clase separada JDOM basada en XML que analiza un modelo intermedio. Ofrece convertidores para convertir entre los modelos de noticias (feed) intermedio y el modelo de fuente de distribución abstracto. Incluye además el manejo de los objetos involucrados en la autenticación de usuario.

Se entiende por autenticación a cualquier proceso por el cual se verifica que alguien es quien dice ser. La autorización es cualquier proceso por el cual a una persona se le permite estar donde quiere ir, o tener información que desea tener "Authentication and Authorization" [9]. La autorización requiere un proceso de autenticación previo.

Es importante destacar que existen distintos tipos de autenticación: en "Authentication Types" [10], Robert Moskowitz, describe los diferentes flujos. La autenticación existe para establecer la confianza entre dos partes, o entre las entidades de autenticación. Estas entidades consisten en una identidad y una llave. La autenticación se establece mediante la realización de una operación criptográfica entre las partes, identidades y claves. La operación de

cifrado o algoritmo de autenticación, establece la naturaleza de la confianza entre las partes. Una red de transporte o el flujo de autenticación proporcionan la conexión entre las partes para el algoritmo de autenticación. Algunos flujos de autenticación soportan un mecanismo de control de autenticación normalizado, para simplificar el soporte para múltiples algoritmos.

En la mayoría de los sistemas de autenticación las capas están estrechamente unidas y no se ha previsto ningún canal. En realidad, la aparición del canal de autenticación hizo posible un amplio espectro de soluciones de autenticación. En el modelo de capas, no todas las combinaciones son viables. El modelo de capas sirve de marco para el seguimiento del trabajo de clasificar varias combinaciones de flujos, canales, algoritmos, entidades y el modelo de confianza que proporcionan.

En "Authentication Types" [10], Robert Moskowitz señala que existen tres grupos de flujos de autenticación: unidireccional, bidireccional, y unidireccional acoplado. En el flujo de autenticación unidireccional una parte siempre debe iniciar la autenticación. En el flujo de autenticación bidireccional aunque una de las partes siempre inicia la autenticación, no hay un rol predeterminado de iniciador y el respondedor, y los roles se pueden invertir en cualquier momento. Hay una distinción clara entre la autenticación unidireccional y bidireccional. En la unidireccional, hay máquinas de estado exclusivas para el iniciador y el respondedor. En bidireccional, hay una máquina de estado que puede ser el iniciador de autenticación o respondedor. Esta máquina de estado sólo se encarga del cambio de sentido y de las condiciones de carrera (cuando ambas partes inician o ambos tratan de obligar a la otra a ser el respondedor). Unidireccional acoplado es una situación en la que hay dos flujos unidireccionales que se tienen que completar antes de que cualquiera de los flujos se considere terminado. El flujo de acoplamiento crea un ambiente muy similar a la autenticación bidireccional y se ocupa de la condición de carrera haciendo caso omiso de ella. No maneja cambios de dirección, esperando que cada parte

asuma plenamente tanto en el rol de iniciador como de respondedor. Ejemplos de protocolos de autenticación unidireccional incluyen a IEEE 802.1x, y TLS. Ejemplos de autenticación bidireccional incluyen IKE y HIP. IEEE 802.1aa se utiliza en flujos acoplados unidireccionales para la autenticación IEEE 802.11i AdHoc.

Observando las aplicaciones web, las autenticaciones y autorizaciones ya no son otorgadas a una sola aplicación. En "Authentication and Authorization in the Google Data Protocol" [11] se introduce como aplicaciones de terceros requieren a menudo un acceso limitado a la cuenta de Google del usuario para determinados tipos de actividad. Para asegurarse de que los datos de usuario están protegidos, todas las solicitudes de acceso deben ser aprobadas por el titular de la cuenta. El control de acceso tiene dos componentes, la autenticación y la autorización.

Los servicios de autenticación permiten a los usuarios iniciar sesión en su aplicación utilizando una cuenta de Google. Algunos servicios también permiten a los usuarios iniciar sesión con otra cuenta, como un inicio de sesión OpenID.

Los servicios de autorización permiten a los usuarios proporcionar acceso a los datos de sus aplicaciones que tiene almacenados en las aplicaciones de Google. Google se toma en serio la privacidad, y cualquier aplicación que requiera acceso a los datos de un usuario debe estar autorizado por el usuario [11].

Los servicios de autenticación y autorización a menudo se conocen colectivamente como auth.

Del análisis de las herramientas disponibles y el estado del arte [3] [12] surge la inquietud de crear un prototipo para publicación de contenido que tenga valor agregado a la sindicación de contenidos:

-Un token de autorización permite al usuario suscriptor acceder a la información a la que tiene acceso en otra aplicación web donde los datos residen.



-Cada usuario puede establecer qué tipo de información desea recibir, esto es, una o más categorías predefinidas (reclamos, tareas, oportunidades, informes).

-Para una categoría o tipo de información predefinida el usuario puede establecer condiciones en base a las cuales desea recibir la novedad o feed.

- Cada usuario puede asignar puntos (del 1 al 10 en cada categoría) lo que determina la importancia de la misma.

- En función de los parámetros mencionados (categorías, peso y condición) el usuario suscriptor podrá recibir un feed o novedad ajustado a sus preferencias y necesidades de información.

- La presentación del feed o novedad puede customizarse con la incorporación de diversidad de estilo aplicado al html del documento o feed que llega al usuario final.

La figura 1 resume el objetivo del uso del token y la generación de contenido.

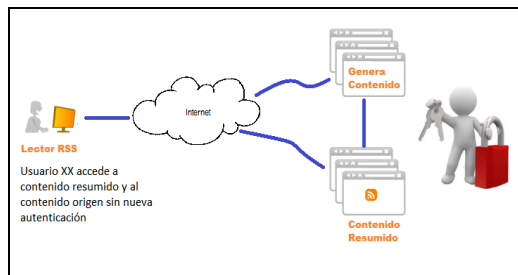


Figura 1 – RSS y Token

La arquitectura del middleware propuesto y el flujo que este middleware facilitará, está basado en la utilización de un token, el cual se constituye en la piedra angular del protocolo de aplicación. En la primera versión el token soporta información variable acerca del negocio, además de la información inherente al protocolo. El middleware permite establecer el algoritmo de cifrado del token mediante configuración.

Es deseable que en el futuro se pueda realizar el cambio de algoritmo de cifrado mediante un servicio estándar. También se requerirá trabajar en la ampliación de la arquitectura del protocolo de aplicación a fin de agregar características de expiración, validación de dominios, etc.

## Líneas de investigación y desarrollo

Se investigan los trabajos realizados. Se trabaja en el diseño y desarrollo de un prototipo denominado pulse (en referencia a pulsar, pedir) utilizando Java como lenguaje de programación, web services, algoritmos de encriptación de claves y Apache Tomcat como servidor de aplicaciones. El prototipo cumple la función de publicar contenido en canales de suscripción asociados a RSS administrando la autenticación y autorización.

## Resultados y objetivo

El objetivo principal es desarrollar un middleware [13] web que permita organizar y distribuir contenido sindicado en entornos Web, escritorio y móvil con la particularidad de incluir soporte para la autenticación relativa a la aplicación Web que contiene los datos que se desea distribuir. El middleware presta servicios mediante el estándar REST [2] para web services, se definirán interfaces claras de interacción con el middleware a efectos de poder consumir sus servicios desde cualquier lenguaje de programación.

Como objetivos secundarios cabe mencionar:

1. Explorar el estado actual del mercado en cuanto al uso de sindicación de contenidos se refiere.
2. Identificar alternativas de integración de aplicaciones web con sindicación de contenidos, en particular considerando el ingreso (usuario y contraseña) automático.
3. Desarrollar un proyecto de publicación de contenidos sindicados que sea independiente en cuanto a plataforma.
4. Proponer un esquema que contemple las distintas especificaciones de sindicación previamente identificadas.
5. Realizar pruebas de validación del esquema propuesto (distintos lectores).
6. Enriquecer el feed o contenido sindicado aportando alertas al usuario final.

7. Establecer una arquitectura y un flujo estándar para el uso de la sindicación integrada con aplicaciones web que requieren el delivery o entrega de contenido personalizado y basado en autenticación.

El resultado obtenido hasta el momento es la publicación de contenido sindicado desde una aplicación web a través del uso de los servicios del prototipo. Cabe destacar que se ha utilizado software open source para el desarrollo del 100% del trabajo.

## Formación de Recursos Humanos

El presente trabajo forma parte de la investigación incluida en la Tesis de Maestría en Ingeniería en Sistemas de Información cuyo plan de tesis fue aprobado en la Comisión de Posgrado de la Universidad Tecnológica Nacional, Facultad Regional Córdoba. La carrera de Maestría en Ingeniería en Sistemas de Información que se dicta en la UTN – FRC está acreditada por Resolución N°: 1277/12 de la CONEAU. Se pretende, más allá del trabajo de tesis propiamente dicho, realizar como producto del trabajo de investigación, seminarios de capacitación y transferencia de conocimientos en la universidad. Esto va a permitir a alumnos y egresados adquirir nuevos conocimientos, nuevas habilidades. Por otro lado, el trabajo de tesis será el puntapié de otros proyectos de investigación, lo que contribuirá a formar nuevos profesionales en este campo.

## Referencias

- [1] “¿Qué es RSS?”  
<http://www.uma.es/secretariageneral/Feed/queesRSS.php>  
 Universidad de Málaga – Secretaría General  
 Fecha de último acceso: 10/11/2012
- [2] “RESTful Web Services”  
 Autores: Leonard Richardson, Sam Ruby  
 Editor: O’Reilly Media  
 Edición: 2007
- [3] “The Rome Project”  
<http://rometools.org/>  
 Autores: Mark Woodman, Martin Kurz  
 Fecha de último acceso: 23/11/2012
- [4] “Java Web Services: Up and Running”  
 Autor: Martin Kalin  
 Editor: O’Reilly & Associates Inc.  
 ISBN: 978-0-596-52112-7
- [5] “XLS Transformation” – W3C Recommendation  
<http://www.w3.org/TR/xslt>  
 Editor: James Clark  
 Fecha de último acceso: 23/11/2012
- [6] “Developing Categorized News Items Using Feeds and Web Services”  
 Autores: Subrata Saha, Atul Sajjanhar, Shang Gao, Robert Dew -  
 School of Information Technology Deakin University -  
 Burwood, VIC 3125, Australia  
 Ying Zhao – School of Information Science and Technology –  
 Beijing University of Chemical Technology – Beijing, 100029,  
 P. R. China  
 2010 10th IEEE International Conference on Computer and  
 Information Technology (CIT 2010)
- [7] “Survey on application tools of Really Simple Syndication (RSS): A Case Study at Klang Valley”  
 Autores: TehPhoey Lee, Abdul Azim Abdul Ghani, Chan Yu Huang -  
 Faculty of Management and Information Technology,  
 UCSI, Malaysia.  
[http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=4631980&tag=1](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4631980&tag=1)  
 Fecha de último acceso: 20/08/2011
- [8] “A Novel Approach: Secure Information Notifying System using RSS Technology”  
 Ladda Preechaveerakul and Wichuta Kaewnopparat  
 Computer Science Department  
 Faculty of Science, Prince of Songkla University  
 Hat Yai, Songkhla, Thailand  
[http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=5189906](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5189906)  
 Fecha de último acceso: 20/08/2011
- [9] “Authentication and Authorization”  
<http://httpd.apache.org/docs/2.2/howto/auth.html>  
 Fecha de último acceso: 22/02/2013
- [10] “Authentication Types”  
 Robert Moskowitz, ICSAlabs  
 Autor: July 16, 2003 - Draft version 1.2  
<http://www.ieee802.org/1/files/.../Authentication%20Types%20%201-2.doc>  
 Fecha de último acceso: 22/02/2013
- [11] “Authentication and Authorization in the Google Data Protocol”  
<https://developers.google.com/gdata/docs/auth/overview#OAuth2>  
 Fecha de último acceso: 22/02/2013
- [12] “Developing feeds with RSS and Atom”  
 Autor: By Ben Hammersley  
 Editor: O’Reilly & Associates Inc.  
 ISBN: 9780596008819
- [13] “A Survey of Middleware”, 18th International Conference on Computers and Their Applications, March 26-28, 2003, Honolulu, Hawaii.  
<http://triton.towson.edu/~karne/research/middlew/surveym.pdf>  
 Autores: Toni A. Bishop y Ramesh K. Karne  
 Fecha de último acceso: 20/07/2012

## Hacia el Fortalecimiento de la Sociedad en el Uso y Aplicación de la Información Geospacial y las TIC

Mabel Álvarez<sup>1</sup>, Blanca Agudiak<sup>1</sup>, Luis Reynoso<sup>2</sup>, Walter Lienqueo<sup>1</sup>, Jones Gwyn<sup>1</sup>, Ester Obreque<sup>1</sup>, Adriana Erblich<sup>1</sup>, Boris Diaz<sup>3</sup>

Grupo de Investigación en Gobierno Electrónico

<sup>1</sup> Departamento de Geografía – Facultad de Humanidades y Ciencias Sociales  
Universidad Nacional de la Patagonia San Juan Bosco (UNPSJB)  
Belgrano y Rawson (9100) Trelew, Chubut

<sup>2</sup> Departamento de Programación – Facultad de Informática  
Universidad Nacional del Comahue (UNComa)  
Buenos Aires 1400, (8300) Neuquén

<sup>3</sup> Departamento de Ciencias Exactas y Naturales – Unidad Académica Río Gallegos  
Universidad Nacional de la Patagonia Austral (UNPA)...  
Lisandro de la Torre 1070 (9400) Río Gallegos, Santa Cruz  
Contactos: mablop@speedy.com.ar

### Resumen

Tanto la concepción tradicional de “Sociedad Civil” como la nueva concepción de la “Sociedad del Conocimiento” tienen como eje central de desarrollo el sentido de la comunicación y el sentido de territorio. Si bien las tecnologías asociadas a cada concepción distinguen características culturales diferentes y propias en las cuales ambos conceptos fueron acuñados respectivamente, hay una comunión en ambos constructos que se basan en la comunicación y el territorio.

El presente proyecto trata de acercar a toda la sociedad al uso, aplicación y conocimiento de la información a través del empleo de tecnologías y tecnologías geoespaciales. Considerando sectores gubernamentales, públicos y privados trata de integrar temáticas relevantes como: tecnologías de acceso, información territorial, ontologías, infraestructura de datos espaciales. Nuestra motivación surge del hecho que el propio territorio ha estado vinculado históricamente a la pertenencia cultural, pero hoy en día esa pertenencia se ha visto modificada con la globalización y las nuevas formas de

vincularnos y comunicarnos. Hoy en día se caracteriza por las facilidades a las cuales tiene acceso el ser humano (o una institución), a qué redes y sobre qué acciones territoriales puede incidir. En el caso del gobierno, está en función también del desarrollo sustentable que fomenta a través de gestión geoespacial.

Palabras Clave: Tecnología de la Información y la Comunicación, Información Geoespacial, Infraestructura de Datos Espaciales.

### Contexto

La coordinación del proyecto “Hacia el Fortalecimiento de la Sociedad en el Uso y Aplicación de la Geoinformación y las TIC” (Proyecto 048/12, Res UNPSJB 493/2012) es llevada a cabo por una Unidad Ejecutora compuesta por integrantes de tres universidades de la Patagonia. Los miembros del proyecto propuesto tienen experiencia en el trabajo conjunto e integran el Grupo de Investigación “TIC e Información Geoespacial (IG)”, con existencia ad hoc desde hace varios años y aprobado formalmente por la UNPSJB, conforme a la normativa vigente OCS N° 131/09.

En materia de Información Geoespacial, se continúa la línea de investigación desarrollada en los siguientes Proyectos de Investigación:

- PI 405 “Aplicación de Geoinformación para optimizar la Administración de la Información Sanitario Ambiental del Ministerio de Salud de la Provincia del Chubut”, UNPSJB.
- PI 360 Aplicación de Geoinformación para optimizar la Administración de la Información Sanitario Ambiental del Ministerio de Salud de la Provincia del Chubut, financiado por el Programa de Vigilancia de la Salud y Control Epidemiológico del Ministerio de Salud de la Nación (VIGI+A), programa PNUD ARG 98/03 en el marco del convenio de préstamo BIRF
- PI 458 “Aplicación de Geoinformación para asistir a organizaciones encargadas de la Administración de Emergencias (ocasionadas por factores climáticos) del Gobierno de la Provincia del Chubut”, UNPSJB.

En la temática de TIC se continúa la línea de investigación de los proyectos de la UNPSJB:

- PI 605 “Plataforma e\_Learning para Geoinformación”.
- PI 803 “Potenciando la utilización de TIC en docencia, investigación y comunicación”.

## Introducción

La importancia del tema para las ciencias y especialidades en las que se encuadra el Proyecto, se basa en el papel que juegan actualmente y las posibilidades que brindan la Información Geoespacial (IG) y las Tecnologías de la Información y la Comunicación (TIC), para el logro de mejoras en la calidad de vida de las personas y la sociedad como un todo.

Tanto la Información Geoespacial como las TIC, requieren de un umbral mínimo

de alfabetización digital para que las personas puedan utilizarlas y sentir que les aportan en la solución de problemas de su vida cotidiana. Esta no es una tarea sencilla puesto que en la composición actual de la sociedad argentina, siguiendo la clasificación propuesta por Piscitelli [5], prácticamente la mitad de población del país, conforme a su edad, serían inmigrantes digitales y la otra mitad serían nativos digitales.

La prospectiva del país para el año 2020, expuesta en el año 2009, explicita las tendencias de las TIC en el país, y las Resoluciones de Cumbres Iberoamericanas relativas a la educación para la inclusión, el gobierno electrónico, la educación para la sostenibilidad, entre otros, orientan a potenciar bajo múltiples estrategias y acciones el uso de las TIC por parte de la población y el desarrollo de servicios apoyados en tecnologías (TIC, de IG, etc.) cada vez más orientados al ciudadano. En tal sentido, tomando en cuenta las especialidades de los miembros de la UE y los proyectos de investigación en los que han ido participando, se cuenta con una rica experiencia aplicada a la región. Las tendencias en materia de IG en general y de Infraestructura de Datos Espaciales (IDE), sumadas a las de TIC, ameritan impulsar un proyecto de las características del propuesto, donde se orientan los esfuerzos, a propiciar acciones para el acercamiento tanto de la IG como de las TIC a la sociedad y hacia una sociedad de trabajo en red [4], [6]. Todo ello, considerando a la vez los aspectos que se relacionan a los ámbitos, académicos, científicos, no gubernamentales y vínculos con áreas del sector público con funciones en IG y TIC. Desde este Proyecto se pretende por una parte contribuir al fortalecimiento de la cultura digital en materia de TIC, potenciando su empleo en actividades de



docencia, investigación y comunicación interna y externa a la comunidad universitaria y por otra aportar a la cultura digital en materia de Información Geoespacial, interna y externa a la comunidad universitaria, convencidos de los beneficios que trae su utilización en la vida cotidiana de las personas.

## Líneas de Investigación y Desarrollo

Para abordar la problemática planteada, se distinguen en el proyecto dos líneas de investigación que la componen, si bien el tratamiento de investigación también es conjunto:

**TIC:** La línea impulsa el estudio e investigación en el empleo de TIC por parte de la sociedad fundamentalmente a través de: herramientas Web 2.0 [1], el uso de espacios virtuales de enseñanza y aprendizaje, experiencias inter-institucionales, utilización de herramientas de software libre, identificación y compilación de contenidos de interés, creación de procesos que agilicen la gestión y/o la comunicación.

**Información Geoespacial:** La línea tiene como objetivo el tratamiento de información geoespacial y difusión de experiencias en el uso y aplicación de Infraestructuras de Datos Espaciales (IDE) [2], Ordenamiento Territorial (OT) [3] y Tecnologías de la Información Geográfica (TIG), en general.

## Resultados y Objetivos

Los miembros del proyecto reúnen experiencia previa como puede inferirse del contexto del proyecto presentado en este artículo. Luego de un año de ejecución del proyecto, los resultados, en relación a los objetivos previstos, más significativos incluyen:

- Producción científica en congresos nacionales e internacionales.
- Difusión de actividades en workshops y newsletters.
- Ejecución de Tareas en la Sociedad Geoespacial Internacional (dependiente de GSDI- Global Spatial Data Infrastructure) de la cual el director del presente proyecto es presidente.
- Participación y tareas conjuntas en la construcción inter-institucional de la IDE de la República Argentina.
- Especificación del proyecto de la UE (Unión Europea) denominado OpenCropMonitoring en conjunto con otros 20 miembros participantes del proyecto, en su mayoría de la comunidad europea. El proyecto tendrá una duración de tres años y el mismo se encuentra en proceso de evaluación.

## Formación de Recursos Humanos

El grupo reúne a 12 investigadores, entre los que se cuentan docentes y alumnos de UNPSJB y un asesor externo. Se cuenta actualmente con 2 doctores y 2 magisters. La formación de recursos humanos constituye una de las razones claves para la formulación de este proyecto, se desarrollará durante toda la etapa de ejecución y se plantea con los siguientes alcances. La Unidad Ejecutora, está integrada por investigadores formados y en formación. La presencia de alumnos nuevos y avanzados se orienta a formar personas en los primeros pasos en investigación o a proseguir los ya iniciados.

Parte de los investigadores están realizando estudios de grado o posgrado,

razón por la que a través del Proyecto se pretende generar el espacio y el ámbito para contribuir a las respectivas formaciones.

Por una parte el Proyecto se orienta al ámbito universitario local, por ser el entorno en el cual se desarrolla la formación presencial, tratando de generar estudios de caso, disponer de datos geoespaciales, de materiales y recursos que puedan utilizarse en cátedras de las carreras de grado y por la otra se pretende ampliar el horizonte local, fortaleciendo actividades del entorno regional patagónico, nacional e internacional, apoyados en el aporte de integrantes de la Unidad Ejecutora con residencia en otros lugares y en la participación en iniciativas nacionales e internacionales.

La formación de recursos humanos, se plantea con alcance a la educación formal en ámbitos de educación reglada y mediante educación no formal e informal, con miras a la sociedad en su conjunto, mediante el empleo de distintas estrategias y el desarrollo de sinergias con otros actores, para lo cual, los proyectos de investigación, las comunidades, las redes de contacto y de trabajo de los miembros de la UE será de vital ayuda.

## Referencias

- [1] Álvarez M., Gallego Gil, D, Dans, Marta I., Rosanigo, Z. B. , Cela, K, Valdivia Guzmán, J., Gatica Zapata, N, López Alvarez, L.A., Training in Web 2.0 tools: a way to bring Spatial Data Infrastructures to people, GSDI 12 World Conference, Singapore 19 -22 de octubre de 2010
- [2] Álvarez, M., Gallego Gil, D., Zerpa, C. (2012). Las IDE y el Gobierno Electrónico: esbozando perspectivas futuras. En Bernabé - Poveda, M.A. y López - Vázquez, C.M. Fundamentos de las Infraestructuras de Datos Espaciales. UPM - Press; Colección Científica. Madrid (España).
- [3] Álvarez, M., Gonzalez, M. E., Zárate, N. C., Schweitzer, A., Díaz, B., Martínez, G.: Articulación de unidades espaciales para el desarrollo y ordenamiento territorial en la Patagonia Austral, en Mapping, Revista Internacional de Ciencias de la Tierra, N°113, octubre de 2006, pp 58-63, Madrid. ISSN: 1.131-9.100”.
- [4] Álvarez, M., Lara, L.: Social networks as a means of spreading SDI to society, GSDI 12 World Conference, Singapore 19 -22 de octubre de 2010.
- [5] Piscitelli, A.: Nativos digitales. Dieta cognitiva, inteligencia colectiva y arquitecturas de la participación. Buenos Aires, Editorial Santillana, 2008.
- [6] Reynoso, L., Alvarez, M. (2012) Estructuras de Redes Sociales y Globales en Procesos Mesosociales y Macrosociales. VII Congreso Internacional de Tecnologías para la Educación y el Conocimiento. Tecnologías Emergentes. Universidad de Educación a Distancia, España.
- [7] Reynoso L., Álvarez, M., Berti Umbert, A., Gatica H. “Exploración de Información del Catastro Multipropósito: Una experiencia en la determinación de la Valuación de la tierra de parcelas rurales extensivas”, 1er Congreso Internacional sobre Catastro Unificado Multipropósito: Investigación Formación y Empresa. Jaén. España, 16 al 18 de junio de 2010.

# Administración y Control de Proyectos

## Extendiendo el Concepto de Indicadores

Edgardo Bernardis, Germán Montejano, Luis Roqué, Hernán Bernardis  
Facultad de Ciencias Físico-Matemáticas y Naturales  
Universidad Nacional de San Luis  
Ejército de los Andes 950 – San Luis – Argentina  
{ebernardis, gmonte, hbernardis}@unsl.edu.ar, araroq@yahoo.com

### Resumen

Hoy en día se manejan extensas cantidades de información, siendo esto un problema complejo a la hora de la toma de decisiones por parte de personas o empresas, particularmente en la industria del software. Para dicho fin, existen diferentes metodologías y herramientas que ayudan a la hora de manejar estas cantidades de información y poder extraer los aspectos más importantes que permiten tomar decisiones respecto del plan que se lleva a cabo. Dichas metodologías y herramientas se basan en controlar los tiempos y recursos para la gestión del plan; siendo un factor restrictivo a la hora de la toma de decisiones.

Actualmente no existe ninguna herramienta totalmente integrada, abierta y lo suficientemente flexible que permita diseñar un plan y controlarlo a través de indicadores.

Por esta razón se propone desarrollar un modelo con su posterior prototipo que subsane estas restricciones.

**Palabras clave:** Ingeniería del Software, Administración de Proyectos, Control de Proyectos, Herramientas CASE.

### Contexto

La línea de investigación descrita en este artículo se desarrolla en el Laboratorio de Calidad e Ingeniería de Software (LaCyS) de la Universidad Nacional de San Luis; y se encuentra enmarcada dentro del proyecto: *Ingeniería del Software: Aspectos de Alta Sensibilidad en el Ejercicio de la Profesión del Ingeniero de Software*, perteneciente a la universidad antes mencionada. Dicho proyecto, es reconocido por el programa de incentivos, y es la continuación de diferentes proyectos de investigación.

### Introducción

La administración y control del avance de proyectos es de máxima utilidad en aquellas organizaciones que quieran definir una estrategia y llevarla a cabo con un seguimiento sistemático.

El control de gestión requiere de una interpretación amplia en torno a los elementos básicos que intervienen en el proceso, por un lado se tiene que el control parte desde un objetivo definido dentro de parámetros de alcance de logros, y por el otro, el control exige técnicas específicas para llevarlo a cabo

de una manera efectiva dentro de un contexto organizacional concreto. Lo primero lo define un determinado marco de planificación y lo segundo la sistematización operativa del control [4], [5], [6].

La planificación estratégica como sistema, contiene un determinado número de procesos que hacen de ella una actividad dinámica, flexible y continua. Esta complejidad no viene dada por las dificultades propias del proceso en torno a la aplicación de sus métodos y técnicas, sino más bien corresponde a la dinámica de la realidad que desborda los más elaborados análisis del entorno y hacen de ésta, un instrumento racionalizador de las aspiraciones organizacionales en un mundo de incertidumbres constantes.

La planificación estratégica asume un rol importante en la construcción de escenarios factibles dentro de cierta certidumbre esperada. Para ello, establece un conjunto de procesos integrados capaces de sistematizar acciones que permitan lograr dentro de un contexto específico, objetivos y metas productivas.

En la actualidad, para las organizaciones resulta crítica la toma de decisiones basadas en información relevante, exacta y actual. Dada la vasta cantidad de información que las organizaciones generan hoy en día, es necesario contar con una herramienta que permita manipular tal cantidad de información y que la presente en una forma resumida y con extrema facilidad de uso.

Si bien los Sistemas de Información existen desde hace varias décadas, su

incursión en las organizaciones, desde el punto de vista estratégico, y su subsecuente establecimiento como ventaja competitiva en las mismas es algo más reciente. Los Sistemas de Información se usan para recoger y almacenar información, para producir reportes específicos para los distintos integrantes y directivos de la organización. Sin embargo, los directivos subutilizan estos sistemas directamente y piensan que la información les resulta de poco provecho si no tienen la posibilidad de explorar detalles ocultos.

Una vez definidos los objetivos y metas, se definen los indicadores para medir el éxito o fracaso del logro de estos objetivos y metas. Estos indicadores permiten medir la estrategia, con lo cual se logra:

- Confirmar las relaciones causa-efecto definidas entre objetivos: esto es importante ya que en el caso que dicha relación no se cumpla se debería replantear la estrategia para evitar errores posteriores.
- Confirmar o desaprobado la habilidad de la organización para alcanzar lo que había planteado: esto permitirá realizar acciones correctivas en el momento adecuado para contribuir al éxito de la aplicación de la estrategia.
- Contribuir al compromiso de los integrantes con la estrategia: comunicar el éxito o fracaso del alcance de los objetivos ayuda a enfocar a los integrantes de la organización con los temas que interesan.

Hay dos tipos de indicadores:

- Resultado: reflejan la eficiencia de la organización en el alcance de los objetivos.



- Desempeño: reflejan el progreso de las acciones que propician el logro de los objetivos.

Debe existir una relación armoniosa entre indicadores de resultado y desempeño para cada objetivo.

La definición de las características de los indicadores es vital para el éxito de la obtención de los objetivos. Un buen indicador debe:

- Ser operacional: enfocado a la acción y que suministre información útil para el control.
- Medirse contra un estándar de logro: definir los valores esperados para cada indicador.
- Ser confiable: que pueda conseguirse y evaluar fenómenos de tipo cualitativo y cuantitativo.
- Medirse fácilmente.
- Tener bien especificado el procedimiento de obtención de datos, incluyendo las fuentes de información.

## Líneas de investigación y desarrollo

En este punto se plantea la propuesta de un plan de proyecto como técnica subyacente para la implementación del Balanced Scorecard en su etapa de ejecución [1], [8], [9], [10], [11].

Una vez planificado el proyecto, construidos la estrategia y el mapa estratégico, asignado a cada unidad de negocio o servicio compartido (incluso a nivel de cada persona o actor) su propio Balanced Scorecard para generar la sinergia correspondiente, comienza la etapa de ejecución del proyecto.

Las técnicas más conocidas para control de proyectos se basan en el control de tiempos, y en casos de perfeccionamiento de éstas, usan el control de recursos. Sin embargo, ninguna de ellas satisface plenamente las necesidades de medición tales como las plantea el Balanced Scorecard [2], [3], [7].

Esta técnica la hemos dado en llamar "Plan de Proyecto" ("Project Plan") y comprende la etapa de planificar y construir la estrategia y su correspondiente Balanced Scorecard, tal como se describió en el segundo párrafo del presente punto.

A partir de allí, comienza la etapa que hemos llamado "Ejecución del Proyecto" ("Project Running") y consiste en la implementación propiamente dicha del Balanced Scorecard, evaluando sistemáticamente el cumplimiento (o no) de los indicadores de gestión (si sus valores se encuentran dentro del rango aceptable) y culminando con el logro de las metas (o no) controlándolo a través de los indicadores de resultado.

Para cada uno de los actores (*stakeholder*) se plantea su propio plan de proyecto en donde se encuentra contemplada su meta (*goal*) a lograr, en el tiempo planificado y con los indicadores de gestión y de resultado que servirán para evaluar el avance del proyecto en su implementación.

En el modelo de implementación propuesto, la primer parte llamada Plan de Proyecto, responde a la pregunta: ¿Cómo debería llevarse a cabo la estrategia de manera tal que el proyecto sea exitoso?. En la segunda parte del modelo propuesto, llamada Ejecución del Proyecto, se responde a la pregunta: ¿Los valores reales de los indicadores están

reflejando que la estrategia se está llevando a cabo según lo planificado para que el proyecto sea exitoso?.

En este modelo cada stakeholder, por ejemplo el  $stakeholder_i$  puede ser una instancia de un cliente, un accionista, un empleado, un proveedor, alguien de la comunidad o, principalmente, una unidad de negocio o servicio soporte (compartido) de la organización. No necesariamente es una persona, sino que puede ser un equipo de personas trabajando en la misma dirección y produciendo sinergia para alcanzar una meta.

En los próximos tres párrafos se destacan algunas restricciones que se especifican, actualmente en este primer diseño, con el fin único de dar mayor simplicidad al modelo. Como toda restricción es molesta desde el punto de vista del manager que implementa el Balanced Scorecard, aunque también, por la misma razón, tiene el beneficio de simplificar la implementación por la reducción en la cantidad de variables que debe manejar.

Si una meta es tal que debe medirse por medio de varios indicadores, en el modelo propuesto, éstos deberán agruparse en un único indicador que será una conjunción de todos los necesarios.

Similarmente, si un stakeholder tiene varias metas a cumplir, en este modelo, deberán conjugarse en una única meta que no será otra cosa que la conjunción generalizada de todas las metas componentes.

Por otro lado, si un stakeholder cumple varios roles, por cuestión de simplicidad del modelo, deben darse tantas instancias de stakeholder como roles cumpla ese stakeholder.

Es por eso que en el modelo propuesto hay una implicación directa de stakeholder a goal.

Pasando a comentar los aspectos referidos a los indicadores, en el modelo, a cada indicador de gestión  $j$  se lo planifica con un valor ideal  $V_{ji}$  en un momento dado del avance de la ejecución del proyecto  $t_i$ , con un rango de aceptabilidad, es decir, valores de desviación  $d_{ji-k}$  y  $d_{ji+l}$ .

El indicador de gestión brindará el valor real actual  $AV_{ji}$  obtenido en un momento dado  $t_i$ , y se evaluará si está dentro del rango de aceptación respecto de su valor ideal. A partir de este valor, el stakeholder sabrá si necesita tomar acciones correctivas para ajustarse al valor deseado en la próxima medición del indicador o, si estando en el camino correcto, pero viendo una desviación en más o en menos (a derecha o izquierda respectivamente) le va dando una pista de que este indicador comienza a desviarse. Es decir, si bien en el momento actual está dentro de los límites aceptables, por la tendencia que está teniendo, probablemente en las próximas mediciones quede fuera del rango de aceptación.

Con los indicadores de resultado, el tratamiento tiene el mismo estilo aunque no es iterativo, sólo ocurre una única vez, y su resultado booleano determina si se cumplió la meta en el tiempo esperado dentro de los valores deseables.

## Resultados y Objetivos

Se espera obtener como resultado, un prototipo en el cual se implemente el Plan de Proyecto propuesto como mecanismo subyacente para soportar el Balanced

Scorecard en sus etapas de diseño y de ejecución.

Uno de los tópicos destacados dentro de los objetivos es la construcción de un sistema automático que implemente la planificación y su posterior control de stakeholders, metas e indicadores de desempeño y de resultado, proveyendo de información de planes/acciones alternativos/correctivos según los estados actuales de los indicadores respecto de los estados planificados.

## Formación de Recursos Humanos

Las tareas realizadas en el contexto de la presente línea de investigación están siendo desarrolladas como parte de trabajos para optar al grado de Licenciado en Ciencias de la Computación. En el futuro se piensan generar diferentes tesis de maestría y doctorado a partir de los resultados obtenidos en los trabajos de licenciatura en curso.

## Referencias

- [1] G. Montejano, N. Debnath, M. Pérez Cota; "Can avoid the present "Babel Tower" effect, caused by the uncontrolled proliferation of management methods?", 10th IEEE International Conference on Industrial Informatics (INDIN), 2012, Beijing, China, 25-27 July 2012.
- [2] Robert Kaplan, David Norton, "Cómo utilizar el Cuadro de Mando Integral: para implementar y gestionar su estrategia", Barcelona – España, Ed. Gestión 2000 S.A., 2001.
- [3] Nils-Göran Olve, Jan Roy, Magnus Wetter, "Implantando y Gestionando el Cuadro de Mando Integral (Performance Drivers)", Ed. Gestión 2000, 2000.
- [4] Robert Kaplan, David Norton, "The Strategy Focused Organization", Harvard Business School Publishing Corporation, 2001.
- [5] R. Uzal, G. Montejano, D. Riesco, N. C. Debnath, A. Dasso, A. Funes, "A General Architecture for Decision Support Systems Based on Kaplan & Norton Concept of Balanced Scorecard" *Proceedings of the International Conference on Computer and Information Science*, Orlando – Florida – USA, October 2001.
- [6] John Rockart, "Chief Executives define their own data needs", *Harvard Business Review*, March-April 1979.
- [7] Paul Arveson, "What is the Balanced Scorecard?", The Balanced Scorecard Institute, <http://www.balancedscorecard.org>, 1998.
- [8] G. Montejano, R. Uzal, D. Riesco, N. C. Debnath, "Towards a Formal Specification for Balanced Scorecard Domain", *Proceedings of the International Conference on Computer Science, Software Engineering, Information Technology, e-Business, and Applications*, Iguazu, Brazil, June 2002.
- [9] G. Montejano, R. Uzal, D. Riesco, "Formalizing the Underlying Project Plan of the Balanced Scorecard", *Proceedings of the 8<sup>th</sup> Argentine Conference in Computer Science*, Buenos Aires - Argentina , October 2002.
- [10] N. C. Debnath, G. Montejano, R. Uzal, D. Riesco, "Formalization of the Balanced Scorecard Implementation Viewed as Project Plan", *Information, An International Interdisciplinary Journal*, Vol. 7 Pag. 69-80, Japan, January 2004.
- [11] N. Debnath, R. Uzal, G. Montejano, D. Riesco, "Integrating Balanced Scorecard Domain Formalization with Underlying Project Plan Formalization", *Proceedings of the 3<sup>rd</sup> IEEE / ACS International Conference on Computer Systems and Applications*, Cairo – Egypt, January, 2005.

# METODOLOGÍA PARA EL PROCESO DE ANÁLISIS DE ADOPCIÓN DE CLOUD COMPUTING PARA PYMES

**Bernal, L., Vegega, C., Pytel, P., Pollo-Cattaneo, M.**

Grupo de Estudio en Metodologías de Ingeniería de Software

Universidad Tecnológica Nacional. Facultad Regional Buenos Aires.

Medrano 951 (C1179AAQ) Ciudad Autónoma de Argentina. Buenos Aires Tel +54 11 4867-7511

[bernal.luciano@gmail.com](mailto:bernal.luciano@gmail.com) , [fpollo@posgrado.frba.utn.edu.ar](mailto:fpollo@posgrado.frba.utn.edu.ar)

## RESUMEN

El paradigma de Cloud Computing ayuda a optimizar los procesos de almacenamiento y manejo de datos, haciendo más eficaz la toma de decisiones en una organización. Se ha observado la falta de una metodología homogénea que permita analizar la conveniencia y la viabilidad de la adopción de esta tecnología dentro de las Pequeñas y Medianas Empresas (PyMEs). Por lo tanto, el presente proyecto tiene como objetivo proponer una metodología (con sus técnicas y herramientas asociadas) considerando las características de las PyMEs, los distintos modelos de costos de Cloud Computing y las oportunidades del mercado en general. De esta forma, los encargados de la toma de decisiones podrán analizar, en forma sistematizada, si la organización se encuentra capaz de migrar a este nuevo paradigma.

**Palabras Clave:** *Cloud Computing, Metodología, Toma de decisiones, Proceso de Análisis de Adopción, PyMEs.*

## Contexto

Este proyecto articula líneas de trabajo del Grupo de Estudio en Metodologías de Ingeniería de Software (GEMIS) de la Facultad Regional Buenos Aires de la Universidad Tecnológica Nacional (UTN-

FRBA).

## INTRODUCCIÓN

Según el Instituto Nacional de Estándares y Tecnología de Estados Unidos, el concepto de Cloud Computing hace referencia a un “modelo que permite habilitar acceso a la red, de forma conveniente y en demanda, a un fondo compartido de recursos computacionales configurables (redes, servidores, almacenamiento, aplicaciones y servicios) que puede ser provisto rápidamente y con un mínimo esfuerzo de administración o interacción con el proveedor” [Mell & Grance, 2011].

También se puede agregar lo que propone [Armbrust, 2009], incluyendo en este concepto al hardware y los sistemas software en los centros de datos (o datacenters en inglés) que proveen los servicios entregados por demanda. Estos servicios se denominan normalmente ‘Software como Servicio’ (o SaaS, por sus siglas en inglés), mientras que los recursos IT (hardware y software del centro de datos) necesarios es lo que se llama ‘Cloud’ o ‘la Nube’. Esta Nube se basa normalmente en la virtualización de recursos de hardware, cuya comercialización está acompañada de sistemas de software que permiten gestionar la arquitectura subyacente.

Para este trabajo se toman en cuenta los tres escenarios propuestos en [Espino Barrios, 2009] para su implementación en los centros



de datos. Específicamente se trabajará sobre el enfoque externo ('public clouds' o arquitecturas de nube públicas donde se comparten los recursos) y el interno ('private clouds' o nube privada).

Si bien existen herramientas que faciliten el cálculo de costos y beneficios económicos de migrar los datos desde servidores internos de la empresa a los servicios de almacenamiento en la Nube [Microsoft Corporation, 2010], éstas dejan de lado otros aspectos importantes a la hora de considerar moverse al nuevo modelo de aprovisionamiento. Por ejemplo, tanto la herramienta Windows Azure TCO de [Microsoft Corporation, 2011] como la RDS Cost Comparison Calculator de [Amazon Web Services, 2011] permiten estimar los costos operacionales de los servicios que ofrecen las empresas mencionadas y determinar los beneficios económicos generados debido a la diferencia entre el mantenimiento de un centro de datos privado y los precios de los servicios en la Nube. Aunque los costos y los beneficios económicos es información útil para la toma de decisiones, sólo representan una porción de los factores que deben ser considerados en el proceso de migración a la Nube. También es importante analizar los factores organizacionales que afectan la infraestructura misma de la empresa y otros, como el impacto en los recursos humanos que van a lidiar con la nueva tecnología. En general, se observa una falta de procesos como el propuesto por [Fourcade] que definan tanto los requisitos de seguridad y portabilidad de las aplicaciones que se desean migrar, así como la estructura interna de la organización.

Por otro lado, en [Khajeh-Hosseini *et al.*, 2010] se especifica un modelo para investigar y analizar las decisiones en adopción de Cloud Computing considerando las repercusiones tanto técnicas como sociales. Pero, este conjunto de técnicas y

herramientas se focaliza en empresas de gran escala, que manejan gran cantidad de datos y aplicaciones, pasando por alto las peculiaridades de las organizaciones de menor tamaño como son las Pequeñas y Medianas Empresas (PyMEs).

Por tales motivos, este trabajo tiene como finalidad desarrollar una metodología (constituida en fases, con sus herramientas y técnicas asociadas) que permita analizar la conveniencia y viabilidad de la adopción de la tecnología de Cloud Computing dentro de una PyME.

### OBJETIVOS E HIPÓTESIS DE INVESTIGACION

Este trabajo se inscribe en una línea de investigación que pretende dar soporte a las personas que tomen decisiones sobre el modelo de IT empleado en un PyME principalmente facilitando el análisis para la adopción de Cloud Computing.

Entre los supuestos (o hipótesis) que guían este proyecto se encuentran los siguientes:

- *Hipótesis I:* Los riesgos y problemas encontrados durante la migración a la Nube pueden ser reducidos mediante su detección temprana a través de técnicas y herramientas correspondientes.
- *Hipótesis II:* Un proyecto de migración a la Nube posee características muy distintas a las de los proyectos de desarrollo de software tradicional. En consecuencia, los métodos existentes en la Ingeniería en Software para el estudio de Viabilidad no pueden ser aplicables.
- *Hipótesis III:* A pesar de existir metodologías para analizar la adopción de Cloud Computing, muchas de estas dejan de lado el estudio de factores propios de las PyMEs.

Teniendo en cuenta dichas hipótesis se buscará como *Objetivo General* proponer una metodología para realizar el análisis de adopción de Cloud Computing teniendo en cuenta las particularidades de las Pequeñas y

Medianas Empresas. Dentro de dicha metodología se incluye la definición de un conjunto de técnicas y herramientas, entre las que se encuentran el cálculo de costos para estimar el uso de recursos virtualizados y la creación de los precios según las necesidades de la empresa.

### RESULTADOS ESPERADOS

La finalidad de este proyecto es desarrollar una metodología que permita facilitar el análisis de la adopción de Cloud Computing en una PyME y así brindar soporte a la hora de decidir adoptar el modelo de recursos virtualizados escalables y flexibles.

En primer lugar, se pretende subsanar la falta de una metodología integradora que aúne herramientas de costeo y otras referidas al análisis de variables organizacionales. De esta manera, se evita el estudio detallado de cada caso en particular, reduciendo costos a la hora de decidir si adoptar o no el modelo de computación en la Nube.

Por otro lado, se busca facilitar de forma general, sistematizada, y racional el proceso de toma de decisiones para todas aquellas empresas que cumplan con las características de una PyME. Para ello se deberá haber definido antes qué es lo que distingue a las PyMEs de otro tipo de organizaciones, en cuanto a la arquitectura que manejan para el almacenamiento y la administración de sus datos.

Al mismo tiempo, se espera que el proyecto logre motivar a los implicados, promoviendo así a los dueños de PyMEs sobre los beneficios de la Nube y la conveniencia de migrar a dicha arquitectura.

### FORMACIÓN DE RECURSOS HUMANOS

Este proyecto pretende tanto la obtención de nuevos conocimientos, como la mejora en la formación del recurso humano. El grupo de trabajo se encuentra formado por dos investigadores tesistas de doctorado, un

investigador tesista de maestría y un investigador en formación.

Además se encuentra en desarrollo un Trabajo Final de Grado que busca formar especialistas en el análisis de adopción de procesos vinculados con la Ingeniería de Software. Por otro lado, en el marco de este proyecto de investigación se prevé la radicación de una Tesis de Maestría en Ingeniería en Sistemas de Información

### REFERENCIAS

Amazon Web Services, 2011, *User Guide: Amazon RDS Cost Comparison Calculator*. Disponible en: <http://aws.amazon.com/economics>

Armbrust, M., 2009, *Above the Clouds: A Berkeley View of Cloud Computing* Electrical Engineering and Computer Sciences, University of California at Berkeley, California.

Espino Barrios, L., 2009, *Cloud Computing como una red de servicios*, Instituto Tecnológico de Costa Rica, Costa Rica. Reporte Técnico MC-6006-200869282.

Fourcade, G., S/A, *Seis variables para analizar antes de saltar a la nube*. IBM. Disponible en: [http://www.ibm.com/bo/services/pdf/final\\_seis\\_variables\\_para\\_analizar\\_antes\\_de\\_saltar\\_a\\_la\\_nube.pdf](http://www.ibm.com/bo/services/pdf/final_seis_variables_para_analizar_antes_de_saltar_a_la_nube.pdf)

Khajeh-Hosseini, A., Greenwood, D., Smith, J. & Sommerville, I., 2010, *The Cloud Adoption Toolkit: Supporting Cloud Adoption Decisions in the Enterprise*. Cloud Computing Co-laboratory, School of Computer Science. University of St Andrews, UK.

Mell, P. & Grance, T., 2011, *The NIST Definition of Cloud Computing*, National Institute of Standards and Technology. Estados Unidos.

Microsoft Corporation, 2010, *The economics of the cloud*. Disponible en: <http://www.microsoft.com/en-us/download/details.aspx?id=5166>

Microsoft Corporation, 2011, *Microsoft Calculator*. Disponible en: <http://www.windowsazure.com/es-es/pricing/calculator>

# Métodos y Tecnología Informática aplicada al desarrollo de Sistemas de Gerenciamiento Energético en apoyo a ISO 50001

Leopoldo Nahuel<sup>1</sup>, José Maccarone<sup>2</sup>, Javier Marchesini<sup>1</sup>, Marcelo D Ambrosio<sup>2</sup>, Laura Cantalops<sup>1</sup>

<sup>(1)</sup> Departamento de Ingeniería en Sistemas de Información

<sup>(2)</sup> Departamento de Electrotecnia

Facultad Regional La Plata – Universidad Tecnológica Nacional – Av. 60 esq. 124 s/n – CP 1900

{lnahuel, jmacarone, jmarchesini, mdambrosio, lcantalops}@frlp.utn.edu.ar

## RESUMEN

El objetivo de este proyecto es realizar actividades de I&D en temas relacionados a Ingeniería de Software Basado en Modelos (ISBM) y Common Information Model (CIM) en conjunción con métodos de producción de sistemas software para gestión de eficiencia energética, siguiendo bases de la norma internacional ISO 50001. Con el interés particular de evaluar e implementar mecanismos que automatizan actividades del gerenciamiento energético, resulta de valor importante explorar distintas áreas de la disciplinas de ingeniería eléctrica para una mejor captura de requerimientos software – de aplicación genérica – en ámbitos industriales de la región, con fuerte necesidad de administración de valores energéticos medibles en términos económicos. En este contexto es necesario establecer un marco de trabajo para construir herramientas informáticas que ofrezcan mecanismos para automatizar alertas y controles sobre dominios específicos, en un ambiente de administración, control y gerenciamiento energético. Considerando aspectos que incluyen extensión de funcionalidad para distintos tipos de energía y evolución de la reciente norma ISO 50001, se presenta en este trabajo una propuesta de herramienta software de soporte a la gestión y ahorro de energía eléctrica. Resulta finalmente un valor importante en este proyecto, integrar actividades I+D+i multidisciplinario entre distintos departamentos de carrera: Electrotecnia & Sistemas de Información.

**Palabras clave:** *Herramienta Software, Common Information Model (CIM), Gestión Automatizada, Eficiencia Energética, Norma internacional ISO 50001.*

## CONTEXTO

Este proyecto de investigación multidisciplinario fue iniciado como parte del Programa anual FORMATEC [1] de la Secretaría de Vinculación Tecnológica de la Universidad Tecnológica Nacional, con el fin de ofrecer mecanismos para la mejora de eficiencia energética en procesos industriales de PyMEs regionales, situadas en el Sector Industrial Planificado (SIP) [2] del Polígono Industrial de la ciudad de Berisso, Prov. Buenos Aires. En este contexto, el objetivo general de nuestra propuesta es proveer un marco de trabajo metodológico basado en Common Information Model (CIM) [3,4] para el desarrollo de una herramienta software - extensible en distintos ámbitos energéticos - que permita alertar a los administradores de energía, distintos aspectos valorados en la normativa internacional ISO 50001 [5], como: multas por traspaso de potencias eléctrica contratadas, controles de consumo eléctrico, reportes comparativos de potencias, entre otros.

Este proyecto de investigación y desarrollo, fue iniciado a través de una acción conjunta entre investigadores de la especialidad de Ingeniería en Sistemas de Información e Ingeniería Eléctrica, de la Universidad Tecnológica Nacional (UTN) - Facultad Regional La Plata, persiguiendo importantes aspectos de innovación en sistemas software, tanto en la forma de construcción como su valiosa aplicación al ámbito de mejora continua de la eficiencia energética [6,7].

Este proyecto es actualmente financiado por Programa de Subsidios para Vinculación y Transferencia Científica y Tecnológica (FORMATEC) del Rectorado UTN y por la



Secretaría de Extensión Universitaria de UTN-FRLP. Esta propuesta se encuentra formalmente enmarcada dentro de un PID UTN, que fue presentada para su homologación en el corriente año, ante la comisión de evaluación científica-tecnológica de la Secretaría de Ciencia, Tecnología y Posgrado del Rectorado UTN.

## 1. INTRODUCCION

La metodología Ingeniería de Software Basada en Modelos (ISBM) [8], ofrece un marco de trabajo donde los modelos son considerados entidades centrales durante el ciclo de construcción del software. Estos modelos requieren ser escritos en lenguajes de modelado estándar como lo son UML (Unified Modelling Language) [9,10] y SysML (System Modelling Language) [11], ambos dos estandarizados por el consorcio de desarrolladores de objetos OMG (Object Management Group) [12]. UML es un lenguaje gráfico de propósito general para la construcción de sistemas software, mientras que SysML es un lenguaje también gráfico, pero orientado a la construcción de sistemas computacionales (incluyendo componentes híbridos de hardware y software).

Siguiendo las bases metodológicas de ISBM, resulta interesante hacer uso del estándar definido y publicado por "Distributed Management Task Force" (DMTF) [13] denominado "Modelo de Información Común" (Common Information Model, CIM), que define la forma de gestionar los artefactos esenciales y globales en un entorno de IT, a través de un perfil UML que representa los distintos elementos comunes y su interrelación. CIM está destinado a permitir una gestión coherente de estos elementos comúnmente relacionados, independientemente de su fabricante o proveedor, ofreciendo una definición común de información de gestión para sistemas, redes, aplicaciones y servicios.

El estándar CIM se compone de una especificación de infraestructura y un esquema: el Esquema proporciona las descripciones de los modelos reales, mientras que la especificación define los detalles para la integración con otros modelos de gestión.

- La **Especificación de Infraestructura CIM** define la arquitectura y los conceptos de CIM, incluyendo un lenguaje mediante el cual se define el esquema CIM (incluyendo cualquier esquema de extensión), y un método para el mapeo de CIM para otros modelos de información, tales como SNMP [14]. La arquitectura de CIM se basa en UML, por lo que es orientado a objetos: los elementos gestionados se representan como clases de CIM y eventuales relaciones entre ellos se representan como asociaciones CIM. La herencia permite la especialización de los elementos básicos comunes en elementos más específicos derivados.
- El **Esquema CIM** es un esquema conceptual que define el conjunto específico de objetos y relaciones entre ellos que representan una base común para los elementos manejados en un entorno de IS. El esquema CIM cubre la mayor parte de los elementos actuales en un entorno de IT, por ejemplo, sistemas informáticos, sistemas operativos, redes, middleware, servicios y almacenamiento. El esquema CIM define una base común para la representación de estos elementos administrados. CIM tiene la característica de ser extensible., permitiendo ampliaciones específicas de los productos, y así se pueden definir los elementos comunes gestionados a través de este esquema...

Sobre los fundamentos antes mencionados, se instanciaron los elementos comunes para la construcción de un sistema software extensible para el ámbito de Gestión Energética, siguiendo lineamientos definidos en la norma ISO 50001.

Además de modelar la arquitectura central del sistema, se implementó un prototipo para administración y control de energía eléctrica, basado en las restricciones de dominio del CIM. Para la implementación se utilizó tecnología J2EE y SWT (Standard Widgets Toolkit) [15] sobre plataforma Eclipse para el desarrollo y MySQL como motor de base de datos.

La primera parte del proyecto, consistió en el desarrollo de un Prototipo de Interfaz de Usuario (PIU) del sistema Software, como posible solución a la problemática. Para llevar adelante dicho desarrollo, se llevaron una serie de tareas correspondientes al proceso de desarrollo de Software. Se partió de la Ingeniería de requerimientos, de tal manera de poder capturar los requerimientos de software y atributos de calidad, explorando el campo de las distintas áreas de las disciplinas de la ingeniería eléctrica. Posteriormente los requerimientos fueron validados en conjunto con lo Stakeholders para continuar la etapa de Análisis. Una vez establecido el comportamiento y teniendo una clara definición de los requerimientos, se procedió a realizar el PIU. Para el desarrollo del PIU se utilizó SWT (Standard Widgets Toolkit), bajo el IDE de desarrollo Eclipse en su versión Galileo, con el apoyo de integrantes especializados en el área de Diseño en Comunicación Visual, permitiendo definir criterios de accesibilidad y usabilidad adecuados, como así también una correcta organización de los componentes que forman parte del PIU.

A continuación, en la Fig. 1 y Fig. 2, se detallan algunas de las pantallas de la Interfaz de Usuario de la herramienta software.

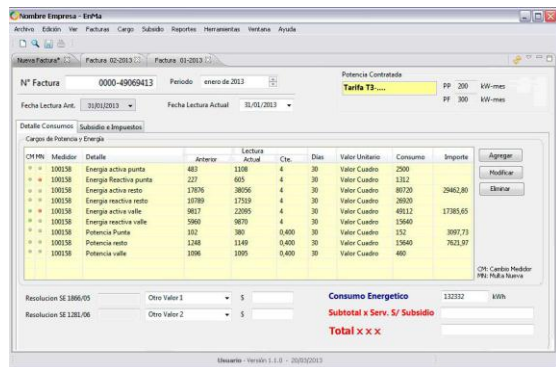


Fig. 1: Pantalla de carga de Factura y consumos

En la Fig. 1 puede observarse la interfaz para carga de datos generales y específicos de usuario (cabecera) y los consumos eléctricos que se registran en la factura (grilla). Se cargan los datos generales de la factura, como el número, el periodo tarifario y las fechas correspondientes a la toma de las mediciones. En la misma pantalla, se colocaron indicadores que permite conocer la existencia de multas por exceso de consumo, o indicarnos la existencia

de un cambio de medidor (círculos de color verde y rojo en la grilla de consumos).

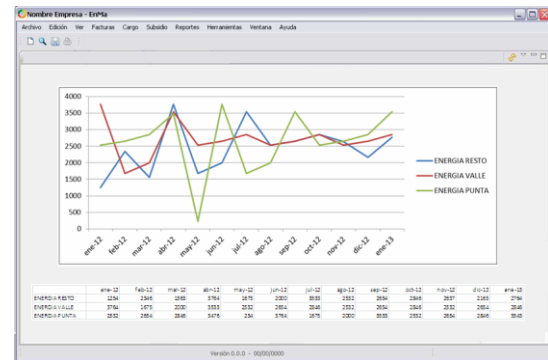


Fig. 2: Pantalla de reportes gráficos comparativos

En la Fig. 2 se observa una pantalla de reporte grafico, correspondiente a los consumos de Energía (Energía Valle, Energía Resto y Energía en Punta), y el ponderado de consumo a lo largo de un periodo. Estos reportes comparativos son de suma importancia para el gerenciamiento energético, ya que nos permite ver la tendencia a lo largo del periodo, y brinda información suficiente para una toma de decisiones sobre la gestión eficiente de la energía.

## 2. LINEAS DE INVESTIGACION y DESARROLLO

En la tarea de investigación que estamos llevando adelante, será necesario, un avance en el estudio y análisis de estándares y herramientas que nos permitan la integración de métodos de producción de sistemas software con los elementos referentes a la gestión energética, bajo la normativa ISO/IRAM 50001. Bajo esta línea de trabajo, serán analizados los siguientes estándares y herramientas:

- Ingeniería de Software Basada en Modelos (ISBM) : un paradigma que sitúa a los modelos como las entidades fundamentales que dirigen el proyecto de construcción de software
- Lenguajes que nos provean lo necesario para el modelado de nuestro sistema en forma correcta, siguiendo buenas practicas de la Ingeniería de Software Basada en Modelos,

en un entorno MDD (Model Driven Development) [16,17]

- **Common Information Models (CIM):** El CIM es un modelo de información estándar para empresas eléctricas, basado en lenguaje UML (Unified Modeling Language). En este modelo se representan objetos del mundo real y sus relaciones, con el propósito de crear un sistema de información que pueda ser utilizado entre diferentes aplicaciones para el manejo e intercambio de datos[ Análisis del estado del arte y de la practica en la aplicación del modelo CIM en empresas eléctricas
- **Domain-Specific Language (DSL)** [18]: es un lenguaje declarativo que ofrece una importante expresión centrada en un dominio determinado de un problema
- **Técnicas de Accesibilidad y Usabilidad para herramientas software.**

### 3. RESULTADOS Y OBJETIVOS

Se planifica avanzar en la capacitación continua de los miembros de la línea de investigación.

Como objetivo general se espera el desarrollo de una herramienta que de soporte tecnológico a la Gestión Energética de Micro y Pequeñas Empresas, con el interés de implementar mecanismos que automaticen las actividades de gerenciamiento energético apuntando a la reducción de consumos y utilización consciente de la energía eléctrica bajo la normativa ISO/IRAM 50001.

En esta primera etapa se concreto un Prototipo de Interfaz de Usuario (PIU) correspondiente al sistema software en cuestión, para administrar, la energía con mayor numero de variables, la energía eléctrica. Se reflejo en el PIU, el conjunto de herramientas necesarias para dar el soporte necesario para administrar la energía eléctrica, desde la carga de datos de facturas hasta reportes gráficos que permiten hacer un análisis de negocio para la toma de decisiones.

Como trabajo futuro se espera poder extender el sistema software, en el campo de gestión energética, a otros tipos de fuentes energéticas que puedan tener las MiPyMEs, como por

ejemplo el Agua, Gas, Combustibles, entre otros; a fin de poder gestionarlos para su reducción de consumos y el uso consciente de las energías.

### 4. FORMACION DE RECURSOS HUMANOS

Este trabajo, es parte de un proceso de incentivación para el desarrollo de actividades I&D, a desarrollar en el Laboratorio LINSI [19] del Departamento de Sistemas de Información en conjunción con el Departamento de Electrotecnia de la UTN - FRLP.

Actualmente el equipo de trabajo esta formado por dos directores, un director por parte del Departamento de Ingeniería Eléctrica y otro por parte del Departamento de Ingeniería en Sistemas, y 3 becarios, uno por parte de Ingeniería Eléctrica y dos correspondientes al área de Ingeniería en Sistemas, todos pertenecientes a la UTN - FRLP.

Esta propuesta se encuentra enmarcada bajo la formalidad de Proyecto de Investigación y Desarrollo (PID) UTN, sometido a homologación por la Secretaría de Ciencia, Tecnología y Posgrado de la Universidad Tecnológica Nacional. De este proyecto participan docentes-investigadores y alumnos becarios de investigación, de dos Departamentos de esta facultad regional: Departamento de Sistemas de Información y Departamento de Electrotecnia.

En este contexto y en pos de formación de recursos humanos en áreas de investigación científica-tecnológica, en el ámbito de esta regional, se realizarán dos Proyectos Finales de carrera grado y una Práctica Profesional Supervisada (PPS) en la especialidad Sistemas, sobre esta línea de investigación y desarrollo aplicado a la industria.

### 5. REFERENCIAS

[1] Programa anual FORMATEC de la Secretaría de Vinculación Tecnológica UTN. <http://www.utn.edu.ar/secretarias/vintecnologica/formatec.utn>

- [2] Información sobre el Sector Industrial Planificado (SIP) de Berisso. [http://www.mp.gba.gov.ar/sicm/agrupamientos/agrup\\_detalle/parque\\_detalle.php?id=1290536021](http://www.mp.gba.gov.ar/sicm/agrupamientos/agrup_detalle/parque_detalle.php?id=1290536021)
- [3] Sanchez Lopez, J.; Espinosa Reza, A., Garcia Espinosa, R.; *Análisis del estado del arte y de la practica en la aplicación del modelo CIM en empresas eléctricas*; Revista Tendencias Tecnológicas, Abril-Junio 2010.
- [4] CIM. [www.cimug.com](http://www.cimug.com)
- [5] IRAM / ISO 50001 (2011): *Sistema de Gestión de la Energía*, IRAM, Buenos Aires, Año 2011
- [6] *Estrategia de Ahorro y Eficiencia Energética 2004-2012* - Ministerio de Medio Ambiente - Madrid – IDEA, Año (2005).
- [7] *Las herramientas de monitoreo para la eficiencia energética en Europa*, París, ADEME-SAVE, Año 2000
- [8] Douglas C. Schmidt, *ISBM.Model-Driven Engineering*, Vanderbilt University, 2006 IEEE. <http://www.cs.wustl.edu/~schmidt/GEI.pdf>
- [9] J. Rumbaugh, I. Jacobson, Grady Booch. “*El Lenguaje Unificado de Modelado, Manual de Referencia*“. Addison Wesley, Primera Edición, 2000.
- [10] UML: [www.omg.org/spec/UML/](http://www.omg.org/spec/UML/)
- [11] Friedenthal S., Moore A., Steiner R., “*A Practical Guide to SysML*”, The Systems Modeling Language, Elseiver, Second Edition, 2011.
- [12] OMG: <http://www.omg.org>
- [13] DMTF: <http://www.dmtf.org>
- [14] Douglas Comer, *Redes TCP/IP*, Cap. 26, 3° Edición, Año 2000
- [15] Standard Widgets Toolkit. [www.eclipse.org/swt/](http://www.eclipse.org/swt/)
- [16] Roxana Giandini, Matías Mangano, Lautaro Mendez, Leopoldo Nahuel. “*La Producción de Software Dirigida por Modelos y la Filosofía Agil*”, PIPP, Año 2011
- [17] Pons Claudia, Giandini Roxana y Pérez Gabriela. “*Desarrollo de Software Dirigido por Modelos: conceptos teóricos y su aplicación práctica*”. 1er. edición. EDULP & McGraw-Hill Education, Argentina, ISBN-13:9789503406304
- [18] Arie van Deursen, Paul Klint, Joost Visser, “*Domain-Specific Language: An Annotated Bibliography*”, ACM SIGPLAN , Junio 2000.
- [19] Laboratorio de Innovaciones en Sistemas de Información, LINSI. [www.linsi.edu.ar](http://www.linsi.edu.ar)



# Sistemas de Información Gerencial en los procesos de negocio de las PyMEs de Jujuy

César A. Castillo<sup>1</sup> & Analía N. Herrera Cognetta<sup>1</sup> & Lía G. Rico<sup>1</sup> & Sergio Salva<sup>1</sup> Julián Sarmiento<sup>1</sup> & Mario A. Tejerina<sup>1</sup> & Laura R. Villarrubia<sup>1</sup>

Facultad de Ingeniería, Universidad Nacional de Jujuy

Av. Italia esquina José Martiarena (4600) San Salvador de Jujuy – Argentina

Tel. 0388-4221585

[ce\\_al\\_castillo@yahoo.com.ar](mailto:ce_al_castillo@yahoo.com.ar)

[anihco@yahoo.com.ar](mailto:anihco@yahoo.com.ar)

[liagrigo@hotmail.com](mailto:liagrigo@hotmail.com)

[herman\\_sergio@hotmail.com](mailto:herman_sergio@hotmail.com)

[sarmiento.julian@gmail.com](mailto:sarmiento.julian@gmail.com)

[mariotejerina@gmail.com](mailto:mariotejerina@gmail.com)

[lauraritavillarrubia@yahoo.com.ar](mailto:lauraritavillarrubia@yahoo.com.ar)

## 1 RESÚMEN

El equipo de investigación ha planteado como fundamento para estudio e investigación, la siguiente hipótesis: “la implementación de Sistemas de Información Gerencial en las Pequeñas y Medianas Empresas de la provincia de Jujuy, aportará grandes ventajas para su eficiente administración y crecimiento”. A fin de refutar o confirmar ésta, se ha iniciado el correspondiente relevamiento en las PyMES de la provincia, en lo concerniente a su situación actual, analizando además, sus necesidades de información, restricciones técnicas y/o económicas, para determinar así la factibilidad de implementar en ellas un Sistema de Información Gerencial. Una vez obtenidos los resultados, se propone construir un prototipo ideal de un Sistema de Información Gerencial de propósito general, aplicable a las PyMES de la provincia de Jujuy, específicamente del rubro servicios, e implementarlo a manera de prueba en ellas, para medir así los efectos que esta herramienta provoca en el mencionado ambiente de las PyMES.

**PALABRAS CLAVE:** Sistema de Información Gerencial- Pequeñas y Medianas Empresas de Jujuy- Investigación- Desarrollo- Prototipo – Implementación.

## 2 CONTEXTO

El proyecto se encuentra acreditado y financiado por la SECTER – UNJu (Secretaría de Ciencia y

Técnica y Estudios Regionales de la Universidad Nacional de Jujuy)

## 3 INTRODUCCIÓN

El eje principal del proyecto de investigación es el estudio y desarrollo de un Sistema de Información Gerencial (SIG) en el contexto y desempeño de las PYMES vigentes en la ciudad de S.S. De Jujuy. Basado en esta premisa es consecuente en primer lugar exponer un resumen teórico respecto a los SIG, para luego mencionar su importancia en el marco aplicativo de las PYMES.

Un SIG es particularmente importante en el medio social y económico actual, en el cual los gerentes necesitan tomar decisiones que incluyen aspectos tales como la capacidad financiera, la satisfacción de los clientes y el control de calidad. Los directores mencionados no pueden tomar decisiones que aseguren la supervivencia, ampliación y rentabilidad de sus programas y organizaciones, si el sistema de información gerencial no produce información completa, exacta y regular. Durand y otros (1999). De esta manera un sistema de información gerencial efectivo proporciona los datos necesarios a los administradores para analizar la situación actual, identificar problemas inmediatos y encontrar soluciones, descubrir patrones y tendencias que les permitan formular objetivos apropiados para el futuro y tomar decisiones inteligentes sobre el uso de recursos limitados tanto humanos como financieros y materiales.

En la actualidad las organizaciones necesitan información actualizada, confiable y completa, sobre todo para tomar decisiones acertadas sobre el desempeño y la ejecución de sus tareas. Según Durand y otros (1999) un sistema de información gerencial reúne datos sobre distintas funciones de una organización, lo que permite al administrador planear, hacer seguimiento y evaluar las operaciones y el desempeño de la empresa de manera integral.

Los SIG son una colección de sistemas de información que interactúan entre sí, y que proporcionan información tanto para las necesidades de las operaciones como de la administración. Sin embargo Laudon (2005) recalca que se trata de una colección de sistemas de información y no un sistema total. Agrega además que en teoría, una computadora no es necesariamente un ingrediente de un Sistema de Información Gerencial, pero en la práctica es poco probable que exista un Sistema de Información Gerencial complejo sin las capacidades de procesamiento y herramientas que hoy en día brindan las computadoras.

Se requiere experiencia, tiempo y dinero para crear un SIG que produzca información integrada, útil y completa. El desarrollo o mejoramiento de tal sistema, es, generalmente, una tarea organizacional de gran envergadura. Existen diferentes propuestas como guía de desarrollo, según Laudon (2005) el proceso puede dividirse en cuatro fases:

Fase 1: Conceptualización .La institución debe esclarecer sus necesidades y lleva a cabo una evaluación inicial de las alternativas viables. Al analizar esta fase ésta habrá elaborado un documento estratégico que presentará el curso de acción a seguir.

Fase 2: Evaluación detallada y diseño. La institución evalúa cuidadosamente los sistemas que podrían ser adquiridos. Si ésta ha decidido modificar un sistema existente o diseñar un sistema personalizado tendrá que analizar los problemas de diseño.

Fase 3: Desarrollo del sistema e implementación.

Fase 4: Mantenimiento del sistema y auditoría del SIG .En esta fase la institución examina aquellos aspectos que deberán ser tratados después de que el SIG ha sido desarrollado e implementado el mantenimiento del sistema, las modificaciones y las auditorías periódicas que se deberán llevar a cabo para asegurar que el sistema funciona adecuadamente.

Hoy en día las PyMEs revisten de gran importancia en las economías nacionales, sobre todo de los países en desarrollo, en los cuales no hay grandes y numerosas industrias donde se apoye la estructura económica. Las PyMEs aportan producción y empleo, garantizan una demanda sostenida, otorgan cierta dinámica al progreso y dan a la sociedad un saludable equilibrio.

Las empresas pequeñas tienen la ventaja, debido justamente a su tamaño, de poder abordar aquellas actividades que no pueden ser realizadas por las firmas mayores, lo que les da agilidad, flexibilidad y reflejos que son fundamentales para enfrentar las contingencias de un mercado complejo que se transforma constantemente.

Sin embargo, conducir PyMEs no es una tarea fácil, requiere una rigurosa gestión en la que no se encuentra el tiempo para incorporar competencias o mejorar las habilidades directivas imprescindibles para alcanzar el nivel necesario de eficacia y eficiencia.

La gestión severa demanda entre otras, información confiable en tiempo real de la cadena productiva de las empresas; es aquí donde se observa la importancia del rol de la informática; en la actualidad se ha convertido en uno de los aspectos más significativos a la hora de tomar decisiones certeras para mejorar el rendimiento de las empresas, ya sean grandes o pequeñas.

Aunque los SIG en un principio fueron privativos de las grandes firmas, en la actualidad se ha convertido en una vital necesidad de cualquier tipo de empresa. Los SIG se constituyen en una herramienta necesaria para el correcto funcionamiento de la productividad de las empresas porque ofrecen un control eficaz, permitiendo realizar una adecuada administración de los recursos de las compañías.

El control eficaz implica, entre otras tareas, tener en cuenta la oportunidad de negocios de la empresa, lo que implica poder analizar información real, oportuna y precisa, con la que se logra una toma de decisiones inteligente en el momento oportuno, con el fin de lograr el mejor desempeño de cada uno de los eslabones de la cadena productiva.

La realidad regional de la ciudad de S.S. De Jujuy demuestra la casi inexistencia del uso de SIG. Sin embargo existe una apertura y disposición importante por partes de sus integrantes para el estudio y desarrollo adecuado

de SIG donde el basamento de la información contenida conlleva a la puesta en práctica de manera eficiente del proceso de toma de decisiones, de cara a la crisis suscitada y vivenciada en el país, pero bajo un enfoque y una predisposición conductual de sus integrantes, tanto directivos como personal adscrito al mismo, y todo focalizado al éxito gerencial.

**4 LÍNEAS DE INVESTIGACION Y DESARROLLO**

- Conjunto de PyMEs de la Provincia de Jujuy: relevamiento de las empresas existentes en la Provincia de Jujuy e información relacionada a su actividad económica, rubros, recursos humanos, debilidades y fortalezas.
- Análisis del entorno: existencia de sistemas de información gerencial en la región y el impacto que produce en las PyMEs que los implementa.
- Organización y métodos: modo de operar, gestionar, dirigir y controlar sus respectivos procesos de negocio.
- Gerencia empresarial: modalidad en la toma de decisiones, presentación de la información, uso de estadísticas, reportes u otros métodos. Disponibilidad, fiabilidad, consistencia y precisión de la información usada actualmente para la toma de decisiones.
- Sistemas de Información Gerencial: análisis, interpretación conceptual, e investigación para la innovación y diseño de sistemas de información gerencial para las PyMEs.
- Tecnologías de Información como apoyo a los procesos de negocio: búsqueda y análisis de información de tecnología de información vigente relacionado a la mejora de los procesos de negocio de las PyMEs.
- Nuevos enfoques para mejorar los procesos de las PyMEs: enfoques innovadores usados por otras empresas similares a las investigadas, que aporten de alguna manera a la mejora de los actuales procesos de las mismas.

**5 RESULTADOS Y OBJETIVOS**

Objetivo General:

- Investigar, desarrollar e implementar un Sistema de Información Gerencial de

propósito general para las pymes de la provincia de Jujuy.

Objetivos Particulares:

- Realizar un estudio de viabilidad de la aplicación de los SIG a las PyMEs de la provincia de Jujuy.
- Construir un prototipo ideal de un SIG aplicable a las PyMEs de la provincia de Jujuy.
- Realizar pruebas piloto de implantación del SIG en algunas PyMEs seleccionadas.

Resultados parciales alcanzados:

- Conceptualización de los Sistemas de Información Gerencial.
- Conceptualización de PyME y situación actual en el ámbito de la provincia de Jujuy.
- Cuadros indicadores de clasificación de Pymes: tamaño-sector, industria, servicios informáticos, servicios de la producción y comercio.
- Cuadros indicadores del entrecruzamiento de variables en Pymes: fuentes de información para la toma de decisiones, nivel de conocimiento en tecnologías de información, utilización de SIG y nivel de competitividad.

**6 FORMACION DE RECURSOS HUMANOS**

La estructura del equipo de trabajo de la línea de I/D presentada es la siguiente:

<b>Rol/s e Integrante asignado</b>
<b>Director - Testing</b>
Villarrubia, Laura
<b>Investigador - Desarrollador</b>
Herrera Cognetta, Analia
Tejerina, Mario Alberto
<b>Investigador – Testing</b>
Rico, Lía Gabriela
<b>Desarrollador – Testing</b>
Castillo, Cesar

<b>Programador</b>
Salva, Hernán Sergio
Sarmiento, Julián Alberto Jesús

Los roles a desempeñar por los integrantes del grupo de investigación están acorde a las actividades a realizar en las diferentes etapas de investigación. Los roles y las actividades correspondientes a cada rol son:

#### Investigadores

- Análisis y diagnóstico de la situación actual.
- Evaluación permanente del trabajo.
- Estudios de factibilidad.
- Identificación de situaciones problemáticas.
- Búsqueda de fundamento teórico.
- Elaboración de casos de pruebas.

#### Desarrolladores

- Análisis, evaluación y adecuaciones de la implementación de herramientas.

#### Programadores

- Desarrollo de software específico en lenguajes de alto nivel.

#### Testing

- Evaluar Calidad durante el proceso
- Planificar y controlar reducción de riesgos, optimización de recursos y seguimiento de estándares.
- Prevenir, identificar y corregir fallas.
- Definición y validación de puntos de control en avance de planificación de actividades.
- Control de solapamiento de desarrollo de aplicaciones.

#### Director

- Planificación, Organización, Coordinación, Dirección y Control del proyecto.
- Liderar, motivar y comunicar.

#### *Formación de los recursos humanos*

Integrante: Laura Rita Villarrubia

- Seminario Taller “Herramientas de Marketing en internet” dirigido a empresarios de Pymes del sector

Turismo. 9 y 10 de Agosto de 2012.  
Instructor a cargo: Gustavo Saientz,  
Director Regional de Operaciones  
Resultics.

Integrante: Lía Gabriela Rico

- Cursando la Maestría en Administración de Negocios. Tesis “Planeación Estratégica de los Sistemas de Información”. UCASAL. Cursado: 2010-2011.

Integrante: Analía Herrera Cognetta

- Cursado de la Maestría en Informática- Universidad Santo Tomás de Aquino- Facultad de Ingeniería – Tucumán.

## 7 REFERENCIAS

- CARLOS CLERI, El libro de las PyMES, Edición 2007 - Ediciones Granica S.A. ISBN 978-950-641-500-6
- ALBERTO R LARDENT, Sistemas de Información para la Gestión Empresaria. Planeamiento Tecnología y calidad, Prentice Hall - Brasil, 2001
- KENNETH C. LAUDON Y JANE P LAUDON, Sistemas de Información Gerencial - Administración de la Empresa Digital, PEARSON EDUCACION - México, 10ma Ed, 2008.
- JAMES A. O'BRIEN, GEORGE M. MARAKAS, Sistemas de Información Gerencial, Mc Graw Hill - México, 7ma Ed -2006.
- JAVIER COLLADO, RAUL H. SAROKA, Informática en las organizaciones, EDICON - Argentina, 2010.
- JACOBSON I., El proceso Unificado de Desarrollo de Software, Pearson Educación, Madrid, 2003.
- LARMAN C., UML y Patrones – Una introducción al análisis y diseño orientado a objetos y al proceso unificado, Pearson Educación, Madrid, 2003.



- OMG, Unified Modeling Language Specification, [www.omg.org](http://www.omg.org), 2001
- Accenture Argentina, Aceleración para el desarrollo de sistemas y procesos, Pearson – Argentina, 2010
- Kenneth C. Laudon & Jane P Laudon.. Management Information, Managing the Digital Firm (9th Edition). 2005.
- DURAND, Edward & ESCALONA, José Gregorio & NÚÑEZ, Merlly SISTEMAS DE INFORMACIÓN GERENCIAL. 1999.

## Arquitecturas Distribuidas para Gobierno Electrónico

† Karina Cenci, <sup>II</sup>Pablo Fillottrani, <sup>††</sup>Jorge Ardenghi

Departamento de Ciencias e Ingeniería de la Computación  
Universidad Nacional del Sur

Av. Alem 1253 - (8000) Bahía Blanca - Argentina

Tel/Fax: (+54) 291-4595135/6

‡ E-mail: kmc@cs.uns.edu.ar

<sup>II</sup> Comisión de Investigación Científicas

E-mail: prf@cs.uns.edu.ar

†† E-mail: jra@cs.uns.edu.ar

### Resumen

La inserción de la tecnología en la sociedad, a través de una gran variedad de dispositivos, en especial los inalámbricos con alto poder de cómputo y comunicación, promueven cambios y desarrollos nuevos. La utilización de sistemas de información se incrementa día a día en organizaciones, empresas, agencias, usuarios, ciudadanos. El acceso a la información se puede realizar desde diferentes locaciones y horarios.

Este proyecto propone integrar los conceptos de arquitecturas, frameworks de desarrollo aplicados a sistemas distribuidos para aplicaciones de gobierno electrónico. Este tipo de aplicaciones requiere propiedades que son necesarias para alcanzar aceptación entre los diferentes usuarios.

**Palabras clave:** Sistemas distribuidos, arquitecturas, frameworks, gobierno electrónico.

### Contexto

Este trabajo de investigación se desarrolla en el Departamento de Ciencias e Ingeniería de la Computación de la Universidad Nacional del Sur (UNS). En particular, como parte de las tareas que se realizan en el Laboratorio de Investigación y Desarrollo en Ingeniería de Software y Sistemas de Información (LIS-SI) y en el Laboratorio de Investigación en Sistemas Distribuidos (LISIDI). El proyecto se financia parcialmente con fondos asignados por la UNS a proyectos de investigación.

### Introducción

El proyecto se centra en la exploración e integración de arquitecturas de sistemas, propiedades dese-

bles y frameworks de desarrollo para sistemas distribuidos, considerando especialmente, las aplicaciones de gobierno electrónico.

Los sistemas de información están penetrando día a día en todos los aspectos de nuestras vidas. En el último tiempo, el avance tecnológico en las comunicaciones, en especial, el auge de las redes inalámbricas, ad-hoc, dispositivos móviles; facilitan a los usuarios el acceso a la información. El desarrollo de sistemas y aplicaciones distribuidos de calidad y seguros es altamente requerido. Gobierno electrónico es uno de los más complejos e importantes sectores donde los desarrollos tecnológicos en sistemas de información se pueden aplicar. Este tipo de sistema se puede implementar en pequeña, mediana y gran escala. La relevancia de sistemas de gobierno electrónico está en obtener un mejor resultado de las políticas, una mayor calidad de los servicios públicos, eficiencia en los procesos de gobierno, un uso eficaz de fondos públicos, aumento de la participación ciudadana, etc.

La arquitectura de un sistema de software es el conjunto de las principales decisiones de diseño que se realizan sobre el sistema [27]. Un aspecto relevante para el desarrollo y eficiencia de sistemas distribuidos es el paradigma de diseño o estilo arquitectónico.

Los frameworks son herramientas útiles para incrementar la productividad y calidad en el desarrollo de aplicaciones. Como contrapartida, los frameworks pueden introducir inconvenientes en el desarrollo y evolución de nuevos proyectos relacionados si dependen altamente de la tecnología de las compañías. Esto puede considerarse como una barrera tecnológica en un nuevo proyecto.

## Líneas de Investigación y Desarrollo

Los sistemas de información facilitan a los usuarios la realización de tareas cotidianas. Los usuarios pueden ser empleados de una organización, ciudadanos, empresas, agencias de gobierno, dirigentes, etc. En la actualidad, los usuarios pueden encontrarse ubicados en diversas locaciones, con distintos husos horarios y culturales.

Ejes centrales de investigación del proyecto son los siguientes temas:

- Aplicaciones Distribuidas. Exploración, estudio y análisis de las propiedades requeridas para el desarrollo de las mismas. Considerando como caso especial, las aplicaciones de gobierno electrónico.
- Arquitecturas Distribuidas. Estudio y análisis de arquitecturas que se han aplicado en las aplicaciones de usuarios. Exploración de arquitecturas referentes para la implementación de sistemas distribuidos.
- Búsqueda y análisis de frameworks para modelar la arquitectura de sistemas distribuidos y para el desarrollo de los mismos.

## Resultados y Objetivos

Se está comenzando con el desarrollo de este proyecto. En la primera etapa, se está realizando la exploración de arquitecturas y frameworks para sistemas de software en especial para entornos distribuidos y de gobierno electrónico. Algunos ejemplos representativos son los siguientes.

- Malek y otros [17] proponen un framework extendido para mejorar el despliegue de la arquitectura de un sistema de software distribuido. El objetivo de la actividad de análisis del framework es garantizar un efectivo despliegue del modelado del sistema, tanto antes y durante la ejecución del mismo. Este análisis está basado en un conjunto de algoritmos específicamente desarrollados para este propósito.

- Kang y otros [15] proponen un framework denominado eGovFrame (e-Government Standard Framework). El mismo utiliza un conjunto de herramientas estandarizadas de software para desarrollo y ejecución de aplicaciones de gobierno electrónico. El objetivo es mejorar la eficiencia en las inversiones tecnológicas y la calidad de los servicios de gobierno electrónico.

El compartir la información es beneficioso para los usuarios de sistema de información, en especial esta ventaja se presenta para el desarrollo de gobierno

electrónico [5]. También es necesario considerar las barreras y riesgos que puede introducir en el despliegue. La integración de estos conceptos, (beneficios, barreras y riesgos), con las propiedades deseables de un sistema distribuido (flexibilidad, seguridad, adaptabilidad, interoperabilidad) servirán de soporte para la búsqueda, evaluación y comparación de arquitecturas y frameworks.

La posibilidad de plantear una arquitectura referente que sirva de molde para el modelado, desarrollo y ejecución de aplicaciones distribuidas es un desafío a investigar.

## Formación de Recursos Humanos

Los resultados de investigación serán utilizados para presentar una tesis de doctorado en Ciencias de la Computación. Además, esta línea de investigación permitirá la dirección de tesis de licenciatura y trabajos finales de ingeniería afines. Asimismo, servirán para generar courseware para materias optativas a dictarse en el Departamento de Ciencias e Ingeniería de la Computación de la UNS.

## Referencias

- [1] C. Baum and A. Di Maio. *Gartner's Four Phases of E-Government Model*. Gartner Group, 2000.
- [2] D. Beer, S. Höhne, H. Petersohn, T. Pöhnitzsch, and M. Voigt. Designing a distributed workflow system for e-government. In *Proc. of the 24th IASTED, Int. Conf. on Modelling, Identification, and Control*, pages 583–588, 2005.
- [3] D. Beer, R. Kunis, and G. Rünger. A component based software architecture for e-government applications. In *First International Conference on Availability, Reliability and Security*, 2006.
- [4] L. Caffrey. Information sharing between & within governments. *The International Council for Technology in Government Administration, London*, 1998.
- [5] K. Mendes Calo, K. Cenci, P. Fillottani, and E. Estevez. Information sharing - benefits. *JCST*, 12(2):49–55, 2012.
- [6] State Services Commission. New Zealand e-government interoperability framework (nz e-gif) version 3.3. Technical report, from e-Government in New Zealand: <http://www.e.govt.nz/library/e-gif-v-3-3-comple-te.pdf>, February 2008. Retrieved August 30, 2010.
- [7] S. Dawes. Interagency information sharing: Expected benefits, manageable risk. *Journal of Policy Analysis and Management*, 15:377–394, 1996.

- [8] S. Dawes and L. Prefontaine. Understanding new models of collaboration for delivering government services. *Communications of the ACM*, 46(1):40–42, 2003.
- [9] E. Estevez, P. Fillottrani, T. Janowski, and A. Ojo. Government information sharing - a framework for policy formulation. In *Governance and Cross-boundary Collaboration: Innovations and Advancing Tools*, chapter 2, pages 23–55. IGI Global, 2011.
- [10] E. Estevez, P. Fillottrani, and T. Janowski. Information sharing in government - conceptual model for policy formulation. *10th European Conference on eGovernment, (p.CD). Limerick, Ireland*, 2010.
- [11] P. Eugster, P. Felber, R. Guerraoui, and A. M. Kermerrec. The many faces of publish/subscribe. *Computing Surveys, ACM*, 35(2):114–131, 2003.
- [12] T. Field, E. Muller, and E. Law. *The e-Government Imperative*. Organization for Economic Co-operation and Development, 2003.
- [13] J. M. Gil-García, C. A. Schneider, T. A. Pardo, and A. M. Cresswell. Interorganizational information integration in the criminal justice enterprise: Preliminary lessons from state and county initiatives. In *Proceedings of the 38th Hawaii International Conference on System Sciences, IEEE*, 2005.
- [14] Q. Guo and C. Smidts. A facility framework for distributed application. In *Workshops of International Conference on Advanced Information Networking and Applications, IEEE*, pages 459–466, 2011.
- [15] G. Kang, S. Mun, and Y. Kwon. Development of e-government standard framework through open innovation strategy. In *ICACT 2012*, pages 1117–1122, 2012.
- [16] D. Lansbergen and G. Wolken. Realizing the promise: Government information systems and the fourth generation of information technology. *Public Administration Review*, 61(2), March-April 2005.
- [17] S. Malek, N. Medvidović, and M. Mikic-Rakic. An extensible framework for improving a distributed software system's deployment architecture. *IEEE Transactions on Software Engineering*, 38(1):73–100, 2012.
- [18] V. Meneklis and C. Douligeris. Extending a distributed system architecture with e-government modeling concepts. In *The 18th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, 2007.
- [19] J. L. Müller, A. Büchner, and P. Müller. A framework for the requirements analysis of service-oriented workflows. In *Third International Conference on Next Generation Web Services Practices*, pages 104–109, 2007.
- [20] NIEM Program Management Office NPMO. Niem concepts of operations. Technical report, from National Information Exchange Model: [http://www.niem.gov/files/NIEM\\_Concept\\_of\\_Operations.pdf](http://www.niem.gov/files/NIEM_Concept_of_Operations.pdf), January 2007. Retrieved August 30, 2010.
- [21] Ministry of Economic Affairs and Department of State Information Systems Communications. Estonian it interoperability framework. Technical report, from RISO-STate Information Systems: [http://www.riso.ee/en/files-/framework\\_2005.pdf](http://www.riso.ee/en/files-/framework_2005.pdf), 2005. Retrieved August 30, 2010.
- [22] Australian Government Information Management Office. Australian government information interoperability framework - sharing information across boundaries. Technical report, from Australian Government Information Interoperability Framework: [http://www.finance.gov.au/publications/agimo/docs-/Information\\_InteroperabilityFramework.pdf](http://www.finance.gov.au/publications/agimo/docs-/Information_InteroperabilityFramework.pdf), April 2006. Retrieved August 27, 2010.
- [23] Australian Government Information Management Office. National government information sharing strategy - unlocking government assets to benefit the broader community. Technical report, from Department of Finance and Deregulation: <http://www.finance.gov.au/publications/national-government-information-sharingstrategy/docs/ngiss.pdf>, August 2009. Retrieved August 27, 2010.
- [24] T. A. Pardo, A. M. Cresswell, S. S. Dawes, and G. B. Burke. Modeling the social & technical processes of interorganizational information integration. In *Proceedings of the 37th Hawaii International Conference on System Sciences, IEEE*, pages 1–8, 2004.
- [25] S. Ronaghan. *Benchmarking e-Government: A Global Perspective Assessing the UN Member States*. United Nations Division for Public Economics and Public Administration (UN-DPEPA) and American Society for Public Administration (ASPA), 2002. available at <http://unpan1.un.org/intradoc/groups/public/documents/un/unpan021547.pdf> (visited June 21, 2012).
- [26] K. Smolander. Four metaphors of architecture in software organizations: Finding out the meaning of architecture in practice. In *Proceedings of the 2002 International Symposium on Empirical Software Engineering*, 2002.
- [27] R. N. Taylor, N. Medvidović, and E. M. Dashofy. *Software Architecture. Foundations, Theory, and Practice*. Wiley, 2010.
- [28] E. Turner and P. Nicoll. *Electronic Service Delivery, including Internet Use, by Commonwealth Government Agencies*. Australian National Audit Office, Commonwealth of Australia, 1999. available at [http://www.anao.gov.au/media/Uploads/Documents/1999-2000\\_audit\\_report\\_18.pdf](http://www.anao.gov.au/media/Uploads/Documents/1999-2000_audit_report_18.pdf) (visited June 21, 2012).



- [29] X. Zhang, G. Han, B. Li, K. Zou, and W. Li. An e-government platform based on multi-tier architecture. In *Fifth International Conference on Information Technology: New Considerations, IEEE*, pages 1170–1172, 2008.

## Experiencias derivadas de la representación, modelización y construcción de simuladores en los ciclos lectivos 2011 - 2012

Sonia I. Mariño y María V. López<sup>1,2</sup>

Departamento de Informática. Facultad de Ciencias Exactas y Naturales y Agrimensura. 9 de Julio N° 1449.

3400. Corrientes. Argentina.

Facultad de Humanidades. Las Heras 727. 3500 Resistencia. Argentina

Universidad Nacional del Nordeste.

TE: (03783) 423126. e-mail: msonia@exa.unne.edu.ar; mvlopez@exa.unne.edu.ar

### CONTEXTO

Se presentan los avances logrados entre los años 2011-2012 derivados de la línea de trabajo de informática educativa aplicada a problemas de modelización y simulación enfatizando la formación de recursos humanos quienes participan en la selección de problemas reales, su representación y modelización y simulación. Las actividades se enmarcan en un proyecto de docencia e investigación que se desarrolla en la cátedra “Modelos y Simulación” del Departamento de Informática de la de la Facultad de Ciencias Exactas y Naturales y Agrimensura (Universidad Nacional del Nordeste) (FaCENA-UNNE), descrito en Mariño y López (2008), Mariño y López (2010), Mariño y López (2011).

### RESUMEN

“Modelos y Simulación” es una asignatura optativa de la Carrera de Licenciatura en Sistemas de Información, de la Facultad de Ciencias Exactas y Naturales y Agrimensura (Universidad Nacional del Nordeste), que comprende un proyecto articulador de docencia e investigación aplicada, orientándose a la formación de recursos humanos desde el grado. Este trabajo se compone de cuatro secciones. En la primera sección se caracteriza la asignatura como contexto en que se desarrollan las actividades. La segunda menciona las líneas de trabajo abordadas en el periodo 2011-2012. En la tercera sección se sintetizan los resultados obtenidos. En la cuarta sección se presentan los productos logrados en el periodo 2011-2012 enfatizando la incorporación, integración y formación de recursos humanos desde el grado.

**Palabras clave:** educación superior, modelos y simulación, formación de recursos humanos, producción de simuladores.

## 1. INTRODUCCIÓN

Se coincide con Martínez Rojo y González (2009) respecto a que Universidad “debe adoptar y adaptar un método de enseñanza válido para el mundo actúa” en

donde los sujetos están familiarizado con el uso intensivo de las herramientas comprendidas por las TIC

La asignatura optativa “Modelos y Simulación”, pertenece al Área Ciencias y Métodos Computacionales del Departamento de Informática de la Facultad de Ciencias Exactas y Naturales y Agrimensura de la Universidad Nacional del Nordeste (FaCENA-UNNE). Corresponde al plan de estudios de la carrera de Licenciatura en Sistemas de Información y se ha caracterizado en trabajos previos como Mariño et al, (2011b), Mariño et al, (2012b).

Los contenidos del programa pertenecen al campo de la Matemática Aplicada. Siguiendo a Gil Chaveznava (2007), es posible afirmar que Modelos y Simulación brinda formación complementaria. Es decir, ofrece conocimientos, habilidades y valores que otorgan al estudiante una visión más amplia de su profesión y del mundo.

El plan de estudios de la carrera describe un conjunto de conocimientos, habilidades y actitudes que definen el perfil profesional de los graduados. Entre éstas se encuentra la habilidad para el manejo sistemas de simulación computarizados, que junto a la capacidad para modelizar, constituyen el objeto de estudio de la asignatura.

A fin de lograr la conexión con el campo profesional y disciplinar, se buscan introducir en el desarrollo de las clases, ejemplos basados en situaciones reales de dominio técnico o académico/científico, para ilustrar a los futuros egresados cómo estos problemas pueden resolverse empleando los temas abordados en la asignatura.

En la Figura 1 se ilustra el número de alumnos inscriptos, regulares y promocionales en los ciclos lectivos 2011-2012 de la asignatura “Modelos y Simulación.

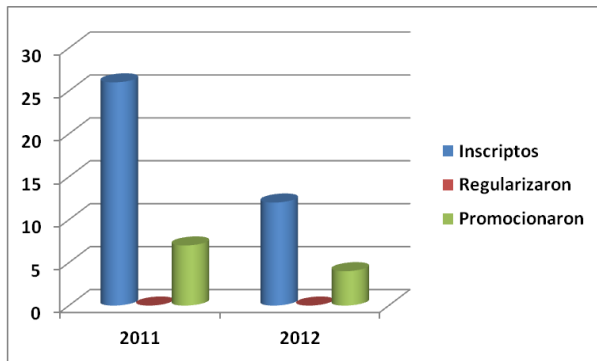


Figura 1. Alumnos inscriptos, regulares y promocionales en las cohortes 2011-2012 de la asignatura “Modelos y Simulación”

Esta asignatura proporciona a los alumnos conocimientos sobre el desarrollo de modelos de tipo matemático, los cuales son utilizados para simular una amplia gama de sistemas reales. Estos conocimientos deben necesariamente ser complementados con los adquiridos en otras asignaturas (lenguajes de programación, paradigmas de desarrollo, técnicas de análisis de sistemas, cálculo de probabilidades y estadística) para poder resolver los Trabajos Prácticos propuestos. Se requiere un razonamiento inteligente por parte de los alumnos para seleccionar aquellos lenguajes y modelos que mejor se adapten a la resolución del problema que se les presenta. Se pretende generar un trabajo original y creativo que propicie en los alumnos la utilización de las distintas herramientas tecnológicas y los conocimientos con los que dispone, a partir de un proceso de aprendizaje que se inició al comenzar la Carrera.

De este modo, se lleva a cabo una integración con otras asignaturas del plan de estudios, logrando así la interconexión entre los contenidos de diversas ciencias, mezclando de manera inteligente los elementos de unas asignaturas con los de otra para el análisis del objeto de estudio, en este caso los modelos de simulación.

En la asignatura se desarrolla un Proyecto orientado a la articulación de actividades de Docencia, Investigación y Extensión, en concordancia con la misiones de la Universidad.

## 2. LINEAS DE INVESTIGACION y DESARROLLO

El grupo ha continuado con la ejecución de las siguientes **líneas generales de investigación**:

1. Rediseño del Entorno Virtual de Enseñanza-Aprendizaje (EVEA) de la asignatura.
2. Construcción de librerías de simulación
3. Implementación de innovaciones pedagógicas.
4. Difusión de resultados.

Estas líneas de trabajo se vinculan con las siguientes actividades:

- Actualización del entorno virtual de enseñanza-aprendizaje (EVEA).

- Selección, estudio y evaluación de herramientas de software libre y propietario para la generación de simuladores educativos y orientados a la modelización de abstracciones de problemas reales.
- Problematización y abstracción de situaciones reales factibles de modelizar y simular.
- Migración a otras herramientas de software, de librerías codificadas en software propietario utilizadas en la asignatura desde el año 1999 como apoyo al desarrollo de clases presenciales.
- Evaluación del desempeño de los algoritmos de simulación estudiados validando diversos lenguajes de programación.

## 3. RESULTADOS OBTENIDOS/ESPERADOS

En esta sección se mencionan los resultados obtenidos en la ejecución del proyecto y los objetivos en curso.

### 3.1. Rediseño del Entorno Virtual de Enseñanza-Aprendizaje (EVEA) de la asignatura.

Los entornos virtuales de enseñanza y aprendizaje (EVEA) son aplicaciones informáticas desarrolladas con fines pedagógicos (Ferreira Szpiniak y Sanz, 2007). Con respecto a la línea 1, se rediseñó el EVEA de la asignatura para incorporar numerosos contenidos y componentes de software generados desde la asignatura, abordando sus cuatros ejes temáticos. Como innovación se señala la selección de los mejores trabajos del seminario integrador, descrito en Mariño et al. (2011a), Mariño et al. (2011b) desarrollados por los alumnos del ciclo anterior, su conversión en algunos casos en publicaciones, y su incorporación al EVEA.

### 3.2. Construcción de librerías de simulación

El desarrollo de herramientas de software educativo para apoyar la educación presencial y su integración al EVEA, como entorno mediático de la modalidad de *b-learning* (Mariño y López, 2007b; Bombelli et al., 2012) implica una constante generación e integración de librerías y simuladores construidos *ad-hoc*. Para el logro de los resultados en esta línea de trabajo se abordó:

**El empleo de herramientas de software libre.** La expansión del software libre inclinó a los integrantes de la cátedra a la selección y estudio de Octave. Se generaron modelos de inventarios como los descritos en Esquivel Forastier (2011) y López et al. (2012).

**El empleo de herramientas de software propietario.** Se emplea principalmente MatLab en la generación de los seminarios integradores. Este lenguaje es uno de los más utilizados en otras asignaturas de la carrera.

**El estudio y selección de lenguajes de programación de propósito específico.**

Previamente se mencionó la construcción de librerías de simulación utilizando lenguajes como MatLab y Octave, se avanza en la elección y estudio de software de simulación como Simulink y Arena. Esta línea se orienta a relevar y aplicar herramientas específicas para proponer futuros estudios comparativos. Algunos trabajos son abordados en el desarrollo de Trabajos Finales de Aplicación (TFA) y con adscriptos.

### **Selección de problemas reales y su modelización**

Con la finalidad de aportar a proyectos de extensión e investigación aplicada, transfiriendo los conocimientos docentes y propios de la técnica de modelado y simulación, se identifican problemas reales que puedan ser resueltos empleando esta técnica. En la mayoría de los casos, son sintetizados para su tratamiento con la metodología de modelado y simulación y la descripta en Mariño y López (2009).

En el año 2011, se estudiaron problemas vinculados a una temporada de fútbol rescatando datos del sitio web de la AFA, y se programó una abstracción del sistema de la represa Itaipú (Barreto et al., 2011). En el año 2012 se desarrollaron un modelo de simulación de colas de espera de una zapatería y modelos de simulación de inventario aplicados a diferentes contextos (productos empleados en una lavandería y bebidas gaseosas comercializadas por un supermercado).

Se continúa avanzando en el desarrollo de simuladores para su procesamiento en paquetes como MatLab, Arena y Simulink.

Cabe aclarar que en las modelizaciones se incentiva la generación de numerosos experimentos, su tabulación, representación gráfica y posterior análisis para elaborar consideraciones pertinentes para apoyar la toma de decisiones y generar un espacio de reflexión relacionando la resolución de problemas que podrían derivarse de prácticas profesionales.

### **3.3. Innovaciones pedagógicas.**

Las producciones derivadas de las actividades de docencia y formación de recursos humanos son compiladas e incorporadas en el EVEA de la cátedra. Se conciben como innovaciones pedagógicas debido a que estas modelizaciones pueden generar estudios en ciclos lectivos posteriores. Además, ilustran el tratamiento de problemas reales, similares a los que podrán enfrentarse en el mundo del trabajo.

### **3.4. Difusión de las actividades y resultados alcanzados.**

Los resultados alcanzados en las distintas etapas de la puesta en marcha del plan de actividades, especialmente las innovaciones pedagógicas propuestas, se difundieron en distintos eventos como congresos o reuniones científicas, reuniones académicas y tecnológicas y revistas, dando a conocer las actividades en el ámbito nacional e internacional. Además, se han publicado artículos donde se detallaron

componentes de software creados abordando los ejes temáticos de la asignatura:

Se mencionan los trabajos descriptos en Barreto et al. (2011), López et al. (2012), Mariño et al. (2012b), Pérez et al. (2012), Primorac et al. (2011a), Primorac et al. (2011b).

A modo de consideraciones finales se menciona que la formación de recursos humanos desde el grado aporta al desarrollo de los futuros profesionales y la vinculación entre lo académico y el mundo laboral. Como perspectiva futura se enfatizará en la utilización y evaluación de herramientas específicas de simulación.

## **4. FORMACION DE RECURSOS HUMANOS**

En la Educación Superior, la formación de recursos humanos es fundamental. En la cátedra, se propicia la incorporación de alumnos plasmándose en el desarrollo de adscripciones y la dirección de TFFA fomentando actividades de docencia, investigación y extensión. A continuación se sintetizan las producciones ejecutadas:

Se plantea un trabajo integrado por docentes-investigadores de la unidad académica, adscriptos y alumnos avanzados de grado, algunos de los cuales se encuentran realizando sus tesinas de fin de carrera en temas que aportan en la línea de los temas abordados en la asignatura.

Se concretó la defensa de una tesina de fin de grado, vinculada con temas de la asignatura (Esquivel Forastier, 2011) denominada “Software Educativo para modelar Sistemas de Colas”.

En el año 2012 se iniciaron dos proyectos de TFA o tesina de la carrera Licenciatura en Sistemas de Información:

Un alumno se encuentra abocado a la búsqueda, selección y estudio de software específico de simulación para su aplicación a resolver un modelos de colas.

Otro alumno está desarrollando una aplicación software para la gestión y control de las existencias de una despensa, basada en un modelo de inventarios y su simulación.

Asimismo, una docente adscripta se encuentra realizando la evaluación y posterior aplicación de la herramienta Arena a problemas de colas.

Cabe aclarar que los productos de TFA y los desarrollados por los adscriptos, siguen los lineamientos y funcionalidades definidas por la cátedra para el EVEA. Esto se fundamenta en su posterior incorporación como complemento del proceso de enseñanza-aprendizaje.

Para propiciar la formación desde el grado en temas de investigación aplicada, se incentiva la lectura de



artículos disciplinares. Además, se fomenta la redacción de artículos derivados del trabajo de seminario integrador, adscripciones y tesinas, entre algunos productos se mencionan Barreto et al. (2011), López et al. (2012), Mariño et al. (2012a), Pérez et al. (2012), Primorac et al. (2011a), Primorac et al. (2011b).

## REFERENCIAS

- Barreto, S. E., Petroff Coloff, N., López, M. V. y Mariño, S. I. (2011): "Modelado y simulación de una represa hidroeléctrica". Anales del XXIV Encuentro Nacional de Docentes en Investigación Operativa (ENDIO) y XXII Escuela de Perfeccionamiento en Investigación Operativa (EPIO): Río Cuarto. Córdoba. Argentina. Mayo de 2011.
- Bombelli, E.; Mella, A., Byorkman, A., Barberis, G. y Cataldi, Z. (2012): "Modelo b-learning para la Enseñanza de la Informática en Ingeniería". Artículos de las Jornadas de Enseñanza de la Ingeniería. Año 2, Volumen 1; 33-38.,
- Esquivel Forastier, S. (2011): "Software Educativo para la modelización y simulación de problemas de colas". Prof. orientadores: Sonia I. Mariño y María V. López. Carrera Licenciatura en Sistemas de Información. FaCENA. UNNE.
- Ferreira Szpiniak, A. y Sanz, C. (2007): "Hacia un modelo de evaluación de entornos virtuales de enseñanza y aprendizaje. La importancia de la usabilidad". Anales del XIII Congreso Argentino de Ciencias de la Computación. CACIC 2007.
- Gil Chaveznava, P. (2007): "Diseño curricular y los diversos modelos educativos". Universidad Autónoma Metropolitana. Unidad Iztapalapa. México D. F. En: [http://cbi.izt.uam.mx/content/eventos\\_divisionales/Seminarios/Seminario\\_Diseño\\_Curricular/Modelo\\_educativo\\_y\\_Plan\\_estudio.pdf](http://cbi.izt.uam.mx/content/eventos_divisionales/Seminarios/Seminario_Diseño_Curricular/Modelo_educativo_y_Plan_estudio.pdf)
- López M. V., Mariño, S. I. y Esquivel Forastier, S. (2012): "Simuladores de problemas de colas". XXV Encuentro Nacional de Investigación Operativa (ENDIO) y XXIII Escuela de Perfeccionamiento en Investigación Operativa (EPIO): Mayo de 2012.
- Mariño, S. I. y López, M. V. (2007b): "Aplicación del modelo b-learning en la asignatura Modelos y Simulación de las carreras de sistemas de la FACENA- UNNE". EDUTEC: Revista Electrónica de Tecnología Educativa. España. Núm. 23. <http://edutec.rediris.es/Revelec2/revelec23/revelec23.html>
- Mariño, S. I. y López, M. V. (2008): "Un proyecto de docencia, extensión e investigación en la asignatura Modelos y Simulación". Anales del X Workshop de Investigadores en Ciencias de la Computación. X WICC. ISBN 978-950-863-863-101-5.
- Mariño, S. I. y López, M. V. (2008b): "Generadores de números aleatorios". Ed. Moglia. ISBN 978-987-05-5025-0.
- Mariño, S. I. y López, M. V. (2009): "Propuesta metodológica para la construcción de software educativo en la asignatura Modelos y Simulación". Anales de XXII ENDIO y XX EPIO.
- Mariño, S. I. y López, M. V. (2010): "Avances del proyecto de docencia, extensión e investigación en la asignatura "Modelos y Simulación"". Anales del XII Workshop de Investigadores en Ciencias de la Computación. XII WICC.
- Mariño, S. I. y López, M. V. (2011): "Experiencias en docencia e investigación en la asignatura "Modelos y Simulación" de la FACENA-UNNE". Anales del XIII Workshop de Investigadores en Ciencias de la Computación.
- Mariño, S. I., López, M. V. y Primorac, C. (2011b): "Los seminarios como una experiencia de integración de conocimientos de modelado y simulación en el ciclo lectivo 2011". Revista IE Comunicaciones N° 14 (2011): julio-diciembre. 37-48, y on-line en: <http://www.adie.es/iecom>
- Mariño, S. I., López, M. V., Longoni, M. G., Velazquez, N. y Roman, A. (2012a): "Simulación y experimentación de modelos de existencias. Un caso orientado a ventas por catálogo". Revista de la Escuela de Perfeccionamiento e Investigación Operativa - EPIO. Edición 33. Aceptado para su publicación.
- Mariño, S. I.; López, M. V. y Alderete, R. Y. (2012b): "Evaluación formativa en la asignatura Modelos y Simulación. Experimentación mediante la generación de software". Revista Iberoamericana de Educación, N° 59/4, de fecha 15/08/12. Publicación editada por la OEI ISSN: 1681-5653. <http://www.rieoei.org/>, Aceptado su publicación
- Mariño, S. I. y López, M. V. (2012c): "Construyendo muestras artificiales con Mathematica". Revista Digital: Matemática, Educación e Internet. Editorial: Instituto Tecnológico de Costa Rica. ISSN 1659-0643 16 págs.
- Martínez Rojo, E. y González Fernández, A. M. (2012): "Renovación, Innovación y TIC en el EEES". REVISTA ICONO 14, 2009, N° 14, pp. 50-63. ISSN 1697-8293. Madrid (España)
- Pérez, C., Mariño, S. I. y López, M. V. (2012): "Desarrollo de generadores de números pseudoaleatorios en Octave". TE&ET. Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología. ISSN: 1850-9959. N° 6. Pgs. 24 a 29. Enero de 2012. <http://teyet-revista.info.unlp.edu.ar/nuevo/files/No6/TEYET6-art03.pdf>
- Pérez, C., (2012): "Software interactivo para el aprendizaje de números pseudoaleatorios y pruebas de hipótesis". Profesor Orientador: Sonia I. Mariño. Carrera Licenciatura en Sistemas de Información. FaCENA. UNNE.

- Primorac, C. R., López M. V. y Mariño, S. I. (2011a):  
"Construcción de una librería de números pseudoaleatorios y muestras artificiales con MatLab". Revista de la Escuela de Perfeccionamiento e Investigación Operativa - EPIO. Edición 32 (digital): 241-258 pp.
- Primorac, C., López, M. V. y Mariño S. I. (2011b):  
"Bondad de ajuste para muestras artificiales de variables aleatorias continuas". Tic Educación y Turismo. Encuentro Internacional BTM 2011. Uruguay. 15 de agosto al 15 de septiembre de 2011.

**Tecnología de Cómputo Ubicua Aplicada a la Educación****Mg. José Luis Filippi<sup>1</sup>, Lic. Guillermo Lafuente<sup>2</sup>, Mg. Carlos Ballesteros<sup>3</sup>, Lic. Gustavo Lafuente<sup>4</sup>, I.S. Daniel Perez<sup>5</sup>, I.S. Sofia Aguirre<sup>6</sup>**LIAU<sup>7</sup> - Facultad de Ingeniería – UNLPam.{filippij<sup>1</sup>, lafuente<sup>2</sup>, balleste<sup>3</sup>, gustavo<sup>4</sup>, perezd<sup>5</sup>, aguirres<sup>6</sup>}@ing.unlpam.edu.ar<sup>7</sup>Laboratorio de Investigación de Ambientes Ubicuos**Resumen**

Las instituciones educativas en general y las universidades en especial, constituyen un ambiente rico en información. La presencia de diferentes actores, con la complejidad de estar en permanente movimiento espacial, desempeñando multiplicidad de actividades, y la necesidad de contar permanentemente con información actualizada; demandan un nuevo ámbito adecuado a éstas características.

El objetivo de éste proyecto busca mejorar los diferentes procesos que se efectúan en las instituciones educativas mediante la aplicación de las tecnologías de la información y la comunicación. En particular, se enfocará la exploración y explotación de las tecnologías de la computación ubicua [1].

Con el fin de explorar las posibilidades que nos brinda el cómputo móvil en el contexto educativo, se efectúa el trabajo experimental en la Facultad de Ingeniería, UNLPam, contando con un escenario de aplicación real.

Este proyecto trata de determinar la utilidad real de los sistemas ubicuos con miras a establecer una implementación a gran escala.

El alcance del proyecto puede medirse en varias dimensiones, formación de recursos humanos de alto nivel para el uso de las diferentes herramientas implementadas, establecimiento de colaboraciones interdisciplinarias e interinstitucionales, desarrollo de herramientas y sistemas para su aplicación en las instituciones educativas y publicación de artículos.

Palabras claves: Aprendizaje Ubicuo. Dispositivo Móvil. Aprendizaje Colaborativo.

**Contexto**

Tipo de Investigación: Aplicada

Campo de Aplicación Principal: 7 1802 Computación, 7 1803 Comunicaciones.

Campos de Aplicación posibles: 13 1040 Ciencia y Tecnología, 7 4399 Otras – Educación – Tecnología Aplicada a la Educación

Institución que Coordina el Proyecto: Facultad de Ingeniería, Universidad Nacional de La Pampa.

**1. Introducción**

En ambientes con necesidades de representación y manejo de conocimiento como lo son los ambientes educativos, existen múltiples necesidades, desarrollo de contenido bibliográfico digital, implementación de programas que permitan la comunicación a través de variados dispositivos tecnológicos móviles entre los diferentes actores de la comunidad educativa, almacenamiento y distribución de material educativo en variados formatos y a través de diferentes plataformas, atención permanente a los actores involucrados dando respuesta a diferentes necesidades, de padres, alumnos, docentes, administrativos y entidades del medio que desean de una u otra forma mantener un contacto fluido con la institución educadora.

Estas relaciones constituyen en sí conocimiento agregado de suma utilidad, rebrota la inquietud de Vannaver Bush (considerado el precursor de la Hipermedia), escrita en “*As we may think*”, donde establecía las dificultades para almacenar y consultar eficientemente la gran cantidad de conocimiento acumulado. La clave para que el material pueda ser consultado eficientemente está en la posibilidad de contar con las relaciones apropiadas que conecten dicho material.

Con la aparición de las redes de computadoras y la conectividad que nos ofrecen los dispositivos móviles, se conforma un ambiente propicio para satisfacer las necesidades de interacción entre los actores mencionados. Mark Weiser[4] menciona que vamos camino a ambientes ubicuos, ambientes poblados de numerosos sensores que gracias a la miniaturización de los dispositivos son invisibles al usuario y están en permanente rastreo de la actividad humana. Y aquí radica nuestro objetivo primario, indagar diferentes herramientas tecnológicas a través de las cuales se pueda ofrecer servicios que satisfagan las necesidades de los usuarios sin intervención humana, características principal de la computación ubicua.

#### Situación Actual del Problema.

Si miramos a nuestro alrededor podemos observar que existe una gran cantidad de componentes electrónicos que están integrados a nuestras vidas, gracias a su reducido tamaño y a su capacidad de cómputo, por ejemplo, computadoras personales, portátiles, teléfonos móviles, PDA, IPod, etc.[2], que logran comunicarse entre sí gracias a la existencia de protocolos de comunicación inalámbrica; a través de redes de celulares, redes de área local (LAN),

redes de área extensa (WAN), y Bluetooth entre otras[3].

Estamos ante la presencia de nuevos dispositivos de comunicación, lo que conlleva un nuevo escenario social y educativo donde los mismos actores, deben redefinir su accionar acorde a las nuevas herramientas tecnológicas que se presentan.

Un nuevo escenario social y educativo se define acorde a los cambios en la nueva organización del conocimiento, que inciden directamente en los procesos mentales. Está cambiando, nuestra forma de conocer, de interactuar con los demás, el volumen de información, su accesibilidad, los nuevos medios de almacenamiento masivo de información, nuevas formas de comunicarse a través de diferentes redes entre personas que se encuentran distantes físicamente.

El concepto de ubicuidad se refiere en general a la presencia de una entidad en todas partes; pero en la computación adquiere la característica de ser, además, invisible. Este paradigma pretende brindar sistemas de cómputo inteligentes que se adapten al usuario, y cuyas interfaces permitan que éste realice un uso intuitivo del sistema. De allí que la meta, de la computación ubicua, de integrar varias computadoras (dispositivos) al entorno físico busca habilitar los beneficios de éstas y de la información digitalizada en todo momento y en todas partes.

La influencia de la computación ubicua es en gran parte de carácter teórico, y dadas las condiciones de integración entre educación y sistemas de cómputo explicada antes, comparte créditos junto a otros conceptos como usabilidad, interactividad e hipertextualidad[5]. Una posible aplicación en educación a distancia, sería que el docente presentara el conocimiento en varias formas que se diferencien en sus características técnicas y didácticas. Así el alumno ejercerá su poder de decisión para optar por una vía



conveniente de aprehensión del conocimiento, con el consecuente mejoramiento de la calidad en el aprendizaje.

Didácticamente, esto implicaría el diseño de distintos materiales o herramientas, teniendo en cuenta los resultados de los antecedentes educativos del alumno. De esta manera se podrán ofrecer cambios significativos en el aprendizaje al conocer los modos de enseñanza que prefiere el estudiante. En otras palabras, al iniciar un curso por Internet las posibilidades didácticas se harán más personales y adecuadas en la medida en que se conozca, a través de sus decisiones anteriores, a ese alumno distanciado físicamente. Para él será imperceptible el hecho de que se reduzcan las opciones gracias a que se le está ofreciendo lo que su atención requiere.

Otro aspecto que tiene relación con el concepto de ubicuidad se encuentra en la ciencia cognitiva. En este sentido las teorías de la percepción retomadas por Norman (1998), nos hablan de la existencia de una percepción periférica donde los objetos solamente son apreciables cuando se centran en nuestra atención. Es decir, que un objeto permanece en un estado tácito hasta que despierta nuestra atención, o es requerido por nuestro interés; de lo contrario seguirá en estado latente.

En los sitios educativos se utilizan lenguajes de programación que posibilitan la “invisibilidad” de los elementos no relevantes en las acciones que ejecuta el usuario en la interfaz. Algunos de estos lenguajes como XML[6], permiten que el receptor de la información pueda acceder a ésta desde cualquier navegador; ya sea el de un asistente personal (PDA) o el de su computadora de escritorio. Todo esto sucede sin percatarnos de los cambios en el diseño, que son necesarios para presentar los datos. Otros lenguajes como Java permiten que un menú

de una página web contenga submenús; visibles únicamente cuando el usuario coloca el cursor encima de un rubro, y desaparecen instantáneamente al abandonar esta zona sensible. A modo de ejemplo y bajo el mismo concepto, los futuros buscadores de contenidos que sigan los lineamientos definidos por la Web Semántica[7] permitirán arribar a contenidos con mayor precisión para el usuario, que los buscadores tradicionales actuales; actual desafío de la llamada Web 3.0.

En el diseño de materiales didácticos para la web, serán también importantes otras características técnicas como por ejemplo la diferencia de peso entre archivos, la facilidad de lectura de un formato, el ancho de banda de la red y sus posibilidades multimediales, etc. [8] Este conjunto de opciones técnicas que permiten la ubicuidad de un sistema de cómputo y las posibilidades conceptuales de diseñar la información, pueden considerarse diseño ubicuo. Trascendiendo teóricamente de los sistemas computacionales a la concepción de aplicaciones didácticas en sitios educativos, el diseño ubicuo nos brinda opciones para mejorar el desempeño de Internet y la Web como medios de comunicación y educación [9].

A partir de establecer una relación estrecha entre el diseño de interfaz y el diseño didáctico, bajo el enfoque de la computación ubicua, ambos ofrecerán al usuario o alumno el conocimiento que requiere sin realizar un gasto cognitivo importante en el medio que transmite [10]. Una integración recíproca entre la educación a distancia y el diseño ubicuo, además de potenciar el uso de Internet en procesos de enseñanza – aprendizaje, contribuirá a la difusión y aplicación de las tendencias tecnológicas que facilitan su uso. Estas perspectivas abren nuevos temas de debate, y al mismo tiempo nuevos puntos de vista en la búsqueda de soluciones a los

problemas que enfrenta la educación en la sociedad.

En éste contexto la escuela debe cambiar, utilizando las nuevas herramientas tecnológicas e instalarlas al servicio de los diferentes actores, generando así un espacio de acción profesional distinto al actual. El entorno educativo cambia [11,3], la educación se presenta como la formación de los educandos en competencias, destrezas, habilidades para desempeñarse en un nuevo espacio social, el digital [12]. Nuevos escenarios educativos, los mismos actores con un nuevo rol, y la implantación de las tecnologías actuales es nuestro desafío actual [13].

## 2. Línea de Investigación y Desarrollo

### Primer Etapa:

- Entrevistar los diferentes actores, directivos, docentes, alumnos, personal administrativo, en su ámbito laboral cotidiano.
- Investigar los requerimientos básicos de TICs para incorporar al proyecto.
- Indagar trabajos de terceros sobre el potencial de la computación ubicua en el ámbito educativo.
- Analizar diferentes herramientas tecnológicas de computación ubicua en ambientes educativos.
- Definir una metodología de desarrollo de trabajo a efectuar.

### Segunda y Tercera Etapa:

- Estudio de la tecnología necesaria para el desarrollo de aplicaciones de cómputo ubicuo.
- Definir prioridades según necesidades para el uso y desarrollo de herramientas de cómputo ubicuo aplicado a la educación.
- Confeccionar herramientas de cómputo ubicuo de interés educativo.

- Probar las aplicaciones desarrolladas en el ámbito educativo adecuado y realizar las pruebas necesarias para la puesta a punto.

### Cuarta Etapa:

- Implantar las herramientas que hayan alcanzado un funcionamiento aceptable.
- Registrar el trabajo desarrollado en sus diferentes etapas.
- Difundir los logros alcanzados a la comunidad entera, para su adopción definitiva en toda la comunidad educativa.
- Presentación de artículos en diferentes congresos de carácter nacional e internacional.
- Difusión y capacitación a instituciones interesadas en aplicar la tecnología presentada.

## 3. Resultados Obtenidos/Esperados

El proyecto se encuentra en las tareas enmarcadas en el punto 4, además se continúa en el desarrollo de nuevos servicios web, que se irán incorporando a la plataforma. El instrumento es accesible a través de la Web, se encuentra conformado hasta la fecha por los siguientes sistemas: [14]

- Sala de trabajo colaborativo
- Repositorio de materiales
- Tablón de anuncios
- Foro de noticias
- Mensajería automática
- Configuración
- Autoevaluación

que suministran servicios a alumnos y docentes de las diferentes carreras de la Facultad de Ingeniería de la Universidad Nacional de La Pampa. Próximamente se desarrollará el modulo para brindar servicios a padres y/o tutores a través del cual podrán realizar un seguimiento sobre la actividad académica desempeñada por sus hijos.

Al concluir el proyecto se espera contribuir en el desarrollo de una plataforma de educación ubicua (u-education), que se implante en diferentes unidades académicas de nuestra Universidad.

## 4. Formación de Recursos Humanos

Director de Proyecto

Co-Director de Proyecto

3 Investigadores

1 Asistente de Investigador

1 Personal de Apoyo

## 5. Bibliografía

[1] Weiser M. (1991), *The Computer for the Twenty-First Century*.

[2] Vedar E, Evans W, Griswold, W, (2009) *Ubibot - Prototyping Infrastructure for Mobile Context-Aware Computing*. Ubicomp 2009.

[3] Carmona M., González S., Castro Ruiz, *Innovación Tecnológica en Comunicaciones Móviles Desarrollada Con Software Libre: Campus Ubicuo* Novática: Revista de la Asociación de Técnicos de Informática, ISSN 0211-2124, N°. 190, 2007

[4] Weiser, M. (1993), *Ubiquitous Computing*, IEEE Computer "Hot Topics", October 1993, <http://www.ubiq.com/hypertext/weiser/UbiCompHotTopics.html>

[5] Nielsen J. (2000), "Designing Web Usability: The Practice of Simplicity", New Riders Publishing, Indianapolis, ISBN 1-56205-810-X

[6] W3C, Extensible Markup Language (XML), <http://www.w3.org/XML/>

[7] W3C, W3C Semantic Web Activity, <http://www.w3.org/2001/sw/>

[8] Girouard J, Horn H, Solovey S, Zibgelbaum, (2008), *Reality Based Interaction, a framework for post-WINP Interfaces*.

[9] Krumm J., (2009), *Creating a New Book of Tutorials: Ubiquitous Computing Fundamentals*, Microsoft Research Redmond, Ubicomp 2009.

[10] Bravo C., Redondo M., Ortega M., Bravo J., *Evolución de un Entorno Colaborativo de Enseñanza Basado en Escritorio hacia la Computación Ubicua*, (2002) COLINE'02 Investigación en Entornos de Interacción Colectiva, Workshop de Investigación sobre nuevos paradigmas de interacción en entornos colaborativos aplicados a la gestión y difusión del Patrimonio cultural. Granada, 11 y 12 de Noviembre del 2002.

[11] Weiser, M. (1998) *The future of Ubiquitous Computing on Campus*. Communications of ACM, 41-1, January 1998, 41-42.

[12] Malani R., Griswold W, Simon B, (2009) *Public Digital Note-Taking in Lectures*. Ubicomp 2009.

[13] Richards M, Woodthorpe J, (2009), *Introducing TUI00 "My Digital Life": Ubiquitous computing in a distance learning environment*. Ubicomp 2009.

# Nuevas Formas de Construcción del Conocimiento: Redes Sociales

Gustavo López, Arturo Servetto, Adriana Echeverría, Patricia Calvo, Ismael Jelder, Osvaldo Clúa, María Feldgen, Mariano Méndez, María Delia Grossi, Elizabeth Jiménez Rey, Alejandra Arriazu, Elena García

Departamento de Computación  
Facultad de Ingeniería – Universidad de Buenos Aires

[glopez@fi.uba.ar](mailto:glopez@fi.uba.ar), [aserve@gmail.com](mailto:aserve@gmail.com), [aechevi@fi.uba.ar](mailto:aechevi@fi.uba.ar), [pat\\_calvo@yahoo.com](mailto:pat_calvo@yahoo.com), [jder@fi.uba.ar](mailto:jder@fi.uba.ar), [oclua@ieee.org](mailto:oclua@ieee.org), [mfeldgen@ieee.org](mailto:mfeldgen@ieee.org), [mdg7501@yahoo.com.ar](mailto:mdg7501@yahoo.com.ar), [marianomendez@gmail.com](mailto:marianomendez@gmail.com), [ejimenezrey@yahoo.com.ar](mailto:ejimenezrey@yahoo.com.ar), [alejandra\\_arriazu@yahoo.com.ar](mailto:alejandra_arriazu@yahoo.com.ar), [eleigar@gmail.com](mailto:eleigar@gmail.com)

## RESUMEN

El objetivo de esta línea de investigación es el estudio del aporte que las redes sociales tienen en la generación y comunicación de saberes. Para ello se estudia el uso de distintas aplicaciones y herramientas tecnológicas y su integración con la enseñanza. Así como también el grado de penetración de dichas herramientas entre los alumnos de las distintas carreras de la FIUBA. Abordando posteriormente la implementación de un entorno de enseñanza con las características de red social.

## CONTEXTO

Esta línea de Investigación forma parte de las tareas de investigación del proyecto por Resolución (CS) 3822/2011 “Innovación Pedagógica vía TIC para la Mejora de la Calidad Educativa en la FIUBA”, correspondiente al Programa UBA-TIC “Potenciar la enseñanza en el nivel superior a través de las nuevas tecnologías”, Resolución (CS) N° 2386/11.

**Palabras Claves:** Redes Sociales, WEB, Enseñanza y aprendizaje colaborativos, desarrollo de creatividad.

## 1. INTRODUCCION

En los últimos años las redes sociales han captado una preponderante importancia en distintos aspectos de la vida social y política de las personas. A tal punto que en el año 2011 redes sociales como Twitter y Facebook jugaron un papel importante en los cambios sociales sufridos en ciertos países. Las redes sociales están formadas por comunidades de usuarios que tienen intereses en común e intercambian información generada por ellos mismos. El formato básico de esta información está compuesto principalmente por: imagen, sonido y texto [1].

En los últimos años las horas consumidas en Facebook por los visitantes promedios en argentina es de 10,5 % (Fuente: comScore Media Metrix, Visitantes Edad 15+ Conectándose desde Hogar/Trabajo, Junio 2012). Esta creciente demanda del



uso de redes sociales lleva a plantearse la necesidad de integrar las mismas a la enseñanza [2]. Además, basados en el contexto actual de la FIUBA, en el cual [3]:

1. Se ha podido constatar que la cantidad de alumnos inscriptos en carreras de informática ha tenido una tendencia creciente en los últimos tiempos.
2. El alumno es captado por el mercado laboral tempranamente debido al buen nivel de las remuneraciones ofrecidas.
3. Las exigencias laborales llevan al alumno al incumplimiento de los requerimientos académicos en cuanto a horarios y dedicación.

Y en los efectos que los anteriores producen:

- A. Aumento de la densidad de cursos en ciertas franjas horarias.
- B. Agotamiento de los recursos físicos debidos a las consecuencias del punto anterior.
- C. El tiempo esperado de finalización de las carreras aumenta.

Por ello es necesaria la aplicación de Nuevas Tecnologías de Información y Comunicación para proveer a los alumnos de herramientas adicionales y alternativas en la construcción y comunicación del conocimiento.

En la actualidad, las Redes Sociales Web forman parte de las actividades sociales de los jóvenes. Por ello, para los docentes, su uso les plantea el reconocimiento de nuevas propuestas para la construcción del conocimiento. Propiciando su utilización en dicha construcción un “puente” entre alumno y docente. Por lo tanto, su implantación en la educación se ve favorecida porque no es un obstáculo, sino un soporte conocido y explorado.

En alentar la actividad colaborativa con participación de docentes y alumnos se basa la naturaleza misma del proyecto desde: el análisis de la situación actual, la planificación del proyecto, la ejecución del mismo, el monitoreo, evaluación y uso.

La mayoría de las plataformas utilizadas en la actualidad no permiten la conexión entre las aulas virtuales, de modo tal que se imposibilita la integración horizontal de contenidos, no permiten la organización de cursos de modo que los alumnos accedan a distinto tipo de material según las características de su modo de aprendizaje (en las aulas de Moodle [5] todos los alumnos acceden al mismo material), teniendo además, varias limitaciones importantes el conjunto de actividades sociales que los miembros de la comunidad universitaria pueden realizar por fuera del aula virtual, tales como comunicaciones en diversas direcciones (entre alumnos, entre docentes, entre alumnos y docentes), o acceso a material sin ser alumno de un curso.

## 2. LINEAS DE INVESTIGACION Y DESARROLLO

Se plantean como temas de estudio:

- Caracterización del uso en el estudio de la tecnología en el alumnado de FIUBA: A partir de los datos obtenidos de encuestas realizadas por los alumnos de distintos cursos del Departamento de Computación de la FIUBA se busca determinar el uso de distintas herramientas tecnológicas dentro del ámbito académico y cómo éstas se aplican para el estudio de las distintas materias.
- Caracterización del uso general de la tecnología en el alumnado de FIUBA: A partir de los datos obtenidos de en-

cuestas realizadas por los alumnos de distintos cursos del Departamento de Computación de la FIUBA se intenta determinar el uso de distintas herramientas tecnológicas fuera del ámbito académico.

- Caracterización y estudio de las herramientas tecnológicas asociadas a la enseñanza: Se realizará un estudio comparativo de distintas herramientas tecnológicas utilizadas en la enseñanza para determinar cuáles de estas deberían ser incluidas en el diseño de una red social del Departamento de Computación de la FIUBA.
- Diseño de una Red Social: diseño de interfaces web adecuadas y la adaptación e integración al entorno de objetos de aprendizaje compartidos en repositorios. Así como la obtención de un conjunto de requerimientos que formen una “base-line” para el producto.
- Implementación de una Red Social: Se propone la implementación de una red social Web para uso de docentes y alumnos del Departamento de Computación de la FIUBA que posibilite la organización de cursos y la instrumentación de modalidades de enseñanza y aprendizaje colaborativos y permita incluir actividades de estimulación de la creatividad. También se brindaría acceso a visitantes WEB con posibilidad de participar en foros de la red y de los cursos para intercambio de ideas y realización de consultas. La implementación del proyecto implica la participación en el desarrollo de un entorno de enseñanza y aprendizaje semipresencial colaborativo con características de redes sociales.

### 3. RESULTADOS OBTENIDOS/ESPERADOS

Los resultados esperados son:

- Fomentar el trabajo colaborativo virtual entre los alumnos. Dada la naturaleza del proyecto, la actividad colaborativa con participación de docentes y alumnos está garantizada desde el análisis de la situación actual, la planificación del proyecto, la ejecución del mismo, el monitoreo, evaluación y uso.
- Determinar una caracterización a nivel de uso de la tecnología de los alumnos de las carreras de la FIUBA.
- Transformar las características presenciales del aprendizaje de las materias del Departamento de Computación de la FIUBA.
- Implementar una plataforma didáctica que resulte útil para el dictado de las materias del Departamento de Computación de la FIUBA.
- Lograr la actualización permanente de los contenidos curriculares pertenecientes a ambas carreras hasta sus últimos avances académicos.
- Articular conocimientos teóricos y prácticos con la participación de los alumnos.

Resultados obtenidos:

- Se desarrollaron distintos prototipos: Alguno de estos prototipos a partir de modificaciones al código mismo de MOODLE, y otros a partir de herramientas de desarrollo de software para redes sociales.
- Definición de requerimientos: se definió un conjunto crítico de requeri-

mientos que habrá de cumplir el sistema producto de este proyecto.

- Análisis e implementación de los prototipos: Se implementaron varios prototipos del sistema. De los obtenidos, se observó que aquellos que se desarrollaron a partir de MOODLE no lograron satisfacer cabalmente los requerimientos críticos, mientras que los desarrollados a partir de herramientas para implementar redes sociales se acercaron a los objetivos del proyecto.
- Encuesta de uso de TICs: Se llevó a cabo una encuesta a docentes y a alumnos sobre el uso de TIC en varios de los distintos cursos que se dictan en el departamento de Computación de FIUBA.

#### **4. FORMACION DE RECURSOS HUMANOS**

Desarrollo de un Trabajo Profesional, que consiste en una plataforma de Enseñanza/Aprendizaje, título es "Campus Virtual FIUBA".

#### **5. BIBLIOGRAFIA**

1. Sánchez Carballido, J. R. (2011). Perspectivas de la información en Internet: ciberdemocracia, redes sociales y web semántica. *Zer-Revista de Estudios de Comunicación*, 13(25).
2. Garcia, I., Peña López, I., Johnson, L., Smith, R., Levine, A., & Haywood, K. (2010). Informe Horizon: edición iberoamericana 2010.
3. Gustavo López, et al. (2012). Implementación de una red social para la Construcción y Comunicación de Saberes en FIUBA. Encuentro Patagónico de educación y tecnología.
4. Helsper, E. J., & Eynon, R. (2010). Digital natives: where is the evidence?. *British Educational Research Journal*, 36(3), 503-520.
5. [www.moodle.com](http://www.moodle.com)

# Indicadores para la Evaluación del Trabajo Colaborativo en Línea en el Área de Electrotecnia

Gustavo Bacino<sup>1</sup>, Stella Maris Massa<sup>1</sup>, Alejandra Zangara<sup>2</sup>

<sup>1</sup> Facultad de Ingeniería, Universidad Nacional de Mar del Plata, Argentina

<sup>2</sup> Facultad de Informática, Universidad Nacional de La Plata, Argentina

[gustavobacino@gmail.com](mailto:gustavobacino@gmail.com), [smassa@fi.mdp.edu.ar](mailto:smassa@fi.mdp.edu.ar), [alejandra.zangara@gmail.com](mailto:alejandra.zangara@gmail.com)

## Resumen

En el presente trabajo se exponen los instrumentos elaborados para la evaluación del proceso de formación mediado por tecnología, en el marco de una experiencia desarrollada en la modalidad de Aula Extendida. Se trata de una matriz de valoración o rúbrica, denominada Matriz de Valoración de Contenidos, pensada para cuantificar los resultados obtenidos por los grupos de alumnos en la resolución de problemas y de una tabla que incluye los indicadores definidos para realizar la evaluación de contenidos que permitan calificar el trabajo colaborativo de los estudiantes que participaron de la experiencia.

Los alumnos deben resolver problemas correspondientes a una unidad temática de la asignatura Electrotecnia 2, perteneciente al Plan de Estudios de las carreras de Ingeniería Eléctrica y Electromecánica, aplicando la metodología de Aprendizaje Basado en Problemas, divididos en pequeños grupos y trabajando en forma colaborativa directamente sobre un Entorno Virtual.

Mediante ambos instrumentos, los estudiantes involucrados cuentan con información previa acerca de cuáles son los criterios de evaluación, permitiendo de este modo lograr una calificación más transparente y uniforme.

**Palabras clave:** Matriz de Valoración; Aprendizaje Basado en Problemas (ABP); Aprendizaje

**colaborativo; Enseñanza en pequeños grupos; Electrotecnia.**

## Contexto

La línea de investigación presentada está inserta en el proyecto denominado "Ambientes virtuales de aprendizaje para la enseñanza de la ingeniería", planificado sobre una duración bianual y que se enmarca en el área de conocimiento interdisciplinaria: Nuevas Tecnologías de la Información y Comunicación (NTIC's) aplicadas a la educación.

La institución a la que pertenece es la Universidad Nacional de Mar del Plata, y sus integrantes son miembros de la Facultad de Ingeniería y del Centro de Investigación de Procesos Básicos, Metodología y Educación (CIMEPB) de la Facultad de Psicología.

Constituye asimismo un avance de una tesis de posgrado en etapa de elaboración, para acceder a la Maestría en Tecnología Informática Aplicada en Educación, dictada por la Facultad de Informática de la Universidad Nacional de La Plata.

## Introducción

Con anterioridad [1] se han descrito las estrategias principales que integran la propuesta: Aprendizaje Basado en Problemas (ABP); Aprendizaje Colaborativo y Enseñanza en Pequeños Grupos.

La modalidad de aula extendida o "extended learning" adoptada, obedece al hecho de que la unidad temática objeto de la experiencia, pertenece a



una asignatura de grado, Electrotecnia 2, del currículum de las Carreras de Ingeniería Eléctrica y Electromecánica, de la Facultad de Ingeniería de la Universidad Nacional de Mar del Plata. Esta característica no permite el reemplazo, aunque sea parcial, de la modalidad presencial. La forma adoptada, en tanto, permite transformar el aula tradicional en un “aula extendida”, ampliando de esa manera sus alcances y brindándole al alumno la posibilidad de complementar su aprendizaje con el desarrollo de actividades en forma virtual, sin que resulte modificado el carácter presencial de la asignatura.

Como metodología de enseñanza y aprendizaje, se ha empleado tradicionalmente, la combinación de encuentros de transferencia de conocimientos teóricos con clases de resolución de problemas y prácticas de laboratorio, todas ellas en la modalidad presencial.

A partir de la utilización de un Entorno Virtual de Enseñanza y Aprendizaje (EVEA), basado en la plataforma Moodle, desarrollado en la facultad para el uso de las cátedras, se han comenzado a realizar actividades de enseñanza y aprendizaje mediadas por tecnología.

En particular, en la experiencia que aquí se trata, dichas actividades han consistido, como se ha mencionado con anterioridad, en la resolución de problemas correspondientes a una de las unidades temáticas de la asignatura, mediante la metodología del ABP, con los alumnos desarrollando la actividad en forma colaborativa, reunidos en pequeños grupos y realizando los encuentros en forma virtual, principalmente a través de los foros.

El proceso de aprendizaje convencional se invierte al trabajar en el ABP. Por lo general, en primer lugar se expone la información, procediéndose a continuación a realizar su aplicación en

la resolución de un problema. En cambio, en el caso del ABP primero se presenta el problema, luego se identifican las necesidades de aprendizaje, se continúa con la búsqueda de la información necesaria y finalmente se regresa al problema.

Son los estudiantes quienes toman la iniciativa para resolver los problemas, lo que permite afirmar que se está frente a una técnica donde ni el contenido ni el profesor son elementos centrales [2] y que incluye, además, una transformación cultural [3].

La necesidad y conveniencia de evaluar el trabajo colaborativo de los estudiantes, divididos en pequeños grupos, es que precisamente esa es una de las principales características que distinguen al ABP.

En cuanto a la colaboración, como concepto de enseñanza, utiliza la interacción social como medio de construcción del conocimiento.

Colaborar es trabajar con otro u otros con el fin de alcanzar los objetivos comunes de aprendizaje. La mayor parte de la responsabilidad de aprender está centrada en los estudiantes y el docente cumple el rol de facilitador [4] - [5].

La utilización de una metodología como el ABP, determina que las herramientas de evaluación que se utilicen no sólo tienen que medir el aprendizaje, deben además conformar un instrumento integrante del proceso de aprendizaje de los alumnos, de modo que ellos puedan percibir cómo se los va a evaluar.

Se separa así el proceso de evaluación en dos etapas, la primera de ellas permite, mediante una Matriz de Valoración de Contenidos (Anexo 1), cuantificar los resultados obtenidos por los grupos de alumnos en la resolución de problemas, en la restante, concebida para la evaluación del trabajo colaborativo, se recurre a un conjunto de indicadores de discurso (Anexo 2), destinados a determinar las habilidades

de trabajo colaborativo de los estudiantes que participan de la experiencia.

Para la generación de indicadores, que permitan describir el trabajo en equipo, basándose en tres dimensiones o mecanismos interpsicológicos: Interdependencia positiva, Construcción de significado y Relaciones psico-sociales, se tomó como modelo el trabajo de Casanova Uribe [6].

Estos indicadores permiten determinar la presencia o ausencia de los mecanismos citados y, a su vez, valorar la existencia o no de aprendizaje colaborativo en el proceso de trabajo grupal [7], posibilitando evaluar la contribución de cada alumno para con el grupo y sus interacciones personales con los demás integrantes del mismo.

Mediante todo el proceso de evaluación descrito en los párrafos anteriores se llega a una calificación final para cada estudiante, que contempla tanto la calidad del trabajo realizado, como su aporte individual a la tarea desarrollada por el grupo al cual pertenece.

### Líneas de Investigación

▲ Aprendizaje Basado en Problemas en un Aula Extendida.

▲ Trabajo colaborativo en pequeños grupos: extensión de sus límites con el auxilio de las TIC's.

La descripción completa del problema de investigación puede leerse en [8].

### Formación de recursos humanos

Se encuentran en desarrollo tres tesis de posgrado de la UNLP en el marco del proyecto de investigación correspondientes a la Maestría en Tecnología Informática Aplicada en Educación y una del Doctorado en Ciencias Informáticas.

Se han realizado numerosas actividades de transferencia: gestión y asesoramiento en el uso de la Plataforma Educativa Moodle de la Facultad de Ingeniería de la UNMDP.

### Referencias Bibliográficas

- [1] Bacino, G., Massa, S. M., & Zangara, A. (2012). "El empleo de una herramienta colaborativa en un entorno Moodle para Aprendizaje Basado en Problemas". XIV Workshop de Investigadores en Ciencias de la Computación. Posadas, Misiones.
- [2] Santillán Campos, Franciso. 2006. "El Aprendizaje Basado en Problemas como propuesta educativa para las disciplinas económicas y sociales apoyadas en el B-Learning". Revista Iberoamericana de Educación 40(2).
- [3] McAlpine, Iain. 2000. "Collaborative learning online". Distance Education 21(1):66-80. Retrieved June 25, 2011 <http://www.informaworld.com>.
- [4] Cenich, Gabriela, and G. Santos. 2006. "Aprendizaje Colaborativo Online: Indagación de las Estrategias de Funcionamiento". Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología (TE&ET) 1(1):79-86. Acceso 25 de junio, 2011. <http://teyet-revista.info.unlp.edu.ar>.
- [5] Barkley, Elizabeth F., K. Patricia Cross, and Claire Howell Major. 2005. "Collaborative learning techniques". edited by John Wiley & Sons Inc.
- [6] Casanova Uribe, M. O., Alvarez Valdivia, I. M., & Gómez Alemany, I. (2009). "Propuesta de Indicadores para Evaluar y Promover el Aprendizaje Cooperativo en un Debate Virtual". Eductec. Revista Electrónica de Tecnología Educativa, 28, 1-18.
- [7] Casanova Uribe, M. O. (2008) "Aprendizaje Cooperativo en un Contexto Virtual Universitario de Comunicación Asincrónica: Un estudio sobre el proceso de interacción entre iguales a través del análisis del discurso". Tesis Doctoral, Universidad Autónoma de Barcelona.
- [8] Bacino, Gustavo, Stella Maris Massa, and Alejandra Zangara. 2011. "Propuesta de Aula Extendida en la educación superior en ingeniería. Aplicación en el área tecnológica básica de Electrotecnia". 9th Latin-American Congress on Electricity, Generation and Transmission - CLAGTEE 2011.

## Anexo 1

## MATRIZ DE VALORACIÓN DE CONTENIDOS

MATRIZ DE VALORACIÓN DE CONTENIDOS					
CRITERIOS	CLASIFICACIONES				
	(100) excelente	(80) muy bueno	60 (bueno)	40 (regular/pobre)	0 (insatisfactorio)
Comprender el problema	Diferencia todos los datos e incógnitas de manera correcta y las restricciones si las hay.	Diferencia todos los datos y la mayoría de las incógnitas y las restricciones si las hay.	Diferencia todos los datos y algunas incógnitas.	Sólo diferencia los datos.	No se ocupa de distinguir datos e incógnitas al principio.
Elaborar un plan	Las ecuaciones planteadas responden exactamente a las necesidades del problema.	Las ecuaciones son planteadas de manera correcta.	En general las ecuaciones son planteadas correctamente.	Se excede en el número de ecuaciones planteadas, algunas son redundantes.	Realiza un planteo equivocado de las ecuaciones.
Ejecutar el plan	Son correctos en su totalidad incluyendo alternativas donde resulta posible. Incluye las unidades físicas asociadas cuando y donde corresponde, respetando las normas ortográficas para los símbolos.	Se presentan soluciones correctas y ordenadas. En general la inclusión de las unidades físicas es correcta así como el respeto por las normas ortográficas para los símbolos.	En general los resultados presentados son correctos aunque se han deslizado errores de cálculo. En algunos casos las unidades físicas no se han incluido correctamente ó no siempre respeta las normas ortográficas.	Algunos errores de cálculo resultan determinantes. En algunos casos las unidades físicas no se han incluido correctamente ó no siempre respeta las normas ortográficas.	Errores producto de un proceder errado o de excesivos errores de cálculo. No demuestra mayor respeto por la inclusión de las unidades físicas, ni por sus normas ortográficas.
Analizar la solución obtenida	Los conceptos físicos asociados son correctos y bien fundamentados. Extrae conclusiones basado en resultados manifiestos.	En general los conceptos físicos involucrados son correctos. Se analizan los aspectos más importantes.	En general los conceptos físicos asociados son correctos aunque se presentan con poca claridad. En general se analizan los aspectos más importantes.	Algunos conceptos físicos son incorrectos o se presentan con poca claridad. El análisis es superficial.	Conceptos físicos difusos o incorrectos. Análisis pobre o nulo.
Informe de presentación	Excelente, se cuidaron hasta los menores detalles. No se detectan errores en aspectos formales (redacción y ortografía)	Formato y presentación bien cuidados. Se han deslizado pocos errores formales.	Formato y presentación bien cuidados en general. Se detectan varios errores formales.	Se descuidó en general este aspecto. El número de errores es relativamente importante	Desprolijo, poco cuidado. Mala redacción y/u ortografía.
<b>Observación:</b> Ninguno de los cinco criterios puede terminar con una clasificación insatisfactoria. Calificación final (máximo 100)					

## Anexo 2

## INDICADORES PARA LA EVALUACIÓN DEL TRABAJO COLABORATIVO

INDICADORES PARA LA EVALUACIÓN DEL TRABAJO COLABORATIVO		
Dimensiones: Mecanismos Interpsicológicos	Indicadores	Definición
Interdependencia Positiva	Contribuye	Ayuda y concurre con el resto del grupo al logro del objetivo.
	Propone	Hace una propuesta tendiente a organizar el trabajo conjunto.
	Solicita	Requiere, gestiona, pide algo de los otros miembros del grupo.
	Consulta	Pide opinión, información, asesoramiento a los integrantes del grupo.
	Aclara	Hace perceptible, manifiesto, inteligible algo para los demás.
Construcción de Significado	Analiza	Examina, compara, detalla resultados obtenidos.
	Simplifica	Facilita, resume, sintetiza algo haciéndolo menos complicado.
	Opina	Juzga, valora acerca de la verdad de algo.
	Demuestra	Prueba, justifica, evidencia una idea o contenido.
	Disiente	Discrepa, no se muestra de acuerdo con el parecer de otro.
	Explica	Aclara, interpreta, desarrolla aspectos de la resolución de problemas.
	Presenta (resultados)	Muestra resultados parciales o totales de la resolución de algún problema.
	Explicita	Manifiesta, expresa la tarea para asegurar su representación compartida.
Relaciones psicosociales	Reconoce	Aprueba, admite, acepta la opinión o aporte de otro.
	Alienta	Anima, estimula, motiva a alguien en su esfuerzo.
	Sociabiliza	Incluye todo tipo de saludos y comentarios o frases humorísticas.
	Comunicación abierta	Expresiones relacionadas con circunstancias personales, afecto y emociones.



## Creación de libros abiertos en América latina: metodología, plataforma y estrategias

Claudia Deco<sup>1</sup>, Cristina Bender<sup>1</sup>, Ana Casali<sup>1,2</sup>, Raúl Kantor<sup>1</sup>

<sup>1</sup> Facultad de Ciencias Exactas, Ingeniería y Agrimensura,  
Universidad Nacional de Rosario, Rosario, Argentina.  
{ deco, bender, acasali, kantor }@fceia.unr.edu.ar

<sup>2</sup> Centro Internacional Franco Argentino de Ciencias de la Información y de Sistemas CIFASIS

### Resumen

En esta línea de investigación se propone el desarrollo de libros de texto abiertos y colaborativos en Latinoamérica. Para esto se trabaja en: el diseño de una metodología colaborativa, la implementación de una plataforma tecnológica adecuada y una estrategia para la generación y adopción de estos libros accesibles vía Web. El impacto esperado es aumentar la accesibilidad de estudiantes de bajos recursos a libros abiertos y así colaborar para disminuir la deserción en la educación superior debido a razones económicas.

**Palabras Claves:** Libros de texto abiertos, metodología colaborativa, plataforma tecnológica.

### Contexto

Esta línea de I+D se está llevando a cabo a través de varios proyectos de investigación.

En el marco del Programa ALFA III de cooperación entre Instituciones de Educación Superior (IES) de la Unión Europea y América Latina, se lleva adelante el proyecto trianual LATIn (Latin American open Textbook Initiative, DCI-ALA/19.09.01/11/21526/279-155/ALFA III (2011)-52), del que forman parte nueve universidades latinoamericanas (entre las cuales se encuentra la UNR) y tres europeas. Este proyecto tiene como objetivo establecer una Iniciativa para la creación colaborativa de libros

de texto abiertos para la educación superior en América Latina. Se propone alentar a docentes de la región a trabajar colaborativamente. En este proyecto se trabaja en conjunto con investigadores de otras universidades latinoamericanas en el marco de la Comunidad Latinoamericana de Objetos de Aprendizaje (LACLO).

Los PIDs de la UNR involucrados son: Búsqueda personalizada de información en Bases de Datos [ING348] dirigido por Claudia Deco (2011-2014); y Sistemas multiagentes y sus aplicaciones en la educación [ING308] dirigido por Ana Casali (2010-2013).

Vinculada a la temática se trabaja también en el proyecto PICTO-2010-0143: "Hacia el desarrollo y utilización de Repositorios de Acceso Abierto para Objetos Digitales Educativos en el contexto de las universidades públicas de la región centro-este de Argentina", dirigido por Patricia San Martín y Ana Casali. Este es un Proyecto de Investigación Científica y Tecnológica Orientados del Fondo Nacional para la Investigación Científica y Tecnológica (FONCYT).

### Introducción

La creación colaborativa de libros de texto abiertos para la educación superior en América Latina apunta a la generación de libros multimedia disponibles libremente en formato electrónico. Su estructura modular posibilita que sean reutilizados permitiendo la creación de libros de texto personalizados. Esto permitiría aportar a la solución de uno de los problemas que limitan la

continuidad de los estudios superiores, que es el alto costo de los libros de texto.

Las bibliotecas universitarias contribuyen a la solución en forma parcial al proveer ejemplares de libros de forma gratuita. Otro camino utilizado es el empleo de fotocopias, pero éstas, además de ser en muchos casos ilegales, en general degradan la calidad de imágenes, gráficos, etc. El uso de apuntes de clase puede ser insuficiente, dado que en general son acotados o incompletos en el desarrollo de los temas. Así se crea una diferencia de posibilidades entre los estudiantes que pueden acceder al libro de texto y los que no. Adicionalmente, la mayoría de los libros de texto son creados fuera de la región y/o por autores no latinoamericanos debido a la dificultad que los autores locales suelen tener para publicar y distribuir sus libros. Una solución para mejorar el acceso a los libros de texto, preservando la libertad académica de cada profesor, ha surgido en el ámbito de las Tecnologías de Aprendizaje, a partir del concepto de recurso educativo abierto los cuales pueden ser libremente copiados, modificados, compartidos, impresos y distribuidos, y han llevado a la creación del concepto de libros de texto abiertos. Se plantea tener una visión amplia de los libros electrónicos, como la que presenta Faquet [1], que los define como documentos virtuales compuestos de fragmentos que se pueden ensamblar para constituir documentos reales que pueden ser leídos o impresos, o hipertextos que pueden leerse mediante la navegación. Estos libros, además, pueden tener contenido multimedia y ser actualizados frecuentemente. Existen varias propuestas de libros abiertos tales como: Wikilibros ([wikibooks.org](http://wikibooks.org)), Connexions ([cnx.org](http://cnx.org)) y el Proyecto California Open Source Textbooks ([opensourcetext.org](http://opensourcetext.org)) entre otras. Aunque en estas propuestas se incluyen secciones de material en español o portugués, la participación de autores latinoamericanos es bastante limitada.

En el marco de la Iniciativa LATIn se propone trabajar en Latinoamérica, en la creación y diseminación de Libros de Texto Abiertos para educación superior. Los libros así creados estarán disponibles en un formato electrónico y no deberán pagarse licencias o derechos para su distribución. Además, la Iniciativa promoverá que profesores de diferentes instituciones de la región trabajen colaborativamente en la creación de estos libros. Para lograr el objetivo, se trabaja en tres ejes principales fuertemente interrelacionados: 1) Diseñar una metodología para la creación

colaborativa de libros de texto abiertos, 2) Diseñar, seleccionar y/o adaptar una plataforma tecnológica para dar soporte a la creación colaborativa, adaptación, combinación y reutilización de estos libros, y 3) Definir un conjunto de estrategias para la implementación y la adopción de esta iniciativa en las instituciones de educación superior de la región. Esto es, atender las cuestiones pedagógicas, tecnológicas y políticas necesarias para que este tipo de iniciativa tenga éxito en países de América Latina.

## Líneas de investigación y desarrollo

Para alcanzar los objetivos que proporcionen una ayuda a los problemas mencionados en la introducción, se han planteado distintas líneas de investigación las cuales se interrelacionan.

En el proyecto LATIn, se trabaja en los tres ejes mencionados en la Introducción. Una descripción detallada de cada uno se presenta a continuación.

1) Diseño de una metodología para la creación colaborativa de libros

Respecto a este eje, se trabaja en el diseño de una metodología que guíe el proceso de creación colaborativa de contenidos, en la que profesores y autores están involucrados. Los principales objetivos son garantizar la calidad de los materiales, facilitar el proceso de colaboración y promover la reutilización de componentes individuales. En este sentido, se intenta determinar el modo de colaboración en la autoría de libros, las herramientas a utilizar y el tipo de licencia para los autores que reutilicen el material modificándolo. Se establece la forma en que estos libros pueden ser personalizados permitiendo la selección de módulos, compaginado y reeditado de sus partes, considerando las características culturales de los diferentes países involucrados, y en particular, cómo se realizarán las traducciones y adaptaciones necesarias. No existe una metodología a medida que sea útil y aplicable en cualquier caso y situación. Cualquier propuesta para una iniciativa de este tipo debe adaptarse a los distintos grupos de trabajo y debería incorporar tipos de colaboración derivados de las tecnologías Web [2].

2) Diseño de una plataforma tecnológica

En este eje se trabaja en el diseño de una plataforma tecnológica basada en Web que brinde las funcionalidades necesarias para soportar la metodología adoptada para la creación

colaborativa de libros. También debe proporcionar herramientas para compaginar módulos de libros para su uso en un curso específico, para la lectura de libros en línea, para la descarga en formato PDF para imprimirlo o para su posterior lectura fuera de línea. Además, se espera que la plataforma permita la creación de nuevas versiones, adaptaciones o traducciones a otros idiomas, así como herramientas para la edición de nuevos módulos. Como primera etapa se realizó un análisis del estado del arte de plataformas existentes, para su posible adopción y adaptación a los objetivos planteados. Se analizaron, entre otros aspectos, si son plataformas abiertas, si permiten la colaboración y la modularización de contenidos, y cómo tratan el control de calidad y la autoría. Los resultados pueden verse en [3].

3) Definición de las estrategias para la implementación y adopción de esta Iniciativa.

En este sentido, es necesario establecer las orientaciones políticas y legales para el funcionamiento de la Iniciativa, así como plantear el tipo de licencias abiertas que mejor se ajusten a las leyes de los diferentes países y la propiedad intelectual de los libros de texto. Por otro lado, se hace indispensable el reconocimiento del estado de generación y utilización de los libros de texto en la región. La investigación de esta situación se planteó por dos vías complementarias: recuperación de información estadística del contexto y estudio de las percepciones y actitudes de los actores (estudiantes y profesores) del proceso educativo. Esto se realizó mediante encuestas en forma de una consulta electrónica abierta utilizando una herramienta en la web ([www.limesurvey.org/](http://www.limesurvey.org/)) y fomentando la participación masiva de alumnos y docentes de las universidades involucradas en LATIn. A partir de las conclusiones del análisis en elaboración de esta información se podrán determinar las mejores estrategias a seguir.

En el proyecto [ING348] se trabaja en la expansión semántica de la búsqueda. El objetivo de esta línea es producir la estrategia de búsqueda temática. Cuando el usuario hace una consulta, ingresa un conjunto de términos que describen el tema de su interés. Luego es necesario un proceso que desambigüe estos términos y los expanda semánticamente incorporando sinónimos y conceptos relacionados. Además, se trabaja en cómo utilizar las características y preferencias del usuario (docentes o alumnos) para que un sistema recomendador no sólo recupere los recursos que

respondan temáticamente a la consulta del usuario, sino que presente aquellos que sean los más adecuados de acuerdo a su perfil. Así, se ayuda al docente en la selección y preparación de material didáctico acorde a la temática a abordar, el tipo de material buscado, el nivel académico del curso y las características de cada grupo de alumnos.

En el proyecto [ING308] se plantea utilizar sistemas multiagentes diseñando a sus componentes con arquitecturas de agentes que los capaciten para actuar de forma flexible y eficiente, en entornos multiagentes. En este sentido se ha propuesto el diseño de agentes BDI graduados (g-BDI) en un entorno social integrado por otros agentes. Resultados preliminares de la arquitectura g-BDI pueden verse en [4]. Se ha utilizado este modelo de agente para diseñar e implementar agentes recomendadores en el dominio de la educación ([5], [6]).

En el proyecto PICTO (Hacia el desarrollo y utilización de Repositorios de Acceso Abierto para ODE en el contexto de las universidades públicas de la región centro-este de Argentina) se trabaja en el diseño y la transferencia de un primer modelo teórico, metodológico y tecnológico de carácter experimental para repositorios de Acceso Abierto de Objetos Digitales Educativos en función de su adopción exitosa por parte de la comunidad académica regional, que permita contribuir al conocimiento de aspectos organizacionales, pedagógicos, sociales y tecnológicos adecuados al contexto de aplicación, y en la optimización de herramientas informáticas de código abierto que contribuyan a la adopción efectiva de dichos repositorios.

## Resultados y Objetivos

El objetivo de este proyecto es la creación y disseminación de Libros de Texto Abiertos para la educación superior. Para lograr este objetivo, se trabaja en el diseño de una metodología para la creación colaborativa de libros de texto abiertos, el diseño de una plataforma tecnológica que soporte la creación, adaptación, combinación y reutilización de estos libros en forma colaborativa, y en la definición de estrategias para la implementación y la adopción de esta iniciativa en las instituciones de educación superior de la región.

Para validar la eficacia de la metodología propuesta, la plataforma tecnológica y las

estrategias de implementación, se plantea desarrollar dieciséis libros abiertos y colaborativos en cada universidad. Los autores serán profesores de las universidades Latinoamericanas participantes y los temas serán sobre áreas consensuadas entre dichas instituciones. Luego, se elegirán cursos en los cuales se utilizarán estos libros creados colaborativamente y se evaluarán los resultados obtenidos. Una vez que esta primera experimentación se haya evaluado, se realizarán los ajustes necesarios y se promoverá esta Iniciativa en las distintas universidades de América Latina.

## Formación de Recursos Humanos

El equipo de trabajo está integrado por la Dra. Ana Casali (investigadora de la Universidad Nacional de Rosario, y del CIFASIS), la Dra. Claudia Deco (investigadora de la Universidad Nacional de Rosario), la M. Sc. Cristina Bender (investigadora de la Universidad Nacional de Rosario), el Dr. Raúl Kantor (profesor de la Universidad Nacional de Rosario) y la Lic. Valeria Gerling (de la Universidad Nacional de Rosario).

Dentro del marco de esta línea de investigación, se desarrollan actualmente las siguientes tesinas de grado de la Licenciatura en Ciencias de la Computación, de la Universidad Nacional de Rosario:

- Extracción automática de metadata de objetos de aprendizaje para su carga en repositorios. Santiago Fontanarrosa. Directoras: Ana Casali, Claudia Deco. En curso.
- Agentes para la recolección de objetos digitales de aprendizaje en la Web. Santiago Beltramone. Directora: Ana Casali. En curso.

## Referencias

[1]. Falquet G., Hurni J., Guyot F. and Nerima L. Learning by creating multipoint of view scientific hyperbooks. En Proceedings del European Perspectives on Computer-Supported Collaborative Learning (CSCL

2001), Maastricht, pp. 222-229. 2001.

- [2]. Casali A., Silva A., Henrique C., Deco C., Frango I., Muñoz Arteaga J., Gomes dos Santos J., Broisin J., Morales R., Hernández Y., Ochôa X. Collaborative methodologies for writing open educational textbooks: a state-of art review. En Proceedings del I Workshop Recursos Educacionais Abertos: Questões para globalização e localização y II International Symposium on OER: Issues for globalization and localization. Congresso Brasileiro de Informática na Educação 2012 - CBIE2012. 26 al 30 de Noviembre 2012. Río de Janeiro. Brazil.
- [3]. Ochôa X., Casali A., Deco C., Gerling V., Frango Silveira I., Fager J., Carillo G., Parra G., Muñoz Arteaga J., Maturana J., Araya E., Motz R. Analysis of existing technological platforms for the collaborative production of Open Textbooks. En Proceedings de EDMedia. World Conference on Educational Multimedia, Hypermedia & Telecommunications. Association for the Advancement of Computing in Education (AACE). June 25 - June 28, 2013 - Victoria, BC, Canada.
- [4]. Casali, A., Godo, Ll. y Sierra, C. g-BDI: A graded intentional agent model for practical reasoning, MDAI 2009. LNAI, Vicenç Torra, Yasuo Narukawa, Masahiro Inuiguchi (eds.), vol. 5861, Awaji Island, Japan, Springer, pp. 5-20, 2009.
- [5]. Casali A., Gerling V., Deco C. y Bender C. Chapter 8: Recommender System for Personalized Retrieval of Learning Objects. In: Book of Educational Recommender Systems and Technologies: Practices and Challenges. ERSAT. Ed. Olga C. Santos and Jesus G. Boticario. aDeNu Research Group. UNED, Spain. ERSAT. 2011.
- [6]. Casali A., Gerling V., Deco C. y Bender C. Sistema inteligente para la recomendación de objetos de aprendizaje \*\*LACLO 2010 Best Papers\*\*. Revista Generación Digital Vol 9, No 1. pp. 88-95. Colombia. 2011.



# Educación / Ingeniería de Software

## La Ingeniería de Software y los Objetos de Aprendizaje como herramientas para las TICs aplicadas a la educación

**Gil, G., Arias Figueroa, D., Gimson, L., Sánchez, E., Ramírez Morales, J., Silvera, J., Rocabado, S.**

**Centro de Investigación y Desarrollo en Informática Aplicada (CIDIA) /  
Facultad de Cs Exactas. / Universidad Nacional de Salta**

**Avenida Bolivia 5150 - Salta - Capital - Argentina A4408FVY**

**Teléfono : 54-0387-4255358**

**Mail: [gdgil@unsa.edu.ar](mailto:gdgil@unsa.edu.ar)**

### RESUMEN

En el Centro de Investigación y Desarrollo en Informática Aplicada (CIDIA) de la Universidad Nacional de Salta venimos trabajando varias líneas de investigación relacionadas con educación por un lado y con la ingeniería de Software por otro lado. Actualmente estas líneas de investigación se consolidaron gracias a la especialización formal de los docentes e investigadores que la conforman.

Los alcances esperados a través de este trabajo es poder lograr fijar pautas que puedan llegar a ser los pilares de una metodología para desarrollar cursos en entornos virtuales y Objetos de Aprendizaje, logrando con ello la estandarización y reutilización de contenidos en la educación apoyada con tecnología, brindando a los docentes de cualquier disciplina una manera accesible y sin necesidad de conocimientos informáticos una fácil elaboración y difusión de contenido que pueda ser reutilizado en diferentes cursos.

#### **Palabras clave:**

*Objetos de Aprendizaje, Entorno virtual, e-learning, ingeniería de software, estandarización de contenido, reutilización de contenido, SCORM, LSM*

### CONTEXTO

El proyecto investigación “*La Ingeniería de Software y los Objetos de Aprendizaje como herramientas para las TICs aplicadas a la educación*”, se desarrolla en la Facultad de Ciencias Exactas de la Universidad Nacional de Salta, el mismo fue acreditado por el Consejo de Investigación de la UNSA.

### INTRODUCCION

Las nuevas tecnologías aportan al campo de la educación aspectos innovadores que suponen una mejora cualitativa en las formas de enseñar y aprender. Su introducción no sólo reduce el coste efectivo de aplicación de teorías y principios de intervención pedagógica sino que abre las posibilidades de exploración de modelos procedentes de diferentes campos, facilitando su compatibilización y permitiendo ofrecer una visión que los englobe. La tendencia tecnológica en la producción de contenido educativo en Web se orienta hacia los Objetos de Aprendizaje, debido a su enorme potencial de reusabilidad, capacidad generativa, adaptabilidad y escalabilidad [12].

En este proyecto nos planteamos continuar con la investigación que se inició como resultado el proyecto de investigación “Desarrollo de un entorno virtual de

enseñanza-aprendizaje para la Universidad Nacional de Salta”, informados en WICC 2010 [4] y 2011 [5], donde se realizaron experiencias de elaboración de Objetos de Aprendizajes con algunas herramientas de autor que soportan el estándar SCORM, que pudieran incorporarse en la plataforma Moodle.

Pretendemos con este proyecto construir los conocimientos para fijar pautas que puedan llegar a ser los pilares de una metodología para desarrollar cursos en entornos virtuales, que sean de uso más generalizado y accesible a través de Objetos de Aprendizaje como una alternativa que combina los aspectos pedagógicos y la facilidad de modelado a través de la Ingeniería de Software.

El enfoque de la investigación es Aplicada ya que con ella emprenderemos la tarea de obtener nuevos conocimientos técnicos de aplicación inmediata a una problemática determinada, además de perseguir una aplicación directa de los conceptos a desarrollar, es decir que se identificarán, priorizarán, planificarán y por último se aplicarán los conceptos.

### **Objetos de aprendizaje**

La definición de Objeto de Aprendizaje que especifica la IEEE, (*Institute of Electrical and Electronics Engineers, Inc.*) [13] es:

*“Un objeto es cualquier entidad digital o no digital que puede ser usada, re-usada o referenciada para el aprendizaje soportado en tecnología”.*

La idea central detrás del uso de los objetos de aprendizaje es la reutilización. El desarrollo de contenido educativo redundante tiene implicaciones de costo y tiempo que en un mercado competitivo resulta inevitable aminorar. Además, compartir cursos completos es difícil e ineficiente debido a que las necesidades y objetivos de aprendizaje varían de acuerdo a cada institución y de una persona a otra. Un enfoque más prometedor es pensar en desarrollar piezas más pequeñas de instrucción que puedan ser compartidas y

reutilizadas en diferentes contextos y que además se puedan combinar para construir bloques de instrucción mayores. Desde este punto de vista, el diseño de contenido educativo no necesariamente tiene que empezar de cero. Un diseñador de cursos sólo tiene que localizar los objetos de aprendizaje que necesita para estructurar un curso a medida.

**Según Morales [14]** un objeto de aprendizaje debe poseer ciertos atributos esenciales que lo distinguen de simples piezas de información:

- ser un objeto educativo.
- proporcionar una cantidad de conocimiento o habilidad relativamente pequeña.
- ser autocontenido.
- ser útil en más de una secuencia de instrucción.
- ser fácil de identificar y por tanto de buscar.
- ser independiente de un sistema administrador del aprendizaje (Learning Management System, o LMS) específico y
- ser accesible desde una gran variedad de plataformas.

Para el modelo SCORM de ADL [1], un objeto reusable de aprendizaje (Sharable Content Object, SCO por sus siglas en inglés) es la unidad mínima de contenido educativo a la que da seguimiento un sistema administrador del aprendizaje. Un SCO está diseñado para poder ser usado en diferentes contextos (por ejemplo, diferentes cursos) por lo que no puede depender de que otros SCO sean presentados al estudiante en distintos momentos.

Existen varias herramientas para “empaquetar” contenidos. Algunas son pagas y otras gratuitas. Estas herramientas son específicas para e-learning, y nos ofrecen un entorno completo en el que diseñar los contenidos no requieren de programación, “empaquetando” de forma automática los contenidos con ellas creados y generando

todo el código necesario para el seguimiento de la actividad del alumno sobre el contenido en cualquier LMS compatible. Así una vez, obtenido el contenido empaquetado, y elegido el LMS según SCORM, se puede subir este archivo. Nuestro grupo se halla abocado a analizar y evaluar diferentes herramientas para generar objetos de aprendizaje.

## **LINEAS DE INVESTIGACION y DESARROLLO**

Los principales ejes temáticos que se están investigando son los siguientes:

- Tecnología Informática Aplicada en Educación.
- Objetos de Aprendizaje.
- Gestores de Contenidos Educativos (**LMS**, *Learning Management Systems*).
- SCORM, *Sharable Content Object Reference Model* (Modelo de Referencia para Objetos de Contenidos Intercambiables).
- Herramientas de Autor para el desarrollo de contenidos SCORM.

## **RESULTADOS ESPERADOS**

Un punto importante que surge de la experiencia realizada por el grupo de investigación hasta el momento, es que las tecnologías que brindan soporte a los diferentes Learning Management Systems, son heterogéneas. Esto conlleva a trabajar con contenidos con características, objetivos y formatos de todo tipo. Del mismo modo, la naturaleza de las personas que interactúan con éstas plataformas, hace que la participación de éstas sea persiguiendo diferentes intereses y objetivos. Por lo anteriormente expuesto creemos que es importante analizar con detenimiento el uso de herramientas de autor y herramientas que permitan la generación de Objetos de Aprendizaje, los cuales nos permiten plantear nuevas forma de desarrollo de material educativo y nuevas metodologías. Sin embargo, actualmente se carece de una

metodología de ingeniería de software específica que asegure la calidad de los objetos de aprendizaje. Por ello, se requiere de un proceso de Ingeniería de Software que permita determinar la secuencia completa de etapas que se presentan durante su diseño y desarrollo hasta su uso y reuso proporcionarle a los desarrolladores de contenido educativo una guía que facilite el proceso de producción de los objetos de aprendizaje.

En función a lo expresado en este trabajo, nuestra investigación tiende a un objetivo concreto, poder proponer una metodología, a través de la ingeniería de software, para la construcción de sistemas basados en objetos de aprendizaje, es decir, describir métodos, prácticas, recursos y artefactos necesarios para la construcción de dichos sistemas.

Si se tiene en cuenta que los Objetos de Aprendizaje se congregan en lecciones y que un conjunto de lecciones constituye un curso, se presenta como uno de los principales impactos esperado la estandarización y reutilización de contenidos en la educación apoyada con tecnología, brindando a los docentes de cualquier disciplina una manera accesible y sin necesidad de conocimientos informáticos para una fácil elaboración y difusión de contenido que pueda ser reutilizado en diferentes cursos.

## **4. FORMACION DE RECURSOS HUMANOS**

La estructura del equipo de investigación es de 7 (siete) miembros incluidos el Director y Co-director.

Cinco miembros del proyecto están realizando el trabajo de Tesis de Posgrado, dos en Ingeniería de Software y dos en Redes de Datos, todas dependientes de la Universidad Nacional de La Plata, y uno en Administración de Negocio dependiente de la Universidad Católica de Salta.

Además continuamos con la dirección de tesis de grado de los alumnos de la Licenciatura en Análisis de Sistemas dependiente de la Facultad de Ciencias Exactas de la Universidad de Salta.

## 5. BIBLIOGRAFIA

- [1] Advanced Distributed Learning, (ADL). (2004). Sharable Content Object Reference Model (SCORM®) 2004 2nd Edition Overview.
- [2] Azpeitia, I., Monge, S. y Ovelar, R. “Una aproximación del diseño de una guía de buenas prácticas en torno al paradigma de los learning objects”. Universidad del País Vasco, 2005.
- [3] Chan Nuñez, Maria Elena (2002) Revista “Apertura”, Innova, Conferencias Magistrales del X Encuentro Internacional de Educación a Distancia “Hacia la construcción de la sociedad de aprendizaje”, Cuernavaca Morelos, México.
- [4] Gil, Gustavo y otros, “Desarrollo de un entorno virtual de enseñanza – aprendizaje para la universidad nacional de salta”. XII Workshop de Investigadores en Ciencia de la Computación, realizado del 5 al 6 de mayo del 2010 – El Calafate, Argentina.
- [5] Gil, Gustavo y otros, “Entorno virtual con objetos de aprendizaje”. VI Congreso de Tecnología en Educación y Educación en Tecnología. Salta, Argentina. ISBN 978-987-633-072-5. 14 al 16 de junio de 2011.
- [6] Gil, Gustavo y otros, “Plataformas e-learning y su evaluación”. III Congreso de Tecnología en Educación y Educación en Tecnología. Bahía Blanca, 12 y 13 de junio de 2008.
- [7] Gil, Gustavo y otros, “Una experiencia en la formación de docentes para el dictado de cursos on-line”. II Jornadas Ingreso y Permanencia en Carreras Científicas – Tecnológicas. 19 al 21 de mayo de 2010. Salta, Argentina. ISBN 978-987-633-056-5.
- [8] Lopez, M. G., Miguel, V. y Montañó N. (2005). Sistema Generador de AMBientes de Enseñanza-Aprendizaje Constructivistas basados en Objetos de Aprendizaje (AMBAR). II Simposio Pluridisciplinar sobre Diseño, Evaluación y Descripción de Contenidos Educativos Reutilizables (SPDECE). Barcelona, España.
- [9] Pressman, R. “Ingeniería del Software: Un enfoque práctico”. McGraw-Hill, 2009.
- [10] Rodríguez, F. y Cardona, P. “Metodología para el desarrollo de cursos virtuales basados en objetos de aprendizaje”. Universidad Autónoma de Aguascalientes, 2005.
- [11] Sánchez, Ernesto, “Herramientas e-learning como soporte a la formación presencial universitaria. II Jornadas Ingreso y Permanencia en Carreras Científicas – Tecnológicas. 19 al 21 de mayo de 2010. Salta, Argentina. ISBN 978-987-633-056-5.
- [12] Wiley, David (2001), “Connecting learning object to instructional design theory: a definition, a metaphor, and a taxonomy”. Utah State University.
- [13] *Institute of Electrical and Electronics Engineers*: <http://www.ieee.org/portal/site>. Consultado en marzo de 2011
- [14] Morales G. Rafael y Agüera H. Capacitación basada en objetos reusables de aprendizaje. Boletín IIE, enero-febrero del 2002



## DETERMINACION DE FACTORES RESILIENTES EN LOS ALUMNOS UNIVERSITARIOS

Mag. Graciela Beguerí, Mag. Raúl Klenzi, Mag. M. Alejandra Malberti  
Instituto de Informática (IdeI) / Departamento Informática (DI) / Facultad de Ciencias  
Exactas Físicas y Naturales (FCEFN) / Universidad Nacional de San Juan (UNSJ)  
Av. Ignacio de la Roza 590 (O), Complejo Universitario "Islas Malvinas", San Juan  
{grabeda, rauloscarklenzi, amalberti} @gmail.com

### Resumen

El presente trabajo expone los resultados de la experiencia de modelar la capacidad de resiliencia de alumnos universitarios avanzados mediante el constructo generado por Wagnild y Young, y relevado a alumnos de las carreras del Departamento Informática de la FCEFN-UNSJ. Esta encuesta, reformulada y adaptada, se describe mediante técnicas de análisis multivariable y tareas de segmentación que posibilitan encontrar las dimensiones-factores que caracterizan al perfil resiliente de los alumnos.

**Palabras clave:** Análisis Multivariable, Minería de Datos, Resiliencia, Deserción universitaria

### Contexto

La línea de investigación se enmarca en el proyecto bianual 2011-2012 “**MINERÍA DE DATOS EN LA DETERMINACIÓN DE PATRONES DE USO Y PERFILES DE USUARIO**” código 21/E889 que se desarrolla en el ámbito de la FCEFN-UNSJ, aprobado por el Consejo de Investigaciones Científicas Técnicas y de Creación Artística (CICITCA), financiado por la propia Universidad y ajustado a evaluación externa.

La reciente instancia de acreditación a la que se vieron sometidas las diferentes titulaciones del área informática del país permitió reconocer una problemática común a las distintas carreras, como son el rezago, la cronicidad y la deserción de alumnos universitarios.

En este contexto desde el proyecto se está intentando modelar el perfil resiliente del alumno de las carreras del Departamento de Informática mediante técnicas de análisis multivariable y minería de datos aplicadas a los resultados de una encuesta generada desde el constructo propuesto por Wagnild y Young, revisado en 1993. Mediante este trabajo se pretende modelar y desde allí segmentar en tres posibles valores (escasa, moderada, alta) la capacidad de resiliencia de los alumnos. Este modelo posteriormente se aplicaría a los alumnos ingresantes y aquel alumno que desde la descripción de la encuesta resultare contenido por el segmento de escasa resiliencia merecería la mayor atención por parte del área de tutorías, gabinetes psicopedagógicos de la FCEFN atento a que su perfil se ajustaría al de un potencial alumno desertor.

### Introducción

Alieto Aldo Guadagni en su informe *Deserción, Desigualdad y Calidad Educativa* expone la problemática en este sector y señala “cuatro problemas graves

que debilitan nuestro proceso educativo, a saber: alto nivel de deserción en el nivel secundario, creciente desigualdad educativa entre provincias y tipo de escuelas, grave retroceso en la calidad educacional y finalmente, una Universidad con muy pocos graduados, especialmente en las carreras orientadas al mundo productivo del futuro”.

Específicamente la deserción y rezago de alumnos en carreras universitarias son problemas que aquejan a la mayoría de las carreras del país, en especial a aquellas pertenecientes a las ciencias exactas, impactando negativamente en todo el sistema de educación superior.

Un relevamiento estadístico realizado en las carreras del Departamento de Informática revela que del total de alumnos ingresantes en 2011, grupo de estudio y encuestados en el marco de la determinación del perfil de resiliencia, el 50% no se inscribe para cursar la totalidad de las asignaturas, si bien por su carácter de ingresantes están en condición de hacerlo. Esto refleja una situación de rezago “por opción” o “voluntaria” desde el mismo inicio de la carrera elegida.

Numerosos estudios versan sobre el análisis de resiliencia. En particular Wagnild y Young (1993), expresan que “la resiliencia es una característica de la personalidad que modera el efecto negativo del estrés y fomenta la adaptación”. En tanto que en *Estado del arte en resiliencia*, elaborado por el Centro de Estudios y Atención del Niño y la Mujer, entre otras, se define a la resiliencia como “la capacidad humana universal para hacer frente a las adversidades de la vida, superarlas o incluso ser transformado por ellas. La resiliencia es parte del proceso evolutivo y debe ser promovida desde la niñez”. Parraga (2012) presenta una síntesis de las cualidades reseñadas en los estudios clásicos de resiliencia, realizados por

diferentes autores. En este sentido, los establecimientos educativos y todos los elementos que forman parte de la educación, pueden ser claves para generar procesos de promoción de resiliencia, centrando recursos y aunando esfuerzos de modo de estimular los factores resilientes en aquellos alumnos que se encuentren en un contexto de desventaja.

## Líneas de investigación y desarrollo

Se trabajó según el constructo propuesto por Wagnild y Young, adaptado y reformulado por Beguerí-Malberti. En él se etiquetaron los resultados de la resiliencia en Escasa, Moderada y Mayor. El cuestionario original, que contenía 25 atributos-

<http://wilderdom.com/html/ResilienceScale.doc>, fue traducido y acondicionado a la realidad socio cultural del alumnado concluyendo éste en 23 atributos. El mismo fue aplicado a 50 encuestas de alumnos avanzados. Éstas fueron procesadas y se les aplicó un análisis multivariante, lo que permitió encontrar 9 factores que describen la capacidad de resiliencia de los alumnos.

Factores	1	2	3	4	5	6	7	8	9
Atributos	17	12	<u>20</u>	7	<u>18</u>	19	11	<u>5</u>	<u>4</u>
	9	3	15	<u>13</u>	<u>10</u>	<u>14</u>	<u>21</u>	<u>2</u>	
	1	22			23				
	6	<u>8</u>							
	<u>16</u>								

Figura 1- Factores que describen la capacidad de resiliencia

La Figura 1 presenta los 9 factores que, de acuerdo con los atributos involucrados, reciben las siguientes nominaciones: Capacidad de relacionarse – Sentirse bien solo – Satisfacción personal – Introspección – Independencia – Perseverancia – Iniciativa – Autoestima – Capacidad de relacionarse

Con el propósito de modelar el constructo adaptado por Beguerí-Malberti desde una tarea de segmentación, se agruparon los 50 registros en tres grupos diferentes de modo de contrastar los resultados con los obtenidos anteriormente.

Para ello se utiliza la herramienta de software libre RapidMiner (RM) 3.5.005 que permite, entre sus múltiples capacidades, realizar tareas de segmentación. En este caso se utiliza el algoritmo de segmentación W-SimpleKmeans- que admite trabajar con valores desconocidos en los registros y agrupar los 50 registros en tres grupos diferentes. Se destaca que sólo 9 valores fueron omitidos, voluntaria o involuntariamente, por los encuestados.

La comparación entre la capacidad de resiliencia obtenida desde la expresión de Wagnild y Young versus la tarea de segmentación (cuyos resultados de resiliencia son los valores considerados como de “predicción”) arroja la siguiente performance y matriz de confusión:

Accuracy: 54.00%				
	true Escasa	true Moderada	true Mayor	class precision
pred. Escasa	7	9	0	43.75%
pred. Moderada	1	7	2	70.00%
pred. Mayor	0	11	13	54.17%
class recall	87.50%	25.93%	86.67%	

Figura 2- Matriz de confusión del modelo de segmentación: algoritmo W-SimpleKmeans

Se observa en la Figura 2 que si bien hay una buena aproximación entre uno y otro modelo para la predicción de resiliencia “Escasa” y “Mayor”, no ocurre lo mismo en el valor de predicción ”Moderada”.

Seguidamente se aplica al modelo alcanzado, mediante tareas de segmentación, una instancia de clasificación con el objeto de describir la encuesta y los valores de predicción obtenidos por el modelo. En esta instancia se usa un árbol basado en ganancia de información relativa -W-J48-, sin poda y provisto por el software mencionado. La Figura 3 muestra la información obtenida en el proceso realizado.

**W-J48**

J48 unpruned tree  
-----

```

2 arreglo <= 3
| 10 decidida <= 2: Escasa (6.0)
| 10 decidida > 2
| | 20 vida <= 3: Escasa (3.0)
| | 20 vida > 3
| | | 8 Me sient <= 2: Escasa (1.0)
| | | 8 Me sient > 2
| | | | 14 interés <= 2: Escasa (1.0)
| | | | 14 interés > 2: Moderada (8.0)
2 arreglo > 3
| 16 atraviesa <= 2
| | 5 molesto <= 3: Moderada (1.0)
| | 5 molesto > 3: Escasa (2.0)
| | 16 atraviesa > 2
| | | 13 cumplido <= 2: Escasa (2.0)
| | | 13 cumplido > 2
| | | | 21 insistir <= 3
| | | | | 18 pto vis <= 2
| | | | | 4 interésl <= 3: Escasa (1.0)
| | | | | 4 interésl > 3: Mayor (2.0)
| | | | 18 pto vis > 2: Mayor (21.96)
| | | 21 insistir > 3: Moderada (1.04/0.04)
    
```

Number of Leaves : 12  
Size of the tree : 23

Figura 3- Árbol sin poda del modelo de clasificación: Algoritmo W-J48

A continuación se detallan solo 11 atributos, los cuales pueden ser interpretados como inductores. Éstos aparecen en el árbol y describen la totalidad de la encuesta. La Figura 4

presenta los atributos inductores resultantes.

2 arreglo  
10 decidida  
20 vida  
8 Me sient  
14 interés  
16 atraviesa  
5 molesto  
13 cumplido  
21 insistir  
18 ptos vis  
4 interésI

Figura 4- Atributos Inductores de la capacidad de resiliencia

Es de destacar que estos atributos se encuentran distribuidos en la tabla 1 en los diferentes factores (valores subrayados) como se expresa en la Figura 5.

FACTOR	ATRIBUTOS
1	16
2	8
3	20
4	13
5	18, 10
6	14
7	21
8	5, 2
9	4

Figura 5- Correspondencia entre factores y atributos inductores

## Resultados y Objetivos

### Resultados:

Al momento de la elaboración de este documento, se han obtenido los siguientes resultados:

- El modelado desde tareas de segmentación resulta una buena aproximación entre los resultados

alcanzados en estudios de resiliencia y nuestra adaptación.

- Los atributos surgidos desde el árbol de clasificación, y que describen la totalidad de la encuesta, están contenidos en los factores encontrados anteriormente.

### Objetivos:

Si bien en el grupo de trabajo han surgido variadas conjeturas relacionadas con la problemática abordada, se destacan las vinculadas con los objetivos planteados a continuación:

- Comparar los resultados de las tareas de segmentación de alumnos avanzados e ingresantes, contra la capacidad de resiliencia surgida de la fórmula de Wagnild y Young.
- Dar respuesta a los siguientes interrogantes: Dado que de los alumnos avanzados se dispone de información académica ¿existirá alguna relación entre aquellos alumnos con alto rendimiento académico y un mayor nivel de resiliencia? ¿serán aquellos cuya etiqueta es de escasa resiliencia, los considerados como rezagados o posibles desertores?

## Formación de Recursos Humanos

Los investigadores integrantes del proyecto ponen énfasis en la formación de recursos humanos. En este sentido se están tutelando 7 becarios en el área de Minería de Datos, en el marco de la convocatoria a becas de finalización de carrera de la Agencia Nacional de Promoción Científica y Tecnológica

A su vez se están dirigiendo dos tesis de posgrado en las áreas Minería de Texto e



Inteligencia de Negocios, respectivamente.

## Referencias

- [1] Kantardzic, M. (2003) *Data Mining: Concepts, Models, Methods, and Algorithms* ISBN:0471228524 John Wiley & Sons © (343 pages)
- [2] Hair, J. F.; Anderson, R.E.;Tatham, R.L. y Black, W.C. (1999). *Análisis Multivariante. 5ª Ed.* Prentice Hall Iberia, Madrid.
- [3] Guadagni, A. (2011) *Deserción, Desigualdad y Calidad Educativa* Econométrica S.A economic research and forecasts Informe Especial N°415 [http://www.econometrica.com.ar/productosservicios/archivos/especiales/iespecial\\_415.pdf](http://www.econometrica.com.ar/productosservicios/archivos/especiales/iespecial_415.pdf)
- [4] Han, J; Kamber, M. ( 2006) *Data Mining: Concepts y Techniques. Second Editions.* Morgan Kaufmann Publisher.
- [5] Hand, David; Mannila, Heikki; Smyth, Padhraic ( 2001) *Principles of Data Mining.* The MIT Press.
- [6] Hastie ,T.; Tibshirani, R.; Friedman, J. (2001)*The Elements of Statistical Learning.Data Mining, Inference, and Prediction.* Springer-Verlag,
- [7] Hernández Orallo J., Ramirez Quintana, J, Ferri Ramirez, C. (2008) *Introducción a la Minería de Datos.* Pearson-Prentice Hall.
- [8] Larose, D. (2006) *Data Mining. Methods and Models.* Department of Mathematical Sciences Central Connecticut State University Wiley. A John Wiley & Sons, Inc Publication.
- [9] Larose, D. (2005) *Discovering Knowledge In Data - An Introduction to Data Mining.* John Wiley & Sons, Inc., Publication.
- [10] Manning C, Prabhakar R. Hinrich & Hinrich Schütze, (2009) *An Introduction to Information Retrieval,* Cambridge University Press.
- [11] Neill, James. *Summaries of Instruments* <http://wilderdom.com/tools/ToolsSummaries.html>
- [12] Párraga, G. T. R., & Martínez, A. E. L. (2012). Resiliencia psicológica y dolor crónico. *Escritos de Psicología*, 5(2), 1-11.
- [13] Peña, Daniel (2002), *Análisis de datos multivariantes.* Ed. Mc. Graw Hill, España.
- [14] R <http://www.r-project.org/>
- [15] Rapid-I . <http://rapid-i.com/api/rapidminer-5.1/com/rapidminer/tools.2011>.
- [16] Tolosa G. y Bordignon, F. (2007) *Introducción a la Recuperación de Información. Conceptos, modelos y algoritmos básicos* UNLu, Arg..
- [17] Wagnild, G.M. & Young, H.M. (1993) *Development and psychometric evaluation of the Resilience Scale. Journal of Nursing Measurement.* Vol 1, pp165-178.
- [18] Wolin, S. and S. J. Wolin, (1997) *Shifting Paradigms: Easier Said Than Done Resiliency in Action.*

# La Autorregulación de los Aprendizajes y el Aprendizaje Colaborativo mediado por TIC en el Nivel Superior

Edith Lovos<sup>(1)</sup>, Tatiana Gibelli<sup>(1)</sup>, Rodolfo Bertone<sup>(2)</sup>

<sup>(1)</sup> Universidad Nacional de Río Negro

Sede Atlántica

Av. Don Bosco y Leloir s/n, Viedma, Río Negro

<sup>(2)</sup> Instituto de Investigación en Informática LIDI (III-LIDI)

Facultad de Informática – UNLP

50 y 120 La Plata Buenos Aires

{elovos,tgibelli}@unrn.edu.ar

pbertone@lidi.info.unlp.edu.ar

## Resumen

Este trabajo presenta los avances y resultados obtenidos durante el año 2012, vinculados al área de tecnologías aplicadas en el ámbito educativo de nivel superior.

Las actividades de investigación se desarrollan en la Universidad Nacional de Río Negro en el marco de una beca de investigación de postgrado y como parte de tesis de maestría de las integrantes.

Los procesos de investigación que se desarrollan están vinculados a las siguientes temáticas: autorregulación de los aprendizajes mediados por TIC y trabajo colaborativo aplicado a la enseñanza-aprendizaje de la programación en alumnos ingresantes a través del uso de las tecnologías.

El trabajo de investigación persigue, en una primera etapa, la formación de Recursos Humanos con formación en postgrado, Maestría de Educación. En una etapa posterior, se prevé continuar la línea de investigación dentro de la UNRN formando recursos propios de la carrera.

**Palabras claves:** trabajo colaborativo, TIC, autorregulación de los aprendizajes

## Contexto

Las investigaciones realizadas se desarrollan en el marco del programa de becas de investigación de la Universidad Nacional de Río Negro (UNRN) conducentes a la formación de posgrado y al desarrollo de dos tesis de maestría.

## Introducción

Una de las líneas de investigación toma como referencia el concepto de aprendizaje autorregulado, tema de investigación reciente, con un abordaje cognitivo del aprendizaje, relacionándolo con formas de aprendizaje académico independientes y efectivas que implican meta-cognición, motivación intrínseca y acción estratégica [1]. Varios autores concluyen que es necesario considerar el papel del adulto en el desarrollo de la autorregulación y particularmente, la estimulación para el desarrollo del aprendizaje autorregulado. Existen algunas investigaciones sobre intervenciones y modelos instruccionales diseñados con el objetivo de enseñar los procesos y las estrategias involucradas en el aprendizaje autorregulado [2]. En este marco se propone una investigación en tono a una intervención en que se incorpora el uso de TIC en el proceso.

Pierre Dillenburg, define al aprendizaje colaborativo como toda aquella situación en la que un grupo de individuos establecen sesiones de trabajo mediante las cuales intentan aprender algo en colaboración [3]. Aprender con otros y de otros, hace referencia en lo que la psicología se conoce como zonas de desarrollo próximo [4], y en los enfoques pedagógicos se plantea como estrategia dialógica [5].

Zañartu Correa [6], entiende que el aprendizaje colaborativo, nace y responde a un nuevo contexto socio cultural donde se define el “cómo aprendemos” (socialmente) y “dónde aprendemos” (en red). En este sentido, Estevez et al. [7] sostienen que los ambientes colaborativos pueden ofrecer un importante soporte a los alumnos durante las actividades aprendizaje de la programación. Y agrega que la resolución de problemas a través de la colaboración promueven la reflexión, un mecanismo que estimula el proceso de aprendizaje. Para el desarrollo de una actividad grupal los alumnos necesitan comunicarse, discutir y emitir opiniones a otros miembros del grupo, alentando de esta forma una actitud de reflexión que conduce al aprendizaje.

Por otra parte el desarrollo del software a nivel profesional no escapa a esta realidad, y es común que las actividades de producción de software se realicen en forma colaborativa haciendo uso de las TIC. Resulta entonces, importante la formación de los alumnos de las carreras vinculadas a la informática, desde su inicio, no solo en las áreas específicas sino lograr una formación integral en la que desarrollen habilidades sociales y personales que promuevan un aprendizaje colaborativo.

En este sentido el trabajo de investigación se propone investigar y determinar como la metodología de trabajo colaborativo aplicada a la enseñanza de la programación en los cursos introductorios, potencia su aprendizaje.

## Líneas de Investigación y Desarrollo

A continuación se hace referencia a las principales líneas de investigación y desarrollo en las que se trabaja desde el marco del proyecto. Para cada línea se mencionan los objetivos específicos.

### Autorregulación del Aprendizaje mediado por TIC

- Describir las principales estrategias de aprendizaje en el área de matemática, particularmente las vinculadas a la autorregulación del aprendizaje, utilizadas por los alumnos en primer año universitario.
- Analizar características del trabajo de los alumnos en el aprendizaje apoyado en recursos TIC como uso de internet (en realización de webquest), software matemático específico, tareas colaborativas online (uso de foros, wiki) y cuestionarios de autoevaluación.
- Observar si existen cambios en la autorregulación del aprendizaje de los alumnos luego de un curso desarrollado en modalidad blended learning incorporando distintos recursos TIC.

### Trabajo colaborativo mediado por TIC

- Indagar acerca de los beneficios pedagógicos del trabajo colaborativo aplicado en los ámbitos de enseñanza de nivel superior.
- Identificar interfaces colaborativas que permitan el desarrollo de algoritmos en el ámbito de la programación.
- Realizar una revisión de las herramientas que permitan la programación colaborativa en cursos introductorios.
- Analizar como el uso de un ambiente colaborativo, potencia las habilidades necesarias para el desarrollo de software, en las prácticas de las materias de programación de los primeros años de la Lic. en Sistemas de la UNRN.

## Resultados esperados/obtenidos

### Autorregulación del Aprendizaje mediado por TIC

En el marco de esta investigación se diseñó e implementó una intervención en un curso de matemática de primer año universitario [8]. A partir de la implementación se obtuvieron algunos resultados en que se observaban las ventajas del uso de las TIC y el trabajo colaborativo para el aprendizaje [9]. Se realizó un análisis de las estrategias de aprendizaje de los alumnos y se están realizando análisis estadísticos en base a cuestionarios implementados pre y post curso para observar si hubo cambios en algunas de las variables analizadas. Asimismo, se realizó un análisis de las percepciones y valoraciones de los alumnos acerca de la propuesta implementada [10].

### Trabajo colaborativo mediado por TIC

Vinculado a esta línea de investigación se llevó a cabo una revisión bibliográfica referida al uso de las TIC como mediadoras de estrategias de enseñanza colaborativa en los cursos de programación de nivel superior y se definieron las características deseables para las herramientas, de manera que faciliten la actividad grupal [11].

Se realizó una revisión sobre tres herramientas que permiten implementar los procesos de colaboración en la enseñanza-aprendizaje de la programación a saber: COLLEGE (Edition, Compiling and Execution of Programs), EclipseGavab, Virtual Programming Lab (VPL). [12]

Se formuló una propuesta de enseñanza-aprendizaje colaborativa, haciendo uso de las herramientas provistas por el EVEA Moodle y el laboratorio virtual VPL para las actividades de laboratorio de la cátedra programación I de la Lic. en Sistemas de la UNRN [13]. La misma está prevista implementarla en el primer cuatrimestre del año académico 2013. Al finalizar la misma se realizará una evaluación de las actividades

desarrolladas y un análisis de los resultados alcanzados.

### Formación de Recursos Humanos

Ambas integrantes del proyecto presentado están desarrollando su Tesis de Magister en Tecnología Informática Aplicada en Educación de la Facultad de Informática de la UNLP. Además una de ellas es becaria en el marco del programa de becas de investigación de la Universidad Nacional de Río Negro (UNRN) conducentes a la formación de posgrado 2011. Ambas son docentes de grado en la carrera Lic. en Sistemas de la UNRN.

### Referencias

- [1] Perry, N.E. Introduction: Using qualitative methods to enrich understandings of self-regulated learning. *Educational Psychologist*, 37(1), 1-3. 2002
- [2] Schunk, D.H., y Zimmerman, B.J. (Eds.). *Self-regulated learning: From teaching to self-reflective practice*. New York: Guilford Press. 1998
- [3] Dillenbourg, P. . What do you mean by collaborative learning?. En Dillenbourg (Ed) *Collaborative – learning: Cognitive and Computational Approaches*. 1 – 19. Oxford: Elsevier. P. 1999.
- [4] Vygotski, Lev. S. *El Desarrollo de los procesos psicológicos superior*. Barcelona . Grupo editorial Grijalbo. 1978.
- [5] Freire, Paulo. *La educación como práctica de la Libertad*, Siglo XXI: México. 1970.
- [6] Zañartu Correa Luz M. *Aprendizaje colaborativo: una nueva forma de Diálogo Interpersonal y en Red*. *Revista Digital de Educación y Nuevas Tecnologías*. Año V, Nro 28. 2003.
- [7] Esteves M., Morgado L., Martins P., Fonseca B.. *The use of Collaborative Virtual Environments to provide student's contextualisation in programming*. *Proceedings of m-ICTE 2006*



[8] Gibelli, Tatiana y Chiecher, Analía. Autorregulación del aprendizaje en entornos mediados por TIC. Una propuesta de intervención en matemática universitaria de primer año. Anales del XVIII Congreso Argentino de Ciencias de la Computación. Pag. 627 a 636. 2012

[9] Gibelli, Tatiana. Trabajos colaborativos usando software de cálculo simbólico para el aprendizaje en matemática universitaria de primer año. En Actas del *Primer Congreso Argentino de la Interacción-Persona Computador@*, Telecomunicaciones, Informática e Información Científica. IPCTIIC 2012.

[10] Gibelli, Tatiana y Chiecher, Analía. Percepciones de los alumnos y uso de recursos en una experiencia en modalidad de aula extendida en matemática universitaria de primer año. Aceptado para presentación en I Jornadas Nacionales y III Jornadas de Experiencias e Investigación en Educación a Distancia y Tecnología Educativa organizadas por Proed – UNC – Marzo 2013.

[11] Edith Lovos, Alejandro Gonzalez, Inés Mouján, Rodolfo Bertone, Cristina Madoz. Estrategias de Enseñanza Colaborativa para un Curso de Programación de Primer Año de la Lic. en Sistemas”. I Workshop de Innovación en Educación en Informática. CACIC2012. ISBN: 978987-1648-34-4. Pág. 1534-1543. Bahía Blanca, Buenos Aires, Argentina, Octubre 2012.

[12] Lovos, Edith. Revisión de Herramientas Colaborativas para la Enseñanza de la Programación a Alumnos Novatos. En Actas del *Primer Congreso Argentino de la Interacción-Persona Computador@*, Telecomunicaciones, Informática e Información Científica. IPCTIIC 2012. Córdoba

[13] Lovos Edith, González Alejandro, Fernández Mouján Inés. Combinando ABP y Herramientas Colaborativas para la Enseñanza de Programación en el Primer Año de la Lic. en Sistemas de la UNRN. Aceptado para presentación en I Jornadas Nacionales y III Jornadas de

Experiencias e Investigación en Educación a Distancia y Tecnología Educativa organizadas por Proed – UNC – Marzo 2013

# APRENDIZAJE BASADO EN PROBLEMAS (ABP) EN LOS ALUMNOS DE CARRERAS TÉCNICAS

Claudia Minnaard, Vivian Minnaard

Universidad CAECE

Av. De Mayo 866 – Ciudad Autónoma de Buenos Aires

Facultad de Ingeniería de la Universidad Nacional de Lomas de Zamora.

Instituto de Investigaciones en Tecnología y Educación (IIT&E)

Camino de Cintura y Avda Juan XXIII – Llavallol.

Telefonos: 011-4282-7880

e- mails: [minnaard@uolsinectis.com.ar](mailto:minnaard@uolsinectis.com.ar), [vivianminnaard@gmail.com](mailto:vivianminnaard@gmail.com)

## Resumen

El Aprendizaje Basado en Problemas (ABP) es uno de los métodos de enseñanza - aprendizaje que ha tomado más arraigo en las instituciones de educación superior en los últimos años. El camino que toma el proceso de aprendizaje convencional se invierte al trabajar en el ABP. Mientras tradicionalmente primero se expone la información y posteriormente se busca su aplicación en la resolución de un problema, en el caso del ABP primero se presenta el problema, se identifican las necesidades de aprendizaje, se busca la información necesaria y finalmente se regresa al problema.

Durante el proceso de interacción de los alumnos para entender y resolver el problema se logra, además del aprendizaje del conocimiento propio de la materia, que puedan elaborar un diagnóstico de sus propias necesidades de aprendizaje. El presente trabajo se centra en analizar el grado de aceptación que tiene por parte de los alumnos de carreras técnicas el ABP y que tipo de conductas y participaciones se detectan. Asimismo, se procura indagar las representaciones que tienen los docentes sobre que características tiene un problema adecuado para el ABP.

## Contexto

La Universidad CAECE comenzó su actividad en abril de 1967. Tras obtener rápidamente el reconocimiento provisorio del Ministerio de Educación de la Nación lanzó sus primeras carreras: las Licenciaturas en Matemática y Sistemas. En su momento, la Licenciatura en Sistemas fue la primera de su índole creada en el país. La expansión de la oferta académica continuó con la creación de la Ingeniería en Sistemas y las licenciaturas en Ciencias de la Computación y Gestión de Sistemas y Negocios. En este contexto se plantea la necesidad de investigar el Aprendizaje Basado en Problemas como método de enseñanza - aprendizaje en las carreras técnicas

## Introducción

El ABP es una estrategia de enseñanza-aprendizaje en la que tanto la adquisición de conocimientos como el desarrollo de habilidades y actitudes resulta importante, en el ABP un grupo pequeño de alumnos se reúne, con la facilitación de un tutor, a analizar y resolver un problema seleccionado o diseñado especialmente para el logro de ciertos objetivos de aprendizaje. Durante el proceso de interacción de los

alumnos para entender y resolver el problema se logra, además del aprendizaje del conocimiento propio de la materia, que puedan elaborar un diagnóstico de sus propias necesidades de aprendizaje, que comprendan la importancia de trabajar colaborativamente, que desarrollen habilidades de análisis y síntesis de información, además de comprometerse con su proceso de aprendizaje. (Allen, 2008; Albanese et al, 1993; Ayala-Valenzuela et al; Roldan Borassi, 2012)

El ABP se sustenta en diferentes corrientes teóricas sobre el aprendizaje humano, tiene particular presencia la teoría constructivista, de acuerdo con esta postura en el ABP se siguen tres principios básicos:

- El entendimiento con respecto a una situación de la realidad surge de las interacciones con el medio ambiente.
- El conflicto cognitivo al enfrentar cada nueva situación estimula el aprendizaje.
- El conocimiento se desarrolla mediante el reconocimiento y aceptación de los procesos sociales y de la evaluación de las diferentes interpretaciones individuales del mismo fenómeno.

El ABP incluye el desarrollo del pensamiento crítico en el mismo proceso de enseñanza - aprendizaje, no lo incorpora como algo adicional sino que es parte del mismo proceso de interacción para aprender. El ABP busca que el alumno comprenda y profundice adecuadamente en la respuesta a los problemas que se usan para aprender, lo cual motiva a un aprendizaje consciente y al trabajo de grupo sistemático en una experiencia colaborativa de aprendizaje.

Es importante señalar que el objetivo no se centra en resolver el problema sino en que éste sea utilizado como base para identificar los temas de aprendizaje para su estudio de manera independiente o grupal, es decir, el problema sirve como detonador para que los alumnos cubran los objetivos de aprendizaje del curso.

Dentro de la experiencia del ABP los alumnos van integrando una metodología propia para la adquisición de conocimiento y aprenden sobre su propio proceso de aprendizaje. Los conocimientos son introducidos en directa relación con el problema y no de manera aislada o fragmentada. En el ABP los alumnos pueden observar su avance en el desarrollo de conocimientos y habilidades, tomando conciencia de su propio desarrollo. Cabe destacar que la tecnología permite organizar y facilitar la comunicación con la clase, particularmente con los grupos. Las aulas virtuales en las plataformas educativas incluyen, en los espacios deliberativos de las mismas, foros de discusión, chat, wikis, mail interno, y en los espacios administrativos: archivos, sitios, presentación, programa. Dado que el aprendizaje cooperativo es fundamental para ABP y los tutores son responsables de asegurar los canales de comunicación adecuados para grupos de trabajo, la satisfacción de este objetivo técnico ayuda a lograr el éxito en el aula de ABP. (Hmelo; Kolmos, 2004; Muñiz Solari, 2004; Vernon et al, 1993; Instituto Tecnológico de Monterrey)

En la conclusión de un problema, afirma Watson (2002) los grupos tienen que reportar sus hallazgos a la clase o preparar un producto escrito para la revisión del tutor. Al contar con diversos canales de comunicación con el tutor y con los compañeros se facilita la preparación de este informe. Una herramienta útil son las wikis ya que permiten la redacción conjunta de un informe, permitiendo que el tutor pueda ver el historial de participación. Los recursos en línea pueden apoyar un curso de ABP, proporcionando varios elementos: (a) los ingredientes para escribir los problemas, (b) la inspiración para el diseño del problema y (c) la información para la resolución de problemas. Los estudiantes se dirigen naturalmente a la web en busca de información para resolver los problemas en un curso de ABP. La evaluación de los recursos en línea de manera crítica y ejecutar búsquedas en la Web con eficacia son lecciones importantes que aprender de

los estudiantes, y los prepara para ser aprendices de por vida.

## Líneas de Investigación y Desarrollo

Los objetivos del proyecto se orientan a analizar el grado de aceptación que tiene por parte de los alumnos de carreras técnicas el ABP y qué tipo de conductas y participaciones se detectan. Asimismo, se procura indagar las representaciones que tienen los docentes sobre que características tiene un problema adecuado para el ABP.

Las hipótesis de trabajo preliminares son las siguientes:

a) *El ABP es un proceso de aprendizaje centrado en el alumno, por lo anterior se espera de él una serie de conductas y participaciones distintas a las requeridas en el proceso de aprendizaje tradicional*

b) *Los docentes presentan distintas Representaciones sociales (RS) con respecto a qué se entiende por problema y cuáles son las estrategias que consideran adecuadas para encarar su resolución.*

## Resultados y Objetivos

La investigación se aborda en etapas con una mirada cuali-cuantitativa (Hernández Sampieri, 2009)

En la primera parte de la investigación se trabaja con una muestra de 12 profesionales de distintas disciplinas que en forma voluntaria accedieron a contestar las preguntas, a fin de indagar cuáles eran sus Representaciones sociales con respecto a qué se entiende por problema y cuáles son las estrategias que consideraba adecuadas para encarar su resolución. La muestra es no probabilística por conveniencia. Se realiza una entrevista que consta de las siguientes preguntas:

- 1) ¿Qué se entiende por problema?
- 2) ¿Qué tipos de problemas se pueden plantear en una clase?
- 3) ¿Para qué sirve resolver problemas?
- 4) ¿Cómo se puede enseñar a resolver problemas?

En una primera instancia se realizó una lluvia de ideas a fin de determinar las

variables con mayor frecuencia en cada una de las preguntas. Una vez seleccionadas estas variables se aplicó un Análisis de Componentes Principales (ACP). El ACP es una técnica proveniente del análisis exploratorio de datos cuyo objetivo es la síntesis de la información, o reducción de la dimensión (número de variables). En el gráfico simétrico (Ver Gráfico 1) correspondiente a la pregunta ¿Qué se entiende por problema?, se observan claramente dos subgrupos, por un lado los que destacan que un problema vincula conocimiento, resultados y resolver y por el otro los que consideran decisiones, situaciones y herramientas

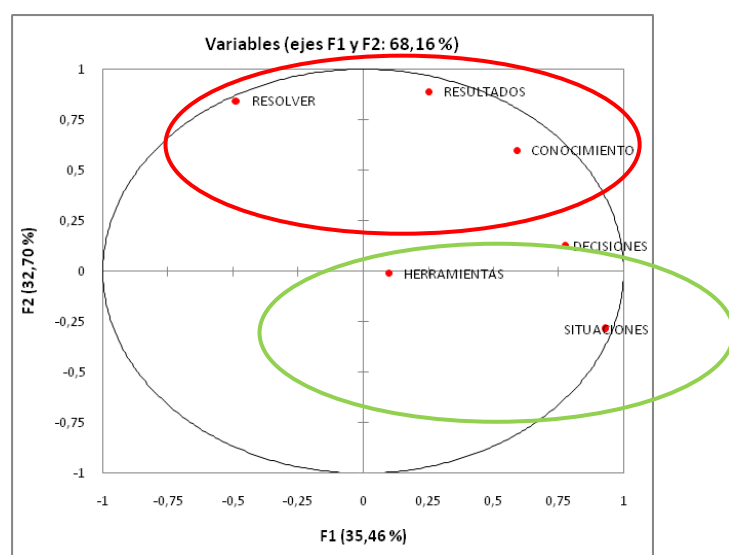


Gráfico 1: Gráfico simétrico

Por otra parte al asociar las variables con los participantes tal como se observa en el Biplot (Ver Gráfico 2) se puede observar que los entrevistados que son docentes e investigadores (P3, P5, P12) se encuentran en el mismo factor.





HERNANDEZ SAMPIERI, R (2009)  
Metodología de la investigación. McGraw-Hill

HMELO,C. The Effect of Problem-based Learning on the Early Development of Medical Expertise En:  
<http://www.cc.gatech.edu/aimosaic/edutech/people/PostDocs/Pubs/hmeloAERA.html>

KOLMOS,A.(2004) Estrategias para desarrollar currículos basados en la formulación de problemas y organizados en base a proyectos. *Educar*33.pp77-96

MUÑIZ SOLARI, O. 2004. Aprendizaje basado en problema (PBL): Beneficios y riesgos. *GeoTrópico*, 2 (2), 51-60, versión pdf En:  
[http://www.geotropico.org/2\\_2\\_Muniz-Solari.pdf](http://www.geotropico.org/2_2_Muniz-Solari.pdf) véase p 52

Instituto Tecnológico de Monterrey El aprendizaje basado en problemas como técnica didáctica  
<http://www.ub.edu/mercanti/abp.pdf>

Instituto Tecnológico de Monterrey. Las Técnicas didácticas en el modelo educativo

ROLDAN BORASSI, M. I. (2012). Introducción al Aprendizaje Basado en Problemas. Curso de Extensión. Facultad de Ciencias Exactas y Naturales. Universidad Nacional de Mar del Plata.

VERNON, D and BLAKE, RL.(1993) Does problem based learning works. A meta-analysis of evaluative research. *Acad Med* 68: 550-63.

WATSON,G (2002) Using Technology to Promote Success in PBL Courses. En :  
[http://technologysource.org/article/using\\_technology\\_to\\_promote\\_success\\_in\\_pbl\\_courses/](http://technologysource.org/article/using_technology_to_promote_success_in_pbl_courses/)

# Aportes para la enseñanza de circuitos RL en corriente continua: análisis de un applet

Claudia Zang<sup>1</sup>, Norah Giacosa<sup>1</sup>, Silvia Giorgi<sup>2</sup>, Jorge Maidana<sup>1</sup>, Alejandro Such<sup>1</sup>

<sup>1</sup>Departamento de Física/Facultad de Ciencias Exactas, Químicas y Naturales/Universidad Nacional de Misiones

Félix de Azara 1552. N3300LQ. Posadas. Misiones. Argentina

Tel/Fax: 54 376 4425414

e-mail: [claudiamzang@gmail.com](mailto:claudiamzang@gmail.com); [norah@correo.unam.edu.ar](mailto:norah@correo.unam.edu.ar); [jamaigms@gmail.com](mailto:jamaigms@gmail.com); [fiale.s@hotmail.com](mailto:fiale.s@hotmail.com)

<sup>2</sup>Facultad de Ingeniería Química/Universidad Nacional del Litoral

Santiago del Estero 2829. S3000AOM. Santa Fe. Argentina

e-mail: [sgiorgi@fiq.unl.edu.ar](mailto:sgiorgi@fiq.unl.edu.ar)

## Resumen

Un número considerable de estudiantes universitarios tiene dificultades para conceptualizar los fenómenos transitorios en circuitos inductivos en corriente continua. En particular, para relacionar las magnitudes físicas que intervienen y las funciones matemáticas que las representan. Se sostiene que algunas de estas dificultades, podrían estar relacionadas con el tratamiento dado al tema en los libros de texto que consultan.

Los resultados de otro trabajo de investigación, que abordó el análisis de una muestra de textos universitarios, indican que el tratamiento dado al tema en los mismos, no facilitaría la construcción de conocimientos, porque, entre otras cosas, se realiza una presentación dominada por un instrumentalismo matemático con baja conceptualización.

Existen otros recursos didácticos, acordes con las tendencias contemporáneas de la Educación Superior, como las Tecnologías de la Información y de la Comunicación, que pueden integrarse curricularmente para promover el aprendizaje deseado.

Por lo mencionado precedentemente, a partir del interés de los autores en la búsqueda de recursos didácticos para mejorar la comprensión del comportamiento de circuitos RL en corriente continua por parte de los estudiantes, se seleccionó y analizó un *applet* de uso libre con el cual se puede complementar el estudio del tema.

Los resultados que se presentan señalan cuestiones de aptitud para su uso en el aula como así también algunas limitaciones sobre las cuales se realizan advertencias a tener en cuenta antes de su utilización.

**Palabras claves:** enseñanza, Universidad, fenómenos transitorios en circuitos resistivo-inductivos, *applet*.

## Contexto

Este estudio se realizó en el marco del proyecto denominado: “El aprendizaje de la Física universitaria: un estudio de las relaciones entre los modelos empleados en la enseñanza y las representaciones de los estudiantes” (16Q479), registrado en la Secretaría de Posgrado de la Facultad de Ciencias Exactas,

Químicas y Naturales (FCEQyN) dependiente de la Universidad Nacional de Misiones (UNaM). Se trata de un proyecto trianual (2011-2013).

## Introducción

Uno de los bloques temáticos de Física, incluido en la mayoría de las carreras científico-tecnológicas de la República Argentina, es Electricidad. En él se estudian, entre otras cuestiones, el comportamiento de los circuitos resistivo-inductivos en corriente continua.

La experiencia docente de los autores muestra que para un número considerable de alumnos, el estudio de fenómenos transitorios no suele ser una tarea sencilla. En general, los estudiantes tienen dificultades para explicar el proceso de aumento y decaimiento de la corriente eléctrica en circuitos de esta naturaleza; y para vincular, por un lado, las diferentes magnitudes físicas que intervienen (corriente eléctrica, diferencia de potencial, tiempo) y por otro, las formas en que pueden representarse las funciones matemáticas que describen las relaciones entre tales magnitudes.

Si bien son múltiples los factores que influyen en el aprendizaje de Física por parte de los estudiantes, los libros de texto que recomiendan los docentes y que utilizan habitualmente los alumnos, desempeñan un papel preponderante en el mencionado proceso.

En este contexto, y reconociendo la importancia que tienen los libros de texto de uso habitual en la enseñanza, se realizó una investigación previa, con el propósito de indagar la perspectiva con la que se aborda en ellos el tema circuitos RL.

Los resultados, del análisis de doce libros de texto universitarios, muestran que el tratamiento que se hace del tema en la mayoría de los ellos no facilitaría el aprendizaje de los fenómenos transitorios en circuitos inductivos. Por un lado, se realiza una presentación dominada por el instrumentalismo matemático;

por otro, se presentan ecuaciones temporales que no se grafican y gráficos cuyas ecuaciones no se explicitan, lo cual dificultaría la comprensión lectora. Además, los problemas resueltos son mayoritariamente cuantitativos y carentes de interés científico. Las imágenes, analogías, referencias históricas y aplicaciones a la vida cotidiana son escasas.

Los libros de texto, desde las restricciones que la palabra escrita y las ilustraciones estáticas imponen, seguirán siendo referentes para guiar la enseñanza de Física, pero existen otros recursos didácticos, acordes con las tendencias contemporáneas de la Educación Superior, como las Tecnologías de la Información y de la Comunicación (TIC), que pueden integrarse curricularmente para promover el aprendizaje deseado [1].

En el ámbito de las Ciencias Físicas, las TIC han mostrado ser un medio apropiado para desarrollar algunas de las competencias profesionales que la sociedad del conocimiento exige [2]. Si bien sus usos son variados: adquisición y procesamiento de datos, obtención de gráficos en tiempos reales, realización de experiencias en entornos virtuales y reales por medio del uso de laboratorios remotos, siendo esta última modalidad de desarrollo e implementación más acotada y reciente [3, 4, 5], las simulaciones computacionales (*applet*) son los recursos educativos más abundantes y accesibles. No obstante, cabe remarcar que algunas simulaciones de acceso libre contienen serios errores.

Se sostiene, concordando con otros investigadores [6] que enseñar a los estudiantes que los libros de textos pueden contener errores e imprecisiones, favorece el desarrollo de una actitud más crítica hacia las deficiencias detectadas e incide en sus concepciones epistemológicas sobre los criterios de validez del conocimiento científico. Desde esta postura, y extendiendo esta afirmación a las simulaciones, en un trabajo anterior [7] se



diseñaron y desarrollaron actividades con un simulador que incurre en fallas operatorias similares a las que es posible reconocer en las resoluciones de problemas lápiz y papel que realizan los estudiantes. Los resultados mostraron que, ante el uso de dos simuladores seleccionados por los docentes, uno de los cuales arrojaba valores equivocados, algunos estudiantes universitarios fueron incapaces de reconocer los resultados erróneos aún luego de haber abordado el tema en las clases correspondientes.

En virtud de lo mencionado, y en el afán de contribuir a mejorar la enseñanza y el aprendizaje de los fenómenos transitorios en circuitos RL, en esta presentación se describen los criterios de selección y el análisis de un *applet* de acceso libre.

## Líneas de investigación y desarrollo

Este trabajo se llevó a cabo en el marco del proyecto de investigación mencionado, el cual se formuló en el contexto de indagación sobre las representaciones mentales de los estudiantes universitarios del ciclo básico relacionadas con determinados tópicos de Física, en este caso el estudio de circuitos RL, y sobre recursos de enseñanza potencialmente útiles para acercar dichas representaciones a los modelos físicos actuales consensuados por la comunidad científica, de manera de hacer más fluidos los procesos de enseñanza y aprendizaje.

Se espera que los resultados obtenidos en el desarrollo del citado proyecto puedan colaborar con el mejoramiento y la necesaria actualización que la enseñanza de la Física demanda; y promover otras investigaciones en esta dirección.

## Resultados y Objetivos

El propósito de este escrito es mostrar resultados preliminares, derivados de un estudio de caso, correspondientes al análisis crítico de

un *applet* con el cual se puede complementar la enseñanza de fenómenos transitorios en circuitos RL en corriente continua en el ciclo universitario básico.

El *applet* que se presenta se escogió en base a categorías que se construyeron, en concordancia con otros estudios, a partir de los siguientes aspectos: criterios para seleccionar información libre de Internet, orientaciones para elegir *software* educativo y recaudos necesarios para la selección de materiales curriculares [8, 9].

El *applet* seleccionado pertenece al curso “Learn Physic using Java” desarrollado por Chiu-king Ng (C. K. Ng). El autor es PhD en Física por la Universidad China de Hong Kong y actualmente se desempeña como docente de Física en el *Yenching Collage* (Hong Kong, República Popular de China). El curso está organizado en tres bloques temáticos: 1) Mecánica, 2) Luz y Ondas y 3) Electromagnetismo y Electrónica; y cuenta con cerca de 50 simuladores. La última actualización, según se indica en la página WEB, data de enero de 2012.

Ingresando al link: [http://ngsir.netfirms.com/englishhtm/RL\\_squar\\_e.htm](http://ngsir.netfirms.com/englishhtm/RL_squar_e.htm) se accede al *applet* elegido, disponible en inglés y chino. El título del simulador, cuya pantalla se muestra en la Figura 1, es: “Square Wave applied to RL in series (Continuous charging and discharging)”.

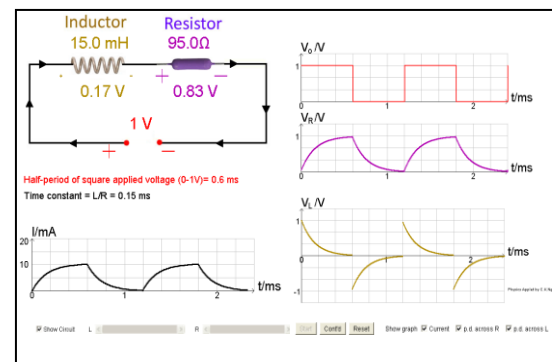


Fig. 1: Pantalla simulador circuito RL C.K. Ng

El objetivo del simulador es estudiar el comportamiento de un circuito RL en serie conectado a un generador de señales cuadradas, en los procesos de crecimiento y decaimiento de la corriente eléctrica en función del tiempo

El semiperíodo de la señal de la fuerza electromotriz (FEM) es constante e igual a 0,6 milisegundos. El valor máximo de la FEM durante el primer semiperíodo, o los consecutivos posteriores impares, es constante e igual a 1 V. En el segundo semiperíodo, o posteriores pares, la FEM se anula. Estas condiciones están explicitadas en el simulador (“*Half-period of square applied voltage (0-1V) = 0.6 ms*”).

En la parte inferior de la pantalla se encuentran los controles de entrada. Ellos son: mostrar animación en circuito (*Show circuit*), asignar valores a  $L$  y  $R$  dentro de un rango establecido por el programa, iniciar la animación (*Start*), pausar o reiniciar la simulación (*Pause-Cont'd*), reajustar valores (*Reset*); y mostrar los gráficos de: corriente, diferencia de potencial en la resistencia y diferencia de potencial en el inductor (*Show graph: Current, p.d. across R, p.d. across L*).

En la Tabla 1 se muestran los valores máximos y mínimos que admiten las dos únicas variables del programa.

Variables	Valor mínimo	Valor máximo
L	5 mL	100 mH
R	50 $\Omega$	100 $\Omega$

**Tabla 1:** Valores mínimos y máximos de  $L$  y  $R$  establecidos en el programa

Una vez seleccionados los valores de  $L$  y  $R$ , lo cual se logra moviendo el cursor destinado a ese efecto, el programa proporciona automáticamente la constante de tiempo inductiva del circuito. Señalando todas las posibles prestaciones del simulador e iniciándolo es posible ver en el circuito la animación del sentido convencional de

corriente de conducción y la evolución de los gráficos mencionados.

Entre las cuestiones de aptitud para su uso en el aula se señalan:

- Utiliza distintos colores para representar los elementos del circuito (resistencia, inductor, fuente de FEM), sus valores y la diferencia de potencial en sus terminales.

- Grafica la corriente, en el proceso de crecimiento y de decaimiento, en función del tiempo en una misma gráfica.

- Muestra la evolución temporal de la diferencia de potencial en las terminales de la fuente de FEM, en la resistencia y en el inductor, en el proceso de crecimiento y de decaimiento de la corriente eléctrica, en tres gráficos separados utilizando los mismos colores que los representan en el esquema del circuito eléctrico.

- Permite plantear actividades cuantitativas acordes al nivel del curso en el que se pretende introducir.

- Proporciona a los estudiantes oportunidades para desarrollar su propia comprensión acerca de los conceptos involucrados en los procesos de crecimiento y decaimiento de la corriente eléctrica en un circuito de esta naturaleza, a través de un proceso de construcción de hipótesis y de prueba de ideas.

- La pantalla se puede transportar a otros programas informáticos lo cual facilita la preservación de la información y la revisión de los procesos de experimentación.

Algunas limitaciones son:

- El espacio del plano que ocupa el simulador, si no se ajusta convenientemente la resolución de la pantalla de la computadora, no se puede ver completo.

- Dado que la FEM es un generador de señales cuadradas, los procesos de crecimiento y decaimiento de la corriente eléctrica se estudian de manera integrada y la decodificación de todos los gráficos mostrados requiere de un trabajo cuidadoso de interpretación.

- En el título del *applet* se hace referencia a “carga” y “descarga” cuando en ningún

elemento de este tipo de circuitos se acumulan o desacumulan cargas eléctricas.

Se sostiene que la incorporación de las TIC en ambientes educativos requiere indefectiblemente de nuevas competencias docentes que sólo se logran, en primera instancia, a través de la capacitación y de una profunda reflexión acerca de su potencial utilidad para promover procesos de construcción de conocimientos científicos; y en segundo lugar con un uso responsable que posibilite ser considerada por los estudiantes como una herramienta más para aprender.

La mediación entre los recursos utilizados, ya sean éstos libros de texto u otros materiales, depende en última instancia del docente.

Saber analizar críticamente los recursos disponibles antes de seleccionarlos para actividades curriculares, en una época caracterizada por la sobreabundancia de información, es una tarea ineludible. Se espera con este trabajo realizar un aporte en esa dirección.

## Formación de Recursos Humanos

El equipo está conformado por docentes del Dpto. de Física de la FCEQyN (UNaM) y alumnos de la Institución. La co-dirección está a cargo de una docente del Dpto. de Física de la Facultad de Ingeniería Química dependiente de la Universidad Nacional de Litoral, con la que se ha compartido trabajo por más de una década. Dos de las integrantes están cursando la Maestría en Enseñanza de las Ciencias Exactas y Naturales (Facultad de Ingeniería, Universidad Nacional del Comahue); y una, la Especialización en Investigación Educativa (Convenio FCEQyN-UNaM con ISFD dependientes del Ministerio de Cultura, Educación, Ciencia y Tecnología de Misiones).

Algunos integrantes del proyecto utilizan regularmente *applets* en cursos de Física, han coordinado un Taller de capacitación docente

en Uruguay durante el año 2012; y diseñado e implementado una Asignatura Optativa destinada a alumnos avanzados del Profesorado en Física (FCEQyN-UNaM) utilizando 34 simuladores de Electromagnetismo y Óptica a partir del segundo semestre de 2012.

## Referencias

- [1] Lapp, D. & Cyrus, V. (2000). Using data-collection devices to enhance students' understanding. *Mathematics Teacher*, 93(6), 504-510.
- [2] Serrano, G. y Fuentes S. (2011) El uso de un simulador virtual para la enseñanza del movimiento de proyectiles. *Encuentro Latinoamericano de Profesores y Estudiantes de Matemática y Ciencias Naturales*, San Rafael, Mendoza. Argentina.
- [3] Llonch, E. y Massa, M. (2006) La hipermedia en el aprendizaje del átomo a la construcción de la materia. *Memorias de CIAEF 2006 – IACPE*. Universidad de Costa Rica.
- [4] Kofman, H. & Concari, S. (2011) Using remote labs for Physics teaching. In: *Using Remote Labs in Education*. Javier García Zúbia and Gustavo Alves (Eds.), University of Deusto Publications, 293-308.
- [5] Pires, M. e Veit, E. (2006) Tecnologías de Informação e Comunicação para ampliar e motivar o aprendizado de Física no Ensino Médio. *Revista Brasileira de Ensino de Física*. V28 N2, 241-248.
- [6] Campanario, J. (2003) De la necesidad a la virtud: cómo aprovechar los errores e imprecisiones de los libros de texto para enseñar Física. *Revista Enseñanza de la Ciencias*. (21), 1, 161-172.
- [7] Giacosa, N.; Giorgi, S.; Maidana, J.; Beck, S. y Zang, C. (2011) Controversias entre modelos físicos y simulaciones computacionales en la clase de electrostática: estudio del funcionamiento de un electroscopio cargado. *Memorias de la REF XVII*. Universidad Nacional de Córdoba. T (70), 1-12.
- [8] Giacosa, N., Zang, C., y Giorgi, S. (2012) TIC en la enseñanza y el aprendizaje de electrostática. *XIV Workshop de Investigadores en Ciencias de la Computación*, 957-961.
- [9] Caamaño, A. (Coord.) (2011) *Didáctica de la Física y la Química*. Barcelona. GRAÓ.

## Actualidad de la capacitación en Inteligencia de Negocios brindada por organizaciones latinoamericanas

H. D. Kuna<sup>1</sup>, E. Zamudio<sup>1</sup>, F. J. Yatchesen<sup>1</sup>, F. Guerrero<sup>1</sup>, M. Lozina<sup>1</sup>, A. Rambo<sup>1</sup>, S. Caballero<sup>1</sup>, M. Marinelli<sup>1</sup>, H. D. Refosco<sup>2</sup>, G. Sequeira<sup>1</sup>, M. Rey<sup>1</sup>, G. C. Escalante<sup>1</sup>

<sup>1</sup> Facultad de Ciencias Exactas, Químicas, y Naturales, Universidad Nacional de Misiones, Posadas, Misiones, Argentina

<sup>2</sup> Facultad de Ingeniería, Universidad Nacional de Misiones, Oberá, Misiones, Argentina

hdkuna@gmail.com

### Resumen

En la actualidad no se ha identificado un programa de capacitación que defina con claridad las estrategias didácticas y pedagógicas adecuadas para la formación de usuarios finales de soluciones de Inteligencia de Negocios (BI) en el ámbito de organizaciones de la provincia de Misiones (Argentina). Este trabajo presenta un relevamiento de las ofertas de capacitación en BI en Latinoamérica, e identifica un conjunto de características de las alternativas de capacitación ofrecidas, con el objetivo de permitir la generación de planes de capacitación adecuados para la realidad de las organizaciones regionales.

**Palabras clave:** Educación, Capacitación, Inteligencia de Negocios Usuarios Finales.

### Contexto

Este trabajo se desarrolló en el contexto del proyecto de investigación "Modelos de capacitación para usuarios finales en el ámbito de Inteligencia de Negocios en la provincia de Misiones", Facultad de Ciencias Exactas, Químicas, y Naturales, Universidad Nacional de Misiones, cuyo objetivo consiste en desarrollar estrategias didácticas y pedagógicas dirigidas a la capacitación de usuarios finales de soluciones en Inteligencia de Negocios en el ámbito de la provincia de Misiones, Argentina. Con este propósito, resultó necesario identificar la oferta de

capacitación en el área mencionada tanto en el ámbito de Argentina, países vecinos, y otros países de Latinoamérica.

### Introducción

Actualmente no se encuentra información sistematizada en relación con la oferta de capacitación en el área de BI. Esta información resulta necesaria para la elaboración de alternativas de formación dirigida a usuarios finales. En este escenario, se optó por la realización de un relevamiento de las características presentadas por las capacitaciones en el ámbito de BI ofrecidas por organizaciones latinoamericanas, identificando los recursos utilizados y las estrategias didácticas y pedagógicas adoptadas por quienes brindan la capacitación.

Se realizó un relevamiento en internet de las organizaciones dedicadas a brindar servicios de formación en el área de BI para determinar: el tipo de capacitación; el público al que están dirigidos; las áreas de BI en las cuales ofrecen educación; la duración de los cursos; los requisitos de formación previa; el idioma en que se imparte; y el tipo de organización que desarrolla el entrenamiento.

Para el estudio se clasificaron las organizaciones en dos grupos. Por una parte, aquellas orientadas a dar consultoría y/o capacitación, y por otra aquellos vendedores de herramientas, los cuales ofrecen capacitación de sus productos. No han sido



considerados para este estudio carreras de educación de posgrado en relación a BI, específicamente especializaciones y maestrías.

Para el relevamiento de organizaciones dedicadas a capacitación en BI se utilizaron los motores de búsqueda y metabuscadores: Google [2], Altavista [3], Mahalo [4], Yippy [5], Ask [6], Lycos [7], Baidu [8], Excite [9], Hispavista [10], Terra [11], y Hotbot [12]. Y se ingresaron conjuntamente los términos de búsqueda *capacitación* e *inteligencia de negocios* en idioma español, y sus traducciones en inglés *training* y *business intelligence*. Las búsquedas se refinaron mediante las herramientas brindadas por cada motor de búsqueda. Se tomaron los primeros 300 resultados de cada buscador y se realizó una revisión manual de los mismos con el fin de determinar si correspondían al tipo de organizaciones relevantes para el estudio.

Como resultado de la búsqueda, se identificaron 71 organizaciones dedicadas a dar consultoría y capacitación en idioma español, y 10 vendedores de herramientas.

A modo de homogeneizar las clasificaciones, se han definido perfiles para cada criterio de análisis.

De acuerdo al **público** al cual está dirigido el curso: Profesional en Tecnologías de Información (IT), Analista, Estudiante.

De acuerdo a la **ubicación geográfica**: Localidades de Argentina, Países extranjeros.

De acuerdo a las **áreas** que capacitan: Conceptual, Datawarehouse y datamarts, Reporting, Procesamiento Analítico En Línea (OLAP), Técnicas y algoritmos de descubrimiento de conocimiento (KDD), Extracción Transformación y Carga (ETL), Gestión de proyectos.

De acuerdo a la **duración**: horas correspondientes a horas reloj, u horas académicas, según se indique.

De acuerdo a las **estrategias** didácticas y pedagógicas: En línea, Presencial, Teórico/Práctico, Flexibilidad del cronograma, Otras metodologías propias de

las organizaciones como B-Learning [13] o Mentored Learning [14].

De acuerdo al **tipo de organizaciones**, es decir, su actividad principal: Capacitación IT, Consultoría IT, Consultoría BI, Consultoría de negocios, Universidad, Fundación sin fines de lucro.

De acuerdo al **idioma** en que se imparte el curso: Español, Inglés.

De acuerdo a las **herramientas** utilizadas para la capacitación: Comerciales (SAP [15], Microsoft [16], O3 [17], IBM [18], Oracle [19], SAS [20], Microstrategy [16].

Comerciales con versión **open source**: Pentaho [21], Jaspersoft [11], QlikView [23], Palo [24].

De acuerdo al **costo** de la capacitación: Se presentan los costos en dólares estadounidenses. Se convirtieron aquellos importes expresados en moneda local. En el caso de presentar varios cursos, se promediaron los importes.

### Líneas de investigación y desarrollo

La investigación se ha enfocado en el relevamiento de las opciones de capacitación en herramientas de BI con el fin de utilizar esta información para el desarrollo de planes de capacitación adecuados a la realidad particular de los posibles usuarios finales de soluciones BI que integran dichas organizaciones.

### Resultados y objetivos

Se pudo determinar que del total de organizaciones relevadas, en un 25% de los casos los cursos determinan como público objetivo tanto al perfil de Profesionales IT, como a Analistas de Negocio. Sin embargo es destacable que un 17% de los cursos están dirigidos exclusivamente a Analistas de Negocio, superando a aquellos dirigidos exclusivamente a Profesionales IT.

La ubicación regional de los centros de capacitación arrojó resultados de Argentina donde se identificaron organizaciones en las ciudades de Buenos Aires, Córdoba, y Puerto Madryn, sin embargo, no se identificó

ninguna organización en la provincia de Misiones, ni en provincias cercanas como Corrientes, Chaco, o Formosa.

La mayor cantidad de cursos relevados estaban dirigidos al aprendizaje de herramientas para el Analista de Negocio, particularmente OLAP (26), Reporting (18), y Dashboards (17). Sin embargo es destacable la cantidad de organizaciones dedicadas a la formación de aspectos conceptuales (15), y de construcción y mantenimiento del datawarehouse (DW). Por otra parte, es importante el número de cursos que no especifican el contenido de los planes de capacitación (27).

Tan sólo la mitad de los cursos indican cuál es la duración de los mismos, aunque aquellos que sí lo informan presentan una gran amplitud en este aspecto, ya que inician desde las 3hs, hasta las 200hs de duración, lo que podría equivaler a una jornada o a todo un semestre de cursado, respectivamente.

Respecto de la estrategia utilizada en el dictado de los cursos, y cómo el capacitado accede a los materiales de los mismos, se plantean diversas alternativas, aunque se destaca una integración entre varias opciones, como por ejemplo: clases presenciales con instructores, combinadas con accesos a plataformas educativas para acceder a los materiales o herramientas de discusión y evaluación. En algunos casos, se presentan alternativas de capacitación donde se traslada el entorno de capacitación a la compañía para aprender directamente con los datos de la organización del cliente.

Otra tendencia es la capacitación con horarios flexibles, donde el capacitado es quien determina en qué momento puede acceder a los recursos educativos.

Un 65% de las organizaciones no indican formalmente la metodología ni las estrategias educativas con las que trabajan.

Un 74% de las organizaciones que se dedican a la capacitación en el área de BI corresponde también al sector dedicado al desarrollo de soluciones IT. El resto de las organizaciones se agrupan en consultoras específicas de

capacitación IT (7%), desarrollo de soluciones BI (6%), y consultoría en temas de negocios (6%). Además, un 4% corresponde a universidades, como la Universidad de Chile [25] que desde la Facultad de Economía y Negocios ofrece un curso de capacitación específico en BI con una duración de 120hs.

También se encontraron fundaciones sin fines de lucro, como la Fundación Egresados U. Distrital (FEUD) [26], la cual es una organización formada por egresados de la universidad colombiana, la cual ofrece un curso de capacitación en BI.

En el caso de las herramientas utilizadas para la capacitación y/o certificación, se destaca la preferencia de herramientas comerciales (55%). En algunos casos se ofrece como alternativa la capacitación con herramientas Open Source, y en otros directamente la capacitación es con este tipo de herramientas. Por otra parte, en un 48% de los casos no se especifica la plataforma con la que se trabaja.

Finalmente, respecto del costo de los cursos se observa una gran amplitud, lo cual está relacionado con la duración de los mismos, más que con las herramientas utilizadas. Únicamente el 17% de los cursos publican información del costo de la capacitación. En aquellos casos donde no están publicados, se solicita que se contacte con un representante de la organización para obtener detalles del mismo.

Los vendedores de las herramientas más utilizadas por las organizaciones de capacitación ofrecen a su vez planes de capacitación oficiales, donde, en algunos casos son implementados por socios comerciales, los cuales suelen ser organizaciones que desarrollan soluciones con dichas herramientas.

Los programas de capacitación oficial ofrecen en general sistemas de rutas de aprendizaje, en los cuales establecen niveles de capacitación y requisitos para poder alcanzar un determinado objetivo, como ser un examen de certificación.

El relevamiento de las ofertas de capacitación en el área de BI a nivel regional permitió

identificar las características de las alternativas de formación ofrecidas por diversas organizaciones vinculadas a BI. De esta manera fue posible determinar el tipo de organizaciones encargadas de la capacitación; las herramientas utilizadas; los aranceles; y el perfil del público objetivo de las capacitaciones.

Por otra parte, se consiguió establecer las alternativas de capacitación ofrecidas por los vendedores de las herramientas tanto comerciales como Open Source, y las alternativas de capacitación brindadas por cada uno.

Resulta necesario fortalecer el rol de las universidades en relación al desarrollo de políticas de capacitación formales y permanentes en el área de BI dirigido a gestores tanto de empresas privadas como de organismos públicos. Estas políticas deben tener en cuenta alternativas Open Source en el contenido de los planes de formación, como también en las herramientas utilizadas, con objeto de garantizar la continuidad y mejoramiento de dichas alternativas tanto en la educación formal en particular, como también en la transferencia de conocimiento en general.

Trabajos futuros dirigidos a asociar los datos obtenidos en esta investigación junto con un relevamiento de las características de las organizaciones locales, permitirán el desarrollo de planes de capacitación adecuados a la realidad particular de los posibles usuarios finales de soluciones BI que integran dichas organizaciones.

#### Formación de Recursos Humanos

En el marco de este proyecto se conformó un equipo de investigación dentro del "Programa de Investigación en Computación" con 12 integrantes (alumnos, egresados y docentes) de la Facultad de Ciencias Exactas Químicas y Naturales de la Universidad Nacional de Misiones, y 1 docente de la Facultad de Ingeniería de la misma institución. De ellos, 3 egresados realizaron su trabajos finales de grado en áreas relacionadas a Descubrimiento del Conocimiento e Inteligencia de Negocio, 3 alumnos se encuentran en proceso de

finalización del trabajo final de grado, y 1 se encuentra iniciando el trabajo final de grado, 1 de los recientes egresados se encuentra cursando una maestría, y 2 de ellos un doctorado.

#### Referencias

- [1] H. D. Kuna, S. Caballero, M. Marinelli, A. Rambo, H. D. Refosco, G. Sequeira, E. Zamudio, F. Guerrero, M. Lozina, M. Rey, F. J. Yatchesen, y G. C. Escalante, «Modelos de capacitación para usuarios finales en el ámbito de Inteligencia de Negocios en la provincia de Misiones». Facultad de Ciencias Exactas, Químicas, y Naturales. Universidad Nacional de Misiones, 2012.
- [2] «Google». [Online]. Available: <http://www.google.com.ar/>. [Accessed: 04-sep-2012].
- [3] «Yahoo! Search - Web Search». [Online]. Available: <http://www.altavista.com/>. [Accessed: 15-oct-2012].
- [4] «Mahalo.com». [Online]. Available: <http://www.mahalo.com/>. [Accessed: 17-oct-2012].
- [5] «Yippy.com - Yippy Search Engine». [Online]. Available: <http://www.yippy.com/>. [Accessed: 17-oct-2012].
- [6] «Ask.com - What's Your Question?». [Online]. Available: <http://www.ask.com/>. [Accessed: 17-oct-2012].
- [7] «Lycos». [Online]. Available: <http://www.lycos.com/>. [Accessed: 17-oct-2012].
- [8] «Baidu». [Online]. Available: <http://www.baidu.com/>. [Accessed: 17-oct-2012].
- [9] «My Excite». [Online]. Available: <http://www.excite.com/>. [Accessed: 17-oct-2012].
- [10] «:: HispaVista: Portal y buscador de Internet. Correo gratis ::» [Online]. Available: <http://www.hispavista.com/>. [Accessed: 17-oct-2012].
- [11] «Terra USA - Noticias, Deportes, Entretenimiento y Estilo de Vida». [Online]. Available:

- <http://www.terra.com/portada/>.  
[Accessed: 17-oct-2012].
- [12] «HotBot». [Online]. Available: <http://www.hotbot.com/>. [Accessed: 17-oct-2012].
- [13] «Capacitación SAP Online - Modalidad B-learning - CVOSOFT IT ACADEMY Centro de formación en tecnología SAP». [Online]. Available: [http://www.cvosoft.com/sistemas\\_sap\\_abap/consultoria\\_sap\\_capacitacion\\_profesional\\_blearning.php#desarrollo](http://www.cvosoft.com/sistemas_sap_abap/consultoria_sap_capacitacion_profesional_blearning.php#desarrollo). [Accessed: 02-sep-2012].
- [14] «Flexible, targeted learning solution - Mentored Learning® - New Horizons Computer Learning Centers». [Online]. Available: <http://www.newhorizons.com/Mentored-Learning.aspx>. [Accessed: 02-sep-2012].
- [15] «Courses & Curricula | SAP BusinessObjects Business Intelligence». [Online]. Available: <https://training.sap.com/us/en/courses-and-curricula/business-objects-business-intelligence>. [Accessed: 17-ago-2012].
- [16] «MicroStrategy Education Solutions». [Online]. Available: <http://www.microstrategy.com/training-events/education/solutions/#2>. [Accessed: 17-ago-2012].
- [17] «Business intelligence Uruguay, BI, Cuadro de mando Integral, Servicios de consultoria Ideasoft | Capacitación BI». [Online]. Available: [http://www.ideasoft.biz/web/?page\\_id=1239](http://www.ideasoft.biz/web/?page_id=1239). [Accessed: 17-ago-2012].
- [18] «IBM - Cognos Business Intelligence: Training and certification». [Online]. Available: <http://www-01.ibm.com/software/analytics/cognos/business-intelligence/training-and-certification.html>. [Accessed: 17-ago-2012].
- [19] «Oracle University Select country». [Online]. Available: <http://www.oracle.com/us/education/selectcountry-new-079003.html>. [Accessed: 20-ago-2012].
- [20] «SAS Training - Business Intelligence». [Online]. Available: <http://support.sas.com/training/us/bi.html>. [Accessed: 17-ago-2012].
- [21] «We offer a range of classroom, online, and on-site training services to meet your Pentaho Business Analytics and BI needs». [Online]. Available: <http://www.pentaho.com/services/training/>. [Accessed: 17-ago-2012].
- [22] «Jaspersoft University Offerings :: Jaspersoft Business Intelligence Software». [Online]. Available: <http://www.jaspersoft.com/training>. [Accessed: 17-ago-2012].
- [23] «QlikView Training - Business Intelligence (BI) Reporting Software Solutions | QlikView». [Online]. Available: <http://www.qlikview.com/es/services/training>. [Accessed: 17-ago-2012].
- [24] «Jedox AG - Academy». [Online]. Available: <http://www.jedox.com/en/academy/>. [Accessed: 17-ago-2012].
- [25] «Diplomado en Business Intelligence». [Online]. Available: [http://www.uejecutivos.cl/index.php?option=com\\_content&view=article&id=160:diplomado-en-business-intelligence-2012&catid=15:tecnologias-de-informacion&Itemid=109](http://www.uejecutivos.cl/index.php?option=com_content&view=article&id=160:diplomado-en-business-intelligence-2012&catid=15:tecnologias-de-informacion&Itemid=109). [Accessed: 24-jun-2012].
- [26] «Inteligencia de Negocios BI». [Online]. Available: <http://www.egresadosudistrital.edu.co/index.php/capacitaciones/diplomados/inteligencia-de-negocios-bi>. [Accessed: 20-ago-2012].
- [27] «Resources and Tools for IT Professionals | TechNet». [Online]. Available: <http://technet.microsoft.com/en-us/ms376608.aspx>. [Accessed: 17-oct-2012].
- [28] «MSDN Library». [Online]. Available: <http://msdn.microsoft.com/es-es/library/ms123401.aspx>. [Accessed: 17-oct-2012].



## Personalización en entornos de u-learning

**Beatriz Fernández Reuter<sup>2</sup>, Elena Durán<sup>1</sup>, Analía Amandi<sup>3</sup>**

<sup>1</sup> Departamento de Informática - Facultad de Ciencias Exactas y Tecnologías (FCEyT)  
Universidad Nacional de Santiago del Estero (UNSE), Santiago del Estero

<sup>2</sup> CONICET, Comisión Nacional de Investigaciones Científicas y Técnicas

<sup>3</sup> ISISTAN, Facultad de Ciencias Exactas, Universidad Nacional del Centro de la Provincia de Buenos Aires (UNCPBA), Tandil

e-mail: bfreuter@gmail.com; eduran@unse.edu.ar; amandi@exa.unicen.edu.ar

### RESUMEN

La personalización se considera una técnica muy potente para mejorar la eficacia de la búsqueda de información y la toma de decisiones. Esta ha permitido la amplia difusión de los sistemas capaces de sugerir información relevante y personalizada a los usuarios, de acuerdo a sus características y preferencias representadas en un Modelo de Usuario. Este modelo en el campo de la educación recibe el nombre de Modelo de Estudiante.

En los últimos años ha surgido un nuevo escenario educativo caracterizado por un conjunto de actividades formativas accesibles en cualquier lugar y desde cualquier dispositivo, el aprendizaje ubicuo (u-learning). Esta nueva forma de aprendizaje, por naturaleza, debe reflejar su capacidad para utilizar tecnologías de la información y la comunicación, tanto desde la perspectiva de las necesidades de los usuarios como de las capacidades que las nuevas tecnologías ofrecen para una formación ubicua.

Por ello, este trabajo propone el desarrollo de un modelo de estudiante que ofrezca la posibilidad de personalización de los servicios que se brindan, facilitando el seguimiento y la realización de las diferentes tareas, haciendo que el usuario se sienta acompañado, con métodos de seguimiento orientados en función de las demandas, y posibilitando los diálogos y la comunicación virtual.

**Palabras clave:** Personalización, aprendizaje ubicuo, agentes inteligentes, Modelo de usuario.

### CONTEXTO

Esta línea de investigación se desarrollará en el marco de una Beca de Postgrado Tipo I otorgada por el CONICET. Además, se inserta en el proyecto “Sistemas de Información Web Personalizados, Basados en Ontologías, para Soporte al Aprendizaje Ubicuo”, que a su vez es parte del Programa “Sistemas de Información Web basados en Tecnologías de Agentes”, correspondiente a la convocatoria 2012-2015 de la Secretaría de Ciencia y Tecnología de la Universidad Nacional de Santiago del Estero (SICYT - UNSE).

### INTRODUCCIÓN

El aprendizaje ubicuo se caracteriza por proporcionar formas intuitivas de identificar a los colaboradores, a los contenidos y a los servicios correctos en el lugar y horario indicado sobre la base del contexto que rodea a los estudiantes (Graf et al, 2009). La personalización juega entonces un papel muy importante en este tipo de aprendizaje, ya que permite proporcionar a los estudiantes el material, las actividades y la información personalizadas, en el lugar correcto y el momento adecuado.

Toda plataforma de aprendizaje personalizada requiere de la creación y mantenimiento de un modelo de estudiante, donde se registren los datos personales del alumno, sus progresos

académicos, sus estilos de aprendizaje, intereses, nivel de conocimientos, y habilidades para resolver problemas, entre otros datos. Sin embargo, para proporcionar una personalización eficiente en un ámbito ubicuo, el modelo del estudiante debe ser capaz de reunir una gran variedad de información acerca de los estudiantes, ya que por ejemplo, ante una consulta, habitualmente no será posible devolverle toda la información disponible en el sistema, debido a las limitaciones de comunicación, procesamiento y presentación de datos que padecen los dispositivos involucrados (Bobbillo et al, 2006). Además, algunos estudios han demostrado que hay herramientas de comunicación que son más adecuadas que otras según sea el tipo de tarea que quiera realizar un estudiante en un ambiente ubicuo (Seppälä y Alamäki, 2003). Algo similar ocurre en relación a sus posibilidades para encarar un trabajo colaborativo, las que seguramente dependerán no sólo de las características del grupo ni de sus habilidades de colaboración, sino de las posibilidades de conectividad social y de su ubicación.

En consecuencia, es preciso registrar en el modelo del alumno información adicional que cobra relevancia en los contextos ubicuos. El registro de toda esta información dará como resultado un modelo de estudiantes integral, que permitirá apoyar a los alumnos a través de una personalización de todos los servicios del entorno de aprendizaje, y a los docentes ofrecer una mejor comprensión del proceso de aprendizaje de sus estudiantes.

Si bien existen diferentes enfoques para llevar a cabo el proceso de modelado del estudiante (Self 1994; Brusilovsky, 1994; Graf et al, 2009; Bobbillo et al, 2006), muchos de ellos no son para entornos ubicuos o no contemplan la posibilidad de personalizar actividades colaborativas en este tipo de entornos. En consecuencia en esta investigación se propone definir un nuevo enfoque rescatando las potencialidades de los enfoques existentes, adaptándolos a los entornos ubicuos y

colaborativos, y a su implementación por medio de agentes de software.

## LINEAS DE INVESTIGACION Y DESARROLLO

El modelado de estudiantes presupone el abordaje de tres cuestiones fundamentales:

- Selección de la estructura que se usará para representar el modelo de estudiante. Esto implica definir el contenido del modelo, y la organización interna de los datos. Para resolver el problema del contenido se analizarán diferentes entornos de aprendizaje ubicuo tales como los que se describen en (Martín et al, 2006; Ortega et al, 2011; Chen et al, 2008) con el fin de identificar los datos que resultan relevantes para personalizar los diferentes servicios que ofrece un entorno ubicuo. En relación al problema de la organización interna de los datos, esta puede adoptar diferentes estructuras de representación tales como: árboles de decisión, redes semánticas, red bayesiana, reglas de decisión, etc. Se estudiarán las ventajas y desventajas de las diferentes formas de representación según el tipo de dato a almacenar.
- Inicialización del modelo de estudiante. La estructura elegida para representar los datos del estudiante debe inicializarse cuando la interacción con el sistema comienza. Existen diferentes enfoques para ello y uno de los más usados es el enfoque basado en técnicas de aprendizaje de máquina. Por lo tanto, se evaluarán estas técnicas para seleccionar las más adecuadas para inicializar los datos que se registren en el modelo.
- Diagnóstico. Es el proceso que permite actualizar el modelo de estudiante. Para ello se estudiarán los enfoques existentes para los diferentes tipos de modelos de estudiante: individual (Brusilovsky, 1994; Khodeir et al, 2010), colaborativo (Peña, 2005; Duran y Amandi, 2011), ubicuo (Graf et al, 2009) y se definirá un enfoque que integre los anteriores y permita crear y mantener un modelo de

estudiante integral para entornos ubicuos. También se analizarán las diferentes técnicas de aprendizaje de máquina, ya que si bien estas han demostrado su utilidad individualmente, si se quiere aprender modelos de estudiante más complejos es necesario realizar varias tareas de aprendizaje o bien combinar distintas técnicas de aprendizaje para una misma tarea de manera que se pueda aumentar su eficiencia.

Para construir modelos complejos y combinar diferentes técnicas de aprendizaje, las arquitecturas basadas en agentes inteligentes proporcionan un marco adecuado para realizar estas combinaciones. Por lo que se definirá un enfoque basado en tecnología de agentes para inicializar y mantener el modelo de estudiante para entornos ubicuos.

## OBJETIVOS Y RESULTADOS ESPERADOS

Si bien existen numerosos desarrollos vinculados al modelado de estudiante que permiten adaptar entornos de e-learning a diferentes aspectos de los alumnos (Brusilovsky, 1994; Khodeir et al, 2010; Gogvadze et al, 2011; Duran y Amandi, 2011) aun no está muy difundida esta tecnología en el campo del u-learning. La mayoría de los modelos desarrollados para estos entornos se limitan a registrar sólo la localización del usuario. Sin embargo, en contextos ubicuos es preciso conocer las características tecnológicas (tipo de red, plataforma, dispositivo, etc), pedagógicas y de interacción propias del aprendizaje ubicuo, así como los contextos en los que estas se manifiestan, para poder personalizar adecuadamente los servicios que presta un sistema de aprendizaje ubicuo.

Dentro de este marco el objetivo general de este trabajo de investigación es abordar y proponer soluciones al problema de la personalización en sistemas de aprendizaje ubicuo.

Los objetivos específicos de esta investigación son:

- Diseñar un Modelo de Estudiante que represente aspectos relevantes (personales, pedagógicos, tecnológicos y de interacción) para personalizar sistemas de aprendizaje ubicuos.
- Definir un enfoque basado en la tecnología de agentes para crear y mantener el modelo de estudiantes para sistemas de aprendizaje ubicuo.

Los resultados que se esperan obtener de esta investigación, conforme a los objetivos, son:

- Estado del arte referido a los diferentes aspectos a incluir respecto de los estudiantes en el modelo de usuario.
- Modelo de Estudiante que integre todos los aspectos relevantes para la personalización de servicios en entornos ubicuos.
- Módulos que inicialicen los datos a registrar en el modelo.
- Módulos basados en agentes que permitan mantener el modelo de estudiante en entornos ubicuos.

## FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo de esta línea se integra por un Becario de Postgrado Tipo I del CONICET, un Director de la Beca y Director del Proyecto de investigación indicado en el contexto y un Co-Director de la Beca y asesor del Proyecto de investigación.

En esta línea de investigación, el becario CONICET desarrollará su tesis para el Doctorado en Ciencias de la Computación de la Universidad Nacional del Centro de la Provincia de Buenos Aires.

## REFERENCIAS

F. Bobillo, M. Delgado, J. Gómez-Romeroer (2006). Uso De Modelos De Restricción Del Contexto Para El Desarrollo De Aplicaciones Móviles Inteligentes.

- P. Brusilovsky (1994). Student model centered architecture for intelligent learning environments, en Proceedings of fourth International Conference on User Modeling, 15-19 de Agosto, Hyannis, MA, USA, User Modeling Inc, 31-36.
- G.D. Chen, C.K. Chang, C.Y. Wang (2008). Ubiquitous learning website: Scaffold learners by mobile devices with information-aware techniques. Elsevier. Computers & Education 50, 77-90.
- E. Duran y A. Amandi (2011). Personalised collaborative skills for student models. Interactive Learning Environment. Ed. Joseph Psotka and Steve Wheeler, Routledge, Taylor & Francis Group, vol. 19, N° 2, 143-162.
- S. Graf, G. Yang, T. Liu, Kinshuk (2009). Automatic, Global and Dynamic Student Modeling in a Ubiquitous Learning Environment. Knowledge Management & E-Learning: An International Journal, Vol. 1, No. 1.
- N. Khodeir, N. Wanas, N. Darwish, N. Hegazy (2010). Utilizing Diagnosing Problems In A Probabilistic Domain To Build Student Models. International journal of computer science & information Technology (IJCSIT) Vol.2, No.4.
- S. Martín, M. A. Castro, J. Peire, A. Colmenar, R. Gil (2006). Nuevas Aplicaciones de la Computación Ubicua en la Enseñanza Personalizada. II Congreso Iberoamericano sobre Computación Ubicua (CICU 06), España, 89-96.
- A. Peña (2005). Collaborative Student Modeling by Cognitive Maps, en Proceedings of the First International Conference on Distributed Frameworks for Multimedia Applications (DFMA '05), Besançon, Francia, IEEE Press.
- J. Self (1994). Formal approaches for student modelling, en J. Greer y McCalla (eds.), Student Modelling: the key to individualized knowledge-based instruction, Springer-Verlag.
- P. Seppälä & H. Alamäki (2003). Mobile learning in teacher training. Journal of Computer Assisted Learning, 19(3), 330-335.
- G. Gogvadze, S. Sosnovsky, S. Isotani, B. M. McLaren. (2011). Evaluating a bayesian student model of decimal misconceptions. In the Proceedings of the 4th International Conference on Educational Data Mining.



## Aprendizaje Organizacional & Aprendizaje Móvil

Gustavo Lopez<sup>1</sup>, Liliana Maria Figueroa<sup>2</sup>, Marilena Maldonado<sup>3</sup>

(1) *Departamento de Matemática, Facultad de Ciencias Exactas y Tecnologías, Universidad Nacional de Santiago del Estero*  
*gustavojlopez@gmail.com*

(2) *Departamento de Informática, Facultad de Ciencias Exactas y Tecnologías, Universidad Nacional de Santiago del Estero*  
*lmvfigueroa@yahoo.com.ar*

(3) *Departamento de Informática, Facultad de Ciencias Exactas y Tecnologías, Universidad Nacional de Santiago del Estero.*  
*marilena@unse.edu.ar*

### Resumen

Los recursos pedagógicos mediados por la tecnología evolucionaron de tal manera que dejaron de lado el modelo aprendiz-maestro, creando contextos más flexibles en distintas comunidades de aprendizaje e impulsando estrategias donde y cuando el aprendiz lo requiera. Por otro lado, cada vez hay una mayor tendencia al uso de dispositivos móviles para manipular la información, para integrar actividades cotidianas, para manejar tecnología de punta y para buscar información. Con esta tecnología los usuarios están aprendiendo a adaptar y realizar sus tareas movilizándose, es decir a ser y estar móviles.

Además, a las organizaciones les interesa y se preocupan por adquirir nuevos conocimientos en el momento justo en el que se necesite para ser aplicado en el contexto adecuado por quien lo requiera, cuando toma decisiones, diseña, planifica, diagnostica y evalúa. Por lo tanto, surge la necesidad de gestionar este conocimiento de una forma especial que brinde apoyo al aprendizaje individual a partir de las experiencias y prácticas realizadas.

Es entonces, un desafío para las organizaciones que hoy sus empleados se capaciten en cualquier momento y en cualquier lugar, y puedan adquirir y compartir conocimiento.

Por lo tanto para las organizaciones, conformar un plan o estrategia de capacitación basado en aprendizaje móvil (M-learning) requiere no sólo contar con tecnología apropiada, sino también

crear una formación apoyada en la demanda de cada uno de los empleados, con aplicaciones diferentes basadas especialmente en dispositivos móviles.

Es por esta razón, que en esta línea de investigación se busca relacionar M-learning con el aprendizaje organizacional.

**Palabras clave:** Aprendizaje Organizacional, M-learning, Organización.

### 1 Contexto

Este informe es parte de un Programa de Investigación en el área de la Informática, que tiene como objetivo realizar investigación aplicada sobre nuevas tecnologías, duras y blandas, para optimizar los procesos de aprendizaje; y en particular en el Proyecto "Optimización de la Calidad del Proceso de Software con Gestión del Conocimiento", Cod 23/C098, que se desarrolla en la Universidad Nacional de Santiago del Estero.

En este proyecto se involucra el desarrollo de aplicaciones concretas de Gestión del Conocimiento y el Aprendizaje Organizacional lo cual contribuirá marcadamente a mejorar la calidad del proceso de desarrollo del software para organizaciones públicas y privadas del medio socio productivo del NOA. Y es financiado por el Consejo de Ciencia y Técnica de la Universidad Nacional de Santiago del Estero, durante el período Enero 2012-Diciembre 2015.

## 2 Introducción

Los avances en ciencia y tecnología han traído consigo un cambio sustancial en las prácticas de todas las esferas de la sociedad. En particular, en la educación somos testigos de que los recursos didácticos mediados por tecnologías son una alternativa para desarrollar procesos de aprendizaje. La implementación de estos medios ha generado el diseño de diversos ambientes, más allá de los presenciales y es común ahora encontrar, con el apoyo de estos recursos tecnológicos (con mayor o menor uso), ambientes multimodales, ambientes combinados o mixtos [1], ambientes digitales en línea por Internet [2], y recientemente, en ambientes de aprendizaje móvil [3].

A partir de esta realidad, la propuesta de esta investigación es plantear al aprendizaje móvil como una estrategia de aprendizaje organizacional.

## 3 Líneas de investigación y desarrollo

El aprendizaje organizacional es la habilidad de una organización para lograr un mejor entendimiento de sus prácticas a partir de la experimentación, la observación, el análisis y la colaboración; y una organización que aprende es una organización que adquiere nuevos conocimientos y es capaz de innovar con la rapidez necesaria para adaptarse y sobrevivir en un entorno dinámico.

En estos últimos años, los avances en TIC han sido aplicados, tanto en las instituciones educativas como en las organizaciones; debido a que factores como tiempo, costo y disponibilidad han comenzado a impactar cada vez más en las mismas; y en la manera en que llevan a cabo todos sus procesos de negocios y en particular la formación de sus empleados.

Hoy las empresas aplican modelos de formación y capacitación basados en nuevos desarrollos tecnológicos y en dispositivos móviles, cuya reciente aplicabilidad se debe también al incremento del gasto en formación de sus recursos humanos.

Este modelo demanda grandes retos, permitiendo que los empleados tengan acceso a su formación, en cualquier momento y en cualquier lugar, cuando lo quieran; ya que el aprendizaje organizacional no tiene una ubicación fija y de esta manera se va logrando que la organización adquiera nuevos conocimientos y experiencias

que le permita adaptarse a las nuevas exigencias del ambiente cambiante.

Además, las tecnologías de comunicaciones móviles también pueden emplearse para sistematizar y organizar los saberes internos de la organización, a fin de promover la interacción espontánea y la formación continua; de esta manera, se podría construir la infraestructura necesaria para la innovación. Por otro lado, la flexibilidad y escalabilidad que caracteriza a las tecnologías del M-learning contribuyen con la generación, intercambio y aplicación de conocimientos de forma permanente.

De esta manera, el objetivo del aprendizaje organizacional recurriendo a tecnologías móviles es desarrollar y compartir una inteligencia colectiva, a través de la comunicación efectiva, con el propósito de desarrollar nuevos conocimientos y así poder innovar.

Por lo tanto, la propuesta en materia de aprendizaje organizacional es que la formación se realice bajo una propuesta de ubicuidad. Esto hace que la formación sea más efectiva, se fomente la formación directa, basada en la demanda; donde lo más significativo es que se permite el acceso a los materiales formativos en el momento en que se lleva a cabo la tarea práctica.

Para un trabajador móvil y en particular para todo aquel empleado que no pueda o no desee mantener una ubicación fija por mucho tiempo, es indispensable contar con soporte tecnológico que le permita tener acceso a la capacitación en todo momento, sin importar el medio de comunicación, la hora o el lugar. Por lo tanto la nueva estrategia de aprendizaje, que se viene presentado desde un comienzo en esta investigación, emerge para satisfacer las necesidades individuales de estos empleados, permitiéndoles acceso a información específica desde cualquier lugar, en cualquier momento y en cualquier dispositivo portátil, ya sea para su utilización inmediata o posterior.

## Resultados y Objetivos

Como resultado de esta línea de investigación, se puede considerar que las organizaciones de nuestro medio, se ha dado un “contagio” por el uso de telefonía móvil, incrementando la demanda en forma personal, pero aún no se la pueda considerar como una propuesta para que las organizaciones puedan aprender y tengan la información y el conocimiento que describan actividades y tareas que deben realizarse

sistemáticamente, conformando de esta manera una estrategia continua de formación y capacitación. Entonces, se requeriría un cambio cultural que fomente la capacitación en el mismo lugar de trabajo y desde las mismas experiencias compartidas, y por otro lado, también los profesionales informáticos deben iniciarse en el desarrollo de aplicaciones para estos entornos tecnológicos.

Además, existen algunas cuestiones que no permiten potenciar la mejora en los procesos de aprendizaje continuos a través de la tecnología móvil, que permita fortalecer el conocimiento de los procesos de negocios; es decir, *plantear al aprendizaje móvil como una estrategia de aprendizaje organizacional*.

Para lo cual, se plantea seguir con las siguientes líneas de investigación:

- Analizar alternativas que permitan representar el conocimiento existente en las organizaciones.
- Definir una estrategia metodológica para diseñar y desarrollar herramientas para capturar conocimientos y experiencias durante el desarrollo de proyecto.

#### 4 Formación de recursos humanos

Los integrantes de esta investigación están desarrollando sus actividades bajo un paradigma científico común atendiendo la problemática que se aborda. También se continuará trabajando en forma colaborativa con los otros miembros del programa de investigación, uniendo esfuerzos para promover un mayor desarrollo científico tecnológico de esta región.

Se aspira que con esta línea de investigación:

- Lograr ser fuente de origen de trabajo de investigación para alumnos de la carrera Licenciado en Sistemas de la Información de la Universidad Nacional de Santiago del Estero y para alumnos de Ingeniería en

Informática de la Universidad Nacional de Catamarca.

- Mantener líneas de trabajo que sirvan a nuestro medio, que busca nuevas estrategias para acrecentar su aprendizaje organizacional.
- Acrecentar la formación continua de los profesionales investigadores.

#### Referencias

[1] Young, J. R. (2002). Hybrid teaching seeks to end the divide between traditional and online. Chronicle of Higher Education, Disponible en: <http://chronicle.com/free/v48/i28/28a03301.htm>

[2] Pastore, R. (2002). Elearning in Education: An Overview. Proceedings of Society for Information Technology and Teacher Education International Conference 2002. (275-276). Chesapeake, VA: AACE.

[3] Laouris, Y.; Eteokleous, N. (2005). We need an Educationally Relevant Definition of Mobile Learning. mLearn 2005 4th World conference on mLearning. Conference theme: Mobile technology: The future of learning in your hands. Disponible en: <http://www.mlearn.org.za/CD/papers/Laouris%20&%20Eteokleous.pdf>

#### Otras Fuentes de Información

- “El Futuro de las Tecnologías Móviles y su Aplicación de Aprendizaje: MOBILE LEARNING. Prof. María Rosa Depetris; Mg. Danya Tavela; Abog. Maria Florencia Castro. Universidad Nacional del Noroeste de la Provincia de Buenos Aires.
- Tecnologías Móviles Aplicadas a la Educación Superior. Susana I. Herrera. Marta C. Fennema. Universidad Nacional de Santiago del Estero. Universidad Tecnológica Nacional, Facultad Regional Tucumán.

# REDES SOCIALES: CONDICIONES INSTITUCIONALES Y COMPETENCIAS TECNOLÓGICAS PARA SU INTEGRACIÓN A LA EDUCACION EN AMBITOS UNIVERSITARIOS

Claudia Minnaard, Noelia Morrongiello, Leandro Rodriguez  
Facultad de Ingeniería de la Universidad Nacional de Lomas de Zamora  
Instituto de Investigaciones de Tecnología y Educación (IIT&E)  
Camino de Cintura y Avda. Juan XXIII –Llavallol  
Teléfonos: 011-4282 7880

e- mails: [minnaard@uolsinectis.com.ar](mailto:minnaard@uolsinectis.com.ar) ; [morrongiello\\_noelia@yahoo.com.ar](mailto:morrongiello_noelia@yahoo.com.ar)

## Resumen

El presente trabajo se orienta a abordar una problemática de actualidad en el campo de integración de TIC a la Educación Superior. Se destacan los avances que hasta el momento, y en forma gradual vienen implementando Instituciones Universitarias, los que en general se ven reflejados en la creación de ámbitos virtuales de enseñanza, bajo el formato: Campus o Plataforma Virtual. Sin embargo y frente a la llegada de nativos digitales al nivel superior de estudios por un lado y el éxito que el uso de redes sociales tienen en este segmento de usuarios, nos preguntamos si las mismas pueden ser utilizadas como herramienta colaborativa para la construcción de conocimiento o solamente tienen potencialidad como herramienta de comunicación. En ambos casos se pretende explorar las condiciones institucionales necesarias para su desarrollo y capacidades con las que debe contar el cuerpo académico para un eficaz aprovechamiento de las mismas.

## Contexto

A lo largo de las últimas dos décadas, la sociedad viene asistiendo a un extraordinario posicionamiento de las Tecnologías de la Información y Comunicación (TIC), en distintos ámbitos: educativos, políticos, económicos, entre otros. Esta circunstancia impone a las

instituciones la necesidad de desarrollar acciones en diversos sentidos. Por un lado planificar actividades para la integración de las TIC a sus respectivos procesos, y por otro, desarrollar estrategias anticipatorias del cambio.

Las Instituciones de Educación Superior, no son ajenas a esta problemática y por lo tanto es fundamental que los escenarios educativos tradicionales estén suficientemente preparados para reconfigurarse en tiempo útil a la luz de estas cuestiones.

Se observa que por un lado, las Universidades reciben una nueva generación de estudiantes “nativos digitales”, para quienes el uso de las TIC en todos los ordenes de la vida es central, con la natural consecuencia para el ámbito de la educación, de que los modelos y estrategias de enseñanza se adapten a este orden.

Este cambio generacional básicamente en su relación con el entorno tecnológico, deriva en una diferenciación fundamental con sus padres y profesores denominados “Inmigrantes Digitales” (Prensky, 2001)

Creemos que esta aptitud de los jóvenes para con la tecnología, que los lleva a adaptarse naturalmente a entornos online, presenta hoy un desafío para las instituciones de educación superior, que podría ser capitalizada por los responsables de gestionarlas, con el objeto de a) contribuir a la calidad de la enseñanza que imparten y 2) mejorar sustancialmente la comunicación institucional.



En este contexto, se advierte el permanente esfuerzo de las instituciones orientado a actualizar el equipamiento informático, pero sin embargo, vemos también que como el cambio es vertiginoso, la obsolescencia física de los recursos se produce cada vez más rápidamente. Solo a modo de ejemplo podemos señalar como las memorias USB o la de los espacios de almacenamiento en la nube, han venido a reemplazar rápidamente dispositivos que no hace tantos años se utilizaban como diskettes y CDs..

En este mismo sentido, las instituciones han habilitado ámbitos virtuales de enseñanza o de apoyo a la función docente en formato de Plataformas Educativas; hoy son muchos los docentes que intercambian materiales con los alumnos a través de estos Entornos Virtuales. Se trata de ámbitos en los que se comparten experiencias y reflexiona sobre las mismas, se acompaña a los alumnos a través de sistemas de tutorías en línea, o se construyen colaborativamente documentos utilizando foros, correo electrónico, wikis, etcétera. Es decir, no se puede negar que las Instituciones vienen adaptándose a las demandas de la sociedad en relación a la inclusión de tecnologías en las aulas.

No obstante, no resulta difícil advertir que estos espacios institucionales reproducen la dinámica del aula presencial, al tiempo que se asiste a un debate en la comunidad de educadores que propicia la búsqueda de nuevos modelos y prácticas basadas en el diálogo, independientemente de la tecnología que se utilice para su mediación.

### **Líneas de Investigación y Desarrollo**

Siguiendo a Torre Espejo (2009), coincidimos que la navegación por Internet, jugar digitalmente o pasar por las redes sociales no implica necesariamente una pérdida de tiempo, sino que pensamos que quienes transitan estos espacios están asimilando competencias tecnológicas y comunicativas muy necesarias para el mundo contemporáneo

Así como a finales del Siglo XX, a los jóvenes se los comienza a llamar “Generación net”, “Generación conectada” o “Generación de la Web”, por reconocer su pertenencia al ciberespacio, hoy se ha comenzado a llamar a los nuevos usuarios como “Generación Red”(Tapscott, 1998), “e-Generación, Millennial Generación y Generación Interactiva” (Bringué y Sabada, 2008) o “Generación Einstein” (Piscitelli, 2009).

Frente a este nuevo escenario, la gran mayoría, de los actuales docentes del nivel universitario, se ven enfrentados a la necesidad de una permanente actualización no solo en materia pedagógica y disciplinar, sino también tecnológica, que solo es una forma parcial de paliar la brecha digital entre ellos y sus alumnos.

Como se señaló, la existencia de Entornos Virtuales de Aprendizaje o Plataformas Educativas disponibles en las distintas Instituciones se han integrado decididamente a las prácticas educativas, sin embargo, hoy los mismos aparecen como ámbitos con formatos y dinámica cercana al de las aulas tradicionales frente a la potencialidad que ofrecen los recursos disponibles hoy en la Web.

La utilización de estas herramientas en el proceso de enseñanza y aprendizaje, encontraría su fundamento pedagógico en las teorías del conectivismo (Siemens, 2004), como una etapa superadora a las conductistas, cognitivistas y constructivistas, y aportaría las bases para el diseño de nuevos escenarios de aprendizaje.

El conectivismo es una resultante de la integración de los principios explorados por el caos, la red y la complejidad y las teorías de auto-organización, donde el aprendizaje - es un proceso que ocurre dentro de los entornos nebulosos de elementos centrales cambiantes - no está totalmente bajo el control del individuo (Siemens, 2004)

El autor señala que el aprendizaje, puede residir fuera de las personas (dentro de una organización o una base de datos) y se centra en la posibilidad de conectar conjuntos de información especializada, de manera que las conexiones que permiten

aprender son más importantes que el propio estado de conocimiento.

Esta nueva corriente de *e-learning* (Meso Ayerdi, 2011) que se conforma por un lado en torno a las concepciones y principios de la Web y que – por otro - tiene su fundamentación pedagógica en las teorías del *conectivismo* se apoyan en los siguientes principios:

- El alumnado es el que crea y modifica los contenidos por medio de las herramientas que pone a su disposición la Web

- Se conforman las experiencias de aprendizaje gracias a la reutilización y las remezclas de los contenidos provenientes de diversas fuentes de acuerdo con las necesidades e intereses del alumnado.

que permiten a los usuarios un mayor grado de comunicación y colaboración de forma flexible y rápida.

Por su parte, Cabero et al (2009) señala que el actual entorno Web potencia espacios virtuales para la interacción social, la participación abierta basada en aplicaciones telemáticas, gratuitas y fáciles de manejar

Las redes sociales, emergen de este entorno y nos preguntamos si su creciente popularidad pone en evidencia la necesidad de incorporar su uso para la docencia. Si la respuesta es afirmativa, creemos necesario abordar la siguiente cuestión a abordar: cuales son las potencialidades y limitaciones percibidas por los docentes para su incorporación al ámbito académico.

El Análisis de las Redes Sociales (ARS) es un conjunto de técnicas y metodologías que permiten la medición y análisis de las estructuras que surgen de la interacción de los actores. Estas técnicas y metodologías consisten en:

- El estudio formal de las relaciones para analizar las estructuras sociales que surgen de la recurrencia de relaciones.

- La asunción es que el análisis de los fenómenos se enriquece estudiando las relaciones entre países, instituciones, individuos, etc.

Asimismo, la interacción entre actores, además de reflejar flujos de información y conocimiento, tiene un impacto particular sobre el comportamiento de los agentes y sobre sus procesos de aprendizaje. El ARS

considera la red como una construcción relacional que puede ser presentada como una matriz y que luego puede ser analizada con herramientas del lenguaje matemático, el álgebra relacional y la teoría de grafos, teniendo origen en el problema de los puentes de Königsberg y la solución planteada por Leonhard Euler en 1741.(Barrere, 2012)

Es posible analizar cuantitativamente las formas en que los individuos o instituciones se conectan, o están vinculados, con el objetivo de determinar:

- La estructura de la red

- Sus flujos de información

- Sus relaciones de “poder”

*“Los patrones de las redes generan oportunidades significativas y restricciones que afectan el acceso de personas e instituciones a recursos e información”* (Sanz, 2003).

## Resultados y Objetivos

El presente trabajo tiene como objetivo general explorar las estructuras de las Redes Sociales y su capacidad para dar respuesta a las necesidades educativas del colectivo de usuarios considerados nativos digitales en el ámbito de la educación superior.

En particular, se pretende indagar la permeabilidad de los docentes para desenvolverse en un ámbito abierto, no planificado como el caso de Facebook.

Por último se busca contribuir al diseño de un instrumento para evaluar las capacidades institucionales que favorezcan la emergencia de un aprendizaje colaborativo e innovador a través de este tipo de entorno y a diagnosticar las competencias tecnológicas del profesorado universitario para su implementación.

Se trata de un estudio de alcance descriptivo, con diseño cualicuantitativo. Se trabajará con vestigios digitales a través de la técnica ARS y se administrará una encuesta a docentes cuyos datos se someterán a tratamiento estadístico.

## Formación de Recursos Humanos

El presente proyecto de investigación se integra al programa de investigación, liderado por el Dr. Ing. Oscar Pascal, a través del programa “Las TIC y su contribución al proceso de enseñanza y aprendizaje en carreras de Ingeniería: Evaluación de Experiencias en la Universidad Nacional de Lomas de Zamora”. En la actualidad el programa integra tres proyectos, que cuentan con aval institucional y participa del programa de Incentivos de la Secretaría de Políticas Universitarias, en los que participan 15 personas, entre las que se cuentan docentes investigadores, personal técnico y personal de apoyo.

La línea de trabajo que se presenta emerge como problemática a partir de los avances tecnológicos que se vienen experimentando en materia de TIC, pero fundamentalmente de los cambios que se advierten en las relaciones de los jóvenes con aquellas.

## Referencias

BARRERE, R. (2012). Módulo de Procesamiento y Análisis de datos. Maestría en Metodología de la Investigación Científica UNLA

BRINGUÉ, X.; SÁDABA, C. (Coords.) (2008): *La Generación Interactiva en Iberoamérica*, Madrid, Fundación Telefónica/Ariel

CABERO, J.; LÓPEZ, E.; LLORENTE, M.C. (2009). *La docencia universitaria y las tecnologías Web 2.0. Renovación e innovación en el Espacio Europeo*. Sevilla: MergabluM.

MESO AYERDI, K. (2011) La implementación de las redes sociales en la enseñanza superior universitaria. En Tejuelo Nro. 12 págs. 137-155

PISCITELLI, A. (2009): “Nativos e inmigrantes digitales. Una dialéctica intrincada pero indispensable”. En R. CARNEIRO et al. (coords.). *Los desafíos de las TIC para el cambio educativo*, Madrid, OEI.

PRENSKY, M. (2001): “On the Horizon”. En *MCB University Press*, 9 (5). Disponible en <<http://www.marcprensky.com/writing/Prensky%20-%20Digital%20Natives,%20Digital%20Immigrants%20-%20Part1.pdf>> [Consultado el 13/3/2011]

SANZ, L. (2003). Análisis de redes sociales: o cómo representar las estructuras sociales subyacentes. En: *Apuntes de Ciencia y Tecnología*, N° 7.

SIEMENS, G. (2004). “Connectivism: A Learning Theory for the Digital Age”. Disponible en <<http://www.elearnspace.org/Articles/connectivism.htm>> [Consultado el 05/3/2013].

TAPSCOTT, D. (1998): *Grown Up Digital: How the Net Generation is Changing*, New York, McGraw-Hill.

TORRE ESPEJO, A. de la (2009): “Nuevos perfiles en el alumnado: la creatividad en nativos digitales competentes y expertos rutinarios”. En *Universidad y Sociedad del Conocimiento*, 1, Vol. 6, pp. 7-14. Disponible en <<http://rusc.uoc.edu/ojs/index.php/rusc/artic/e/view/v6n1-de-la-torre/17>>. [Consultado el 13/3/2011]

## COMUNIDAD DE APRENDIZAJE EN UN CONTEXTO VIRTUAL: RED DE DOCENTES DE LA FI UNLZ

Marta Comoglio, Oscar Pascal, Oscar Campoli

Facultad de Ingeniería de la Universidad Nacional de Lomas de Zamora

Instituto de Investigaciones de Tecnología y Educación (IIT&E)

Camino de Cintura y Avda. Juan XXIII –Llavallol

Teléfonos: 011-4282 7880

e- mails: [mcomoglio@gmail.com](mailto:mcomoglio@gmail.com) ; [oscarmpascal@hotmail.com](mailto:oscarmpascal@hotmail.com)

### Resumen

El presente trabajo estudia la dinámica de un espacio virtual colaborativo: la Red de Docentes de las Carreras de Ingeniería de la Facultad de Ingeniería de la Universidad Nacional de Lomas de Zamora, con el objeto de analizar la forma a través de la que se genera conocimiento en espacios virtuales de nivel superior. Se propone, asimismo, indagar las percepciones de los integrantes de la Red respecto de la actividad colaborativa y su impacto en la función docente.

Se presenta un estudio de carácter descriptivo, desarrollado a través de un diseño cuali-cuantitativo.

### Contexto

La Facultad de Ingeniería de la Universidad Nacional de Lomas de Zamora (FI UNLZ) viene desarrollando, desde el año 2005, un proceso gradual y sostenido, destinado a integrar TIC a la enseñanza presencial. Ha implementado diferentes acciones orientadas a crear y sostener, en la Unidad Académica, una cultura de la Innovación en la Educación. Acompañando esta estrategia, en el año 2009 se crea el Instituto de Investigaciones en Tecnología y Educación (IIT&E), con el objeto de contar con un ámbito institucional que propicie la generación de conocimiento a partir de la reflexión de las propias prácticas educativas.

En este contexto surge la Red de Docentes de la Facultad como un espacio virtual dedicado al desarrollo de las capacidades docentes. Se trata de una Red

de colaboración académica y virtual de pares, que ha permitido canalizar diferentes experiencias de Enseñanza y Aprendizaje. La misma forma parte del proyecto institucional que promueve la integración de las TIC a la planificación, gestión y evaluación de prácticas educativas. Se trata de un espacio dinámico basado en la construcción colaborativa de conocimientos entre colegas, compartiendo logros e inquietudes, experiencias y dificultades; un espacio de encuentro y aprendizaje, que refuerza el sentido de pertenencia y cohesiona al cuerpo de docentes como institución.

### Introducción

De manera vertiginosa, la sociedad se ha visto impactada, durante los últimos años, por un inusitado desarrollo tecnológico, el cual plantea nuevas posibilidades para la educación. Bajo este nuevo paradigma surgen nuevos planteos y desafíos, para los que la colaboración se presenta como una alternativa de solución a los distintos tipos de organización (Duran, 2001). Es así como Hargreaves (2003) señala que la colaboración para la resolución de problemas es una de las piedras angulares de las organizaciones. En este marco, la colaboración en red -reforzada por la incorporación de TIC- en organizaciones educativas es considerada como una de las claves del aprendizaje (Badía y García, 2006; Garrison, 2006; Harasim et al., 2000; Kirschner, 2002; Majó y Marques, 2002). En su informe presentado a la UNESCO, Delors (1997) planteaba las características esenciales para la educación en el siglo XXI



en cuatro premisas que todos los gobiernos, mediante su sistema educativo, debían implementar: “aprender a aprender; aprender a hacer; aprender a ser; y aprender a trabajar en grupo”.

El actual contexto exige a los docentes ser polivalentes: sumar a sus labores de enseñanza las actividades de investigación y extensión; agregar a sus currículos estudios de posgrado en áreas disciplinares; y profesionalizar sus conocimientos en pedagogía. Esto requiere de estrategias activas de gestión, como la articulación de prácticas y experiencias, la creación de nuevos modos de producción de conocimiento, y la inclusión de herramientas facilitadoras como las TIC (ambientes virtuales que favorecen la interacción de los actores y promueven la emergencia de nuevos perfiles y liderazgos en el seno de una comunidad determinada; y potencian los intercambios, la creatividad y los nuevos patrones de acción y de libertad académica entre los docentes).

La conformación de este nuevo perfil de profesores deriva en la necesidad de que las instituciones educativas desarrollen propuestas innovadoras que impacten, positivamente, en las actuales prácticas de enseñanza; requiere de transformaciones estructurales, producto de la combinación de políticas e instituciones, capaces de diversificar las estrategias de formación, reconsiderar el rol del profesor, e incorporar tecnologías y virtualidad a los procesos.

Este proceso de transformación que, como se señaló, se inicia en la Facultad en el año 2005, ha dado lugar al surgimiento de la Red de Docentes, soportada en la Plataforma E-ducativa, cuyas actividades se centran en:

- seguimiento y revisión curricular;
- planificación y evaluación de actividades de formación;
- articulación y prácticas compartidas entre docentes;
- formación docente auto-asistida;
- creación de bases documentales e instrumentales;
- elaboración de boletines, eventos, foros y debates, que permitan socializar los logros alcanzados;

- intensificar la formación práctica;
- desarrollar competencias tecnológicas en los profesores.

## Líneas de Investigación y Desarrollo

No cabe duda que, desde sus comienzos, Internet se ha desarrollado bajo la idea de comunidad. Es así como en el mundo de la red se recrean, permanentemente, situaciones de la vida cotidiana a través de lo que se ha dado en llamar “comunidades en línea”, “comunidades electrónicas” o “comunidad virtual”.

El término “comunidad” proviene del latín *communis* (común) o *communicare* (establecimiento de una comunalidad). Foster (1997) señala que, entre los señalados, el más utilizado es el de “comunidad virtual”, y está conformado por dos nociones con connotaciones diferentes. Por un lado, se puede definir comunidad como “conjunto de personas vinculadas por características o intereses comunes”<sup>1</sup>. En este sentido, Wilbur (1997) sostiene que el término deriva de la idea de común y, por lo tanto, la esencia de la comunidad está vinculada a la noción de un grupo de personas que comparten objetos, ideas, propiedades, identidades, cualidades, entre otros atributos, siempre en un espacio común. La representación del espacio común es la que permite evocar un espacio geográfico donde se ubican o residen las personas integrantes de esa comunidad. Sin embargo, el autor advierte que el significado de la palabra no alude al espacio físico donde los integrantes de la comunidad se reúnen, sino que hace referencia a la cualidad de los grupos, de compartir distintos objetos comunes.

Por otra parte, en relación al término virtual, podemos señalar que, etimológicamente, deriva de *virtus*, que quiere decir “fuerza o energía inicial”; es decir, la capacidad de producir efectos. Más tarde, la física toma el término para denotar “conjunto de los puntos aparentes de

<sup>1</sup> RAE. Real Diccionario de la Lengua Española. XXII Edición.

convergencia de los rayos luminosos que proceden de un objeto, después de pasar por un espejo o un sistema óptico, y que, por lo tanto, no puede proyectarse en una pantalla”<sup>2</sup>.

Podemos concluir que la acepción actual del término emana de esas dos raíces semánticas: por un lado, la virtud de producir un efecto y, por el otro, la de imagen especular.

Siguiendo este hilo encontramos que la noción de lo virtual establece dos dicotomías: una entre lo real/irreal, y otra entre la representación del objeto y el objeto representado. De esta manera, se vislumbra una tercera vertiente: que lo virtual es una simulación de lo real. Entonces, encontramos que *virtual* es aquello que tiene virtud para producir un efecto, aunque no lo produce en presente, por lo que, frecuentemente, se utiliza en oposición a efectivo o real.<sup>3</sup>

Como se señaló, es el surgimiento de las redes las que proporcionan la infraestructura para la comunicación mediada tecnológicamente, lo que permite esta forma de co-presencia virtual, resultante de interacciones electrónicas individuales, que no se encuentran restringidas por limitaciones de tiempo y espacio, comunicación ésta que ocurre en lo que se ha dado en llamar “ciberspacio”, y que permite un acercamiento a la noción de “comunidad virtual” (Miller, 1999).

Por su parte, Foster (1997) recoge la definición de Rheingold, en el sentido de que la comunidad virtual es “la agregación social que emerge de la Red cuando suficiente gente desarrolla discusiones públicas lo suficientemente largas, con suficiente sentimiento humano, formando redes de relaciones personales en el ciberspacio”. Se trata, entonces, de comunidades de personas basadas en los intereses individuales, y en sus afinidades y valores.

Lo señalado en el párrafo precedente permite concluir que “toda comunidad virtual descansa sobre tres pilares: los

aspectos de interés común, los integrantes de la comunidad y el espacio virtual” (Salinas, 2003).

Se coincide que las características fundamentales sobre las que se sustentan las comunidades virtuales, y que, por lo tanto, condicionan su calidad de vida, son: accesibilidad, cultura de participación, colaboración, aceptación de la diversidad y voluntad de compartir, destrezas disponibles entre los miembros, contenidos relevantes, por citar los más importantes (Pazos, 2001).

Cherny (1999) asegura que el sentimiento de comunidad es esencial en la vida de las comunidades virtuales, ya que ellas requieren algo más que el mero acto de conexión; para concluir, “la clave está en la interacción humana a través de ordenadores”.

Es así que, teniendo en cuenta que es el texto el que se convierte en el medio fundamental de intercambio en los ambientes no presenciales, Rourke et al. (2001) destaca la importancia de considerar tres componentes al momento de analizar las dinámicas de estas comunidades: 1) la presencia cognoscitiva; 2) una presencia de la enseñanza; y 3) una presencia social. Este autor define la presencia cognoscitiva como “el grado al cual los participantes, en cualquier configuración particular de una comunidad de la investigación, son capaces de construir el significado a partir de una comunicación sostenida”.

El segundo elemento, presencia de enseñanza (actividad de los tutores y coordinadores), incluye el diseño educativo y la conducción del proceso, facilitando un aprendizaje activo.

El tercer elemento es la presencia social, definida como la capacidad de los participantes de proyectarse social y emocionalmente en una comunidad, cuya función es apoyar los objetivos cognoscitivos y afectivos del aprendizaje, a partir de la capacidad de cuestionar, sostener, apoyar y ampliar el pensamiento crítico de los participantes en una verdadera comunidad de aprendizaje.

Este círculo virtuoso se estructura en las interacciones del grupo que rescatan y dan continuidad a los aportes de cada integrante,

<sup>2</sup> REA- Real Diccionario de la Lengua Española XXII Edición.

<sup>3</sup> Ibidem.

contribuyendo al aumento de la integración académica, social e institucional.

No cabe duda que las posibilidades de comunicación que ofrecen las nuevas tecnologías permiten, en este caso, a los docentes-alumnos realizar aportes, emitir opiniones, compartir reflexiones que difícilmente podrían darse en un ámbito exclusivamente presencial, justamente por la dificultad que significa coincidir en tiempo y espacio.

En síntesis, la Red de Docentes es un espacio virtual de conocimiento-acción, que tiene como objetivos *favorecer la comunicación e interacción entre docentes y coordinadores; facilitar el intercambio de materiales y resultados de experiencias; compartir opiniones e información; y realizar consultas*. El espacio cuenta con varias secciones: a) seguimiento curricular; b) formación pedagógica; c) apoyo a docentes; y d) foros.

Este marco referencial es el que sustenta la implementación de la Red de Docentes, cuya actividad es objeto de la presente investigación.

## Resultados y Objetivos

Desde la perspectiva teórica señalada en los párrafos precedentes, este trabajo concentra su atención en comprender cómo se llevan a cabo el proceso de colaboración entre docentes en contextos virtuales en el ámbito universitario, e indagar las percepciones sobre el aporte que deriva de la colaboración. Se aspira que, a partir de los resultados de la presente investigación, se contribuya a mejorar los procesos de colaboración entre docentes que tiene lugar en un entorno virtual.

Se trata de un estudio de caso, en el que intervienen docentes de distintas disciplinas y categorías, de las carreras de Ingeniería Industrial y Mecánica.

Los registros con los que se trabajará son vestigios digitales de foros de discusión programados como parte de las actividades académicas que se desarrollan en la Red. Asimismo, se trabaja con datos provenientes de una encuesta administrada al colectivo de docentes de la unidad académica.

Se trata de un estudio estructurado sobre la base de un diseño metodológico con elementos propios de la investigación cualitativa: observación participante, etnografía virtual (Hine, 2000), análisis del discurso electrónico (Romaña, 2007), y tratamiento de datos por técnicas estadísticas.

Se utiliza también el forograma, como herramienta alternativa de evaluación de espacios de discusión virtual (Salazar Villegas, 2004).

## Formación de Recursos Humanos

El equipo de trabajo está liderado por el Dr. Ing. Oscar Pascal, a través del programa “Las TIC y su contribución al proceso de enseñanza y aprendizaje en carreras de Ingeniería: Evaluación de Experiencias en la Universidad Nacional de Lomas de Zamora”. Este programa se integra mediante tres proyectos, en el que participan 15 personas, entre las que se cuentan docentes investigadores, personal técnico y personal de apoyo. Una de las líneas de trabajo es la que se presenta en esta oportunidad. Actualmente, está en curso una tesis de maestría vinculada, directamente, con esta línea de I+D.

## Referencias

- Badia, A.; García, C. (2006): “Incorporación de las TIC en la enseñanza y el aprendizaje basados en la elaboración colaborativa de proyectos”. En: Badia, A. (Coord.): *Enseñanza y aprendizaje con TIC en la educación superior* [monográfico en línea], Revista de Universidad y Sociedad del Conocimiento (RUSC). Vol. 3, n° 2. UOC.
- Cherny, L. (1999): *Conversation and Community. Chat in a Virtual World*. CSLI Publications, Standford California.
- Delors, J. y otros. (1997): *La educación encierra un tesoro*. Madrid, Santillana. Ediciones UNESCO.
- Durán D. (2001): “Cooperar para triunfar”. En *Cuadernos de Pedagogía*, n° 298, enero 2001, pp. 73-75. Barcelona. Cispraxis.

- Foster, D. (1996): "Community and Identity in the Electronic Village". En Porter, D. (Ed.): *Internet Culture*. Routledge New York 23-38.
- Garrison, D. R. (2006): *Online collaboration principles*. Journal of Asynchronous Learning Networks, 10(1), 25-34.
- Harasim, L.; Hiltz, S. R.; Turoff, M; Teles, L. (2000): *Redes de aprendizaje. Guía para la enseñanza y el aprendizaje en red*. Barcelona, Gedisa.
- Hargreaves, A. (2003): *Enseñar en la sociedad del conocimiento: La educación en la era de la inventiva*. Barcelona, Ediciones Octaedro.
- Hine, C. (2000): *Virtual Ethnography*. Sage, Londres.
- Kirschner, P. A. & Lai, K-W (2007): *Online communities of practice in education*. Technology Pedagogy and Education 16 (2), pp. 127-132.
- Majó, J.; Marquès, P. (2002): *La revolución educativa en la era internet*. Barcelona, CissPraxis.
- Pazos, M., Perez Garcias, A. y Salinas, J. (2001): "Comunidades virtuales: de las listas de discusión a las comunidades de aprendizaje". Comunicación. Edutec'01. V Congreso Internacional de Tecnología, Educación y Desarrollo Sostenible. 17-19 de septiembre, Murcia.
- Rheingold, H. (1993): *La comunidad virtual. Una sociedad sin fronteras*. Barcelona, Gedisa.
- Romaña, Blay Teresa (2007): "Evaluar el trabajo con foros electrónicos: propuesta de un sistema". Revista de Universidad y Sociedad del Conocimiento (RUSC). Vol. 4, n.º 2. UOC. Disponible en: <http://www.uoc.edu/rusc/4/2/dt/esp/romana.pdf>
- Rourke, I.; Anderson T.; Garrison A.; Archer W. (2001): "Assessing Social Presence In Asynchronous". Text-based Computer Conferencing. Journal of Distance Education/Revue de l'enseignement à distance.
- Salazar Villegas A. M. (2004): "Forograma, una estrategia alternativa para la evaluación de espacios virtuales de discusión". Ponencia VII Congreso Colombiano de Informática Educativa.
- Salinas, J. (2003): "Comunidades Virtuales y Aprendizaje Digital". Disponible en <http://gte.uib.es/pape/gte/sites/gte.uib.es/pape/gte/files/Comunidades%20Virtuales%20y%20Aprendizaje%20Digital.pdf>
- Wilbur, H.M. (1997): Experimental ecology of food webs: complex systems in temporary ponds. Ecology 78, pp. 2279-2302.



# CONDICIONES LINGÜÍSTICAS DE LA PRODUCCION DEL CONOCIMIENTO CON HERRAMIENTAS DE COMUNICACIÓN ASINCRÓNICA: EL FORO EN ESPACIOS VIRTUALES DE APRENDIZAJE

Mariana Fernandez, Paula Penco

Facultad de Ingeniería de la Universidad Nacional de Lomas de Zamora

Instituto de Investigaciones de Tecnología y Educación (IIT&E)

Camino de Cintura y Avda. Juan XXIII – Lomas de Zamora

Teléfonos: 011-4282 7880

e- mails: [mariana.fernandez@yahoo.com.ar](mailto:mariana.fernandez@yahoo.com.ar) ; [paucecipenco@yahoo.com.ar](mailto:paucecipenco@yahoo.com.ar)

## Resumen

El presente trabajo se orienta a estudiar las condiciones lingüísticas de la producción de conocimiento en ámbitos académicos universitarios, en particular la que surge en ámbitos de comunicación asincrónica dentro de espacios virtuales de aprendizaje.

Aborda el análisis de la escritura generada en uno de los espacios más utilizados en el ámbito académico: el foro. Se plantea como un diseño descriptivo correlacional con análisis de vestigios digitales resultante del intercambio escrito en Foros Académicos desarrollados en un Ambiente Virtual de Aprendizaje (EVA), soporte de estudios de nivel superior universitario.

Se trabaja sobre una experiencia de enseñanza en la modalidad Blended Learning, que tiene lugar en la Facultad de Ingeniería de la Universidad Nacional de Lomas de Zamora (FI UNLZ).

## Contexto

Una de las estrategias que la Facultad de Ingeniería de la Universidad Nacional de Lomas de Zamora (FI UNLZ) viene desarrollando desde el año 2005, ha sido la integración de las TIC como complemento a la enseñanza presencial. En la actualidad el 75% de las asignaturas de las carreras de Ingeniería Mecánica e Industrial cuentan con sus correspondientes aulas virtuales,

alojadas en la Plataforma Institucional Educativa operativa en la siguiente dirección <http://www.itc-unlz.com.ar>,

En el año 2010 se crea el Instituto de Investigaciones en Tecnología y Educación (IIT&E) en el ámbito de la Facultad, cuya misión es: a) propiciar un ámbito de reflexión de las propias prácticas docentes y b) generar nuevo conocimiento a partir del estudio sistemático de experiencias educativas mediadas por TIC en el ámbito de la educación superior.

El presente trabajo explora algunas de las problemáticas que se visualizan como resultado del intercambio escrito en ámbitos académicos mediados por tecnologías de la información y la comunicación (TIC).

## Introducción

Con el advenimiento de la era de la información y la comunicación, marcada por el cambio de paradigma de la modernidad a la posmodernidad, y sus múltiples novedades, la sociedad a nivel mundial se encuentra atravesada por una transformación de las funciones y efectos sociales del pensamiento en las prácticas cotidianas, que afectaron la forma de ver el mundo y vivir la vida.

Según el autor Gianni Vattimo (1996), *“hemos entrado en un escenario, el de la postmodernidad, donde la comunicación y los medios adquieren un carácter central, aunque esa abundancia de emisores continuos no aporta una visión unitaria, ni siquiera una visión contextualizada e*

*independiente. Asistimos a una especie de 'babel informativa' que, más que aturdir y violentar, abre caminos a la libertad, a la pluralidad, y se escapa de las visiones unitarias de la racional-modernidad”.*

Así, destacamos el desarrollo y maximización de los medios masivos de comunicación, los cuales han tomado al sujeto desde sus múltiples ámbitos, en lo personal, laboral y académico.

Siendo de nuestro interés abordar el último ámbito mencionado, cabe destacar que el autor Luceli Patiño Garzón (2006) define *“la escritura, entendida como epistemología del conocimiento, que viabiliza el pensamiento y se constituye en una estrategia de aprendizaje que funda la tradición universitaria. A su vez, fortalece una cultura académica centrada en la argumentación, la discusión y el registro. Todos estos son elementos básicos para la producción del conocimiento como fenómeno público”.*

En este marco y a partir de experiencias concretas, nuestra preocupación radica en indagar los cambios emergentes de la producción escrita en los medios virtuales, específicamente, el resultante de foros de debate académico del nivel universitario. El proyecto se orienta al análisis lingüístico de los elementos conversacionales producidos en instancias de comunicación escrita asincrónica, con la finalidad de caracterizar las vicisitudes que experimenta la lengua española durante las mismas.

## **Líneas de Investigación y Desarrollo**

Así como la evolución de los medios masivos, a nivel social, ha generado una maximización de la comunicación y sus posibilidades, también enriqueció a las instituciones educativas, mediante múltiples herramientas digitales que abrieron y siguen abriendo nuevos caminos a la enseñanza y el aprendizaje, como es el caso del foro.

Alexánder Arbey Sánchez-Upegui (2009) define el foro como *“una comunicación grupal, dialógica, temática, asincrónica y argumentativa, orientada a*

*generar un proceso de construcción del conocimiento”.*

Sin dudas, se trata de una herramienta informática, muy utilizada en el ámbito educativo, con la finalidad de debatir sobre una o más temáticas propuestas, y construir nuevos conocimientos de manera cooperativa. Lejos de poner en duda el logro de este objetivo que persigue la herramienta del foro, nuestro interés se centra en el desarrollo de la escritura y su forma. Notamos evidente que, al igual que los medios masivos de comunicación, dentro y fuera del ámbito educativo el espacio del foro vislumbra las mismas problemáticas lingüísticas que en aquéllos.

Francisco Yus (2001) realiza un estudio de las particularidades del uso del lenguaje en la comunicación electrónica, manifestando que *“esta comunicación virtual se realiza a través de ‘mensajes en forma de texto pero con clara vocación oral’ y, por lo tanto, ‘híbridos’, entre la ‘estabilidad’ de la escritura y la ‘espontaneidad’ y lo ‘efímero’ de la interacción oral”.*

En principio, todos los sujetos hemos notado que, en el momento de escribir un texto, formal o informal, contar un suceso, una historia o compartir información, no hay espacio ni tiempo suficientes para ahondar en detalles. En el caso del espacio, muchas veces el mensaje tiene que ser breve y concreto; además, dado nuestro ritmo tan acelerado de vida, el tiempo tampoco es suficiente para escribir toda la información completa. Por estos motivos, un texto completo y correctamente escrito pasa a ser una síntesis que, muchas veces, no sólo deja de lado información o detalles del contenido que se quiere comunicar, sino, más grave, comienzan a aparecer palabras abreviadas e, incluso, cambios de la ortografía, que se alejan de una correcta escritura.

De esta manera, se observa la evolución creciente de los cambios en la escritura, aun en ámbitos académicos, es decir, de formación profesional, se hace cada vez más frecuente debido al contemporáneo uso e incremento de los medios de comunicación que acompañan la vida diaria de cada sujeto. Estas

circunstancias estarían llevando a una inminente transformación del lenguaje escrito dentro de la sociedad, por su creciente aplicación en medios escritos, informales o no.

Es este conjunto de problemáticas que se visualizan como resultado de la comunicación escrita en ámbitos académicos mediados por tecnologías de la información y la comunicación (TIC) es el núcleo de nuestro planteo problemático.

Cabe mencionar que los cambios señalados no se limitan a las redes sociales, sino que este fenómeno se inició con los mensajes de texto SMS, que en 150 caracteres aproximadamente (incluyendo signos de puntuación y espacios) obligaban a reducir la información, con el riesgo de perder el núcleo y el sentido del mensaje. En muchos casos, esto produjo un crecimiento de los mensajes con poco sentido, casi en tono de telegrama y, poco a poco, se volvió una costumbre y se expandió a otros medios. Luego, se trasladó al correo electrónico que, aunque tiene espacio ilimitado de texto, siguió con esa costumbre de reducir los mensajes a su mínima expresión, para invertir menos tiempo y energía posible en transmitir la información que era motivo de dicho correo.

Otro fenómeno influyente, con la llegada y evolución de Internet, fue la convivencia con el lenguaje escrito en inglés, esto es, palabras sin acentos, un solo signo de admiración o interrogación, etc., que aplicamos a nuestros mensajes escritos diarios (sean mensajes SMS, correos electrónicos, chat, redes sociales, etc.). Por último, la tendencia fue más allá y se trasladó al chat, donde, si bien hay espacio ilimitado para redactar información, también se observa un uso inadecuado del lenguaje escrito, al manejarlo incompleto e incorrecto, para perder el menor tiempo, apuntando siempre a la agilidad, rapidez, inmediatez y fluidez del uso de los medios de comunicación.

Adentrándonos en el ámbito académico, consideramos que el foro es una de las herramientas centrales para la construcción colectiva de conocimiento, y en los mismos vislumbramos claramente,

esta problemática. Sin embargo, al margen de la calidad educativa que se desarrolle en un foro, en este caso nuestro interés radica en el estudio de la expresión escrita en este espacio virtual, y evaluar su alejamiento de la normatividad lingüística.

Una indagación preliminar nos permitió identificar algunas transformaciones de la lengua, presentes en forma reiterada en espacios de foros educativos, tales como abreviación informal de palabras, como ser el uso de “q”, “xq”, “xa”, “tb” (en lugar de “que”, “porque”, “para”, “también”, respectivamente), u otras palabras típicas utilizadas en mensajes de texto en teléfonos celulares, que se trasladan al resto de los medios virtuales de escritura. Además, el uso de los signos al modo del lenguaje inglés, es decir, la falta del primer signo de exclamación e interrogación es una constante de la escritura en los espacios virtuales, como también la duplicación o triplicación de dichos signos finales, sea para acompañar una sola palabra, frase u oración, cuya multiplicación parece querer resaltar el sentido de lo dicho, o tener una finalidad afectiva. Otra cuestión notoria es la disminución del uso de signos de puntuación, mayormente la coma (,), y del tilde, aspecto que se encuentra también vinculado a la convivencia con el lenguaje inglés, cuya lengua no posee acentuación ortográfica.

Por otra parte, la presencia de expresiones coloquiales y otras marcas de oralidad, muchas veces vislumbradas en expresiones afectivas, no se corresponderían con el espacio educativo en el cual se está interactuando, donde, también, muchas veces suelen aparecer “emoticones”, que son una secuencia de caracteres del teclado que se utilizan para acentuar un mensaje e intentar expresar emociones de parte del usuario. Por ende, estas secuencias, que podríamos llamar “íconos emotivos”, acompañan dichas expresiones afectivas o aparecen por sí solos, ya que cada uno posee un sentido en sí mismo. De esta manera, los usuarios utilizan una nueva forma de expresarse, que normalmente representa un rostro humano en sus diversas expresiones de emociones, mediante el cual pueden

mostrar su estado de ánimo a través de una forma de expresión iconográfica introducida en un campo textual.

## Resultados y Objetivos

En el marco del planteo problemático señalado, el presente estudio propone: a) comprobar si los foros de debate académico utilizan lenguaje informal, más cercano al oral que al escrito con características más formales b) Identificar, clasificar y establecer tipologías de mutaciones de la lengua española en ámbitos virtuales de enseñanza y aprendizaje colaborativo.

Para alcanzar estas metas se ha diseñado un estudio de carácter descriptivo correlacional. Las unidades de análisis serán las producciones escritas en diversos foros de las asignaturas correspondientes a las carreras: Ingeniería Mecánica e Industrial de la Facultad de Ingeniería de la Universidad Nacional de Lomas de Zamora.

Las hipótesis de trabajo son: 1) “en los foros de debate académico se utiliza un lenguaje informal, más cercano al oral que al escrito”; 2) “la lengua escrita presenta una serie de transformaciones respecto de la normatividad de la lingüística española, con independencia del ámbito -académico o no- en el que se produzcan las transacciones comunicacionales.

## Formación de Recursos Humanos

El equipo de trabajo está liderado por el Dr. Ing. Oscar Pascal, a través del programa “Las TIC y su contribución al proceso de enseñanza y aprendizaje en carreras de Ingeniería: Evaluación de Experiencias en la Universidad Nacional de Lomas de Zamora”. Este programa se integra mediante tres proyectos, en el que participan 15 personas, entre las que se cuentan docentes investigadores, personal

técnico y personal de apoyo. Una de las líneas de trabajo es la que se presenta en esta oportunidad. Actualmente, está en curso una tesis de maestría vinculada directamente con esta línea de I+D. A la vez, una de las autoras se encuentra cursando un Máster en Educación, con especialización en “Las TIC en la Educación”, teniendo como objetivo abordar, investigar y desarrollar estas temáticas, como eje de su tesis.

## Referencias

Arbey Sánchez-Upegui, Alexander (2009): *Nuevos modos de interacción educativa: análisis lingüístico de un foro virtual*, Facultad de Educación de la Universidad de la Sabana, disponible en <http://www.google.com.ar/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CC8QFjAA&url=http%3A%2F%2Fdialnet.uniri.oja.es%2Fdescarga%2Farticulo%2F3084415.pdf&ei=XaU3UdeXNoqc8QSYoIHgDw&usg=AFQjCNGQy1iM20X54AXGPMUXJwcXk98xMg&sig2=ZZ4ZbUybmjqX44j400UghQ&bvm=bv.43287494,d.eWU>

Gianni Vattimo (1996): *Credere di credere*, informe disponible en <http://www.infoamerica.org/teoria/vattimo1.htm>

Patiño Garzón, Luceli (2006): *La escritura académica en la formación del docente universitario*, disponible en <http://aprendeonline.udea.edu.co/revistas/index.php/revistaeyp/article/viewFile/6940/6353>

Yus, Francisco (2001): *Ciberpragmática* (El uso del lenguaje en Internet), disponible en <http://www.scielo.org.ar/pdf/spilquen/n6/n6a12.pdf>



## M-learning en la universidad.

### Marco de análisis. Aplicaciones móviles como objetos de aprendizaje.

Susana I. Herrera<sup>1</sup>, Jorge L. Goñi<sup>1</sup>, Marta C. Fennema<sup>1,2</sup>, María I. Morales<sup>3</sup>

(1) *Instituto de Investigación en Informática y Sistemas de Información, Facultad de Ciencias Exactas y Tecnologías, Universidad Nacional de Santiago del Estero*  
sherrera@unse.edu.ar, jlgoni@live.com.ar

(2) *Departamento de Sistemas, Facultad de Tecnología y Ciencias Aplicadas, Universidad Nacional de Catamarca*  
crisfen@yahoo.com

(3) *Departamento de Matemática, Facultad de Ciencias Exactas y Tecnologías, Universidad Nacional de Santiago del Estero*  
mines\_morales@yahoo.com.ar

### Resumen

Se propone investigar acerca de la mejor manera de implementar experiencias o proyectos de m-learning en la educación universitaria. Para ello, por un lado, se está elaborando un marco que permita analizar, diseñar y evaluar experiencias de m-learning en el posgrado; por otro lado, se construirán objetos de aprendizaje que sean recursos de m-learning para el nivel de grado y posgrado.

A partir de antecedentes propios y de una revisión bibliográfica para indagar sobre la temática, se está elaborando un marco con sustento teórico que permite identificar y caracterizar las experiencias o proyectos de m-learning. Dicho marco se plasma en una serie de grillas orientadoras que identifican y caracterizan dichas experiencias, tornándolas comparables y evaluables.

Además se construirán aplicaciones móviles, considerando cuestiones de desarrollo de sistemas móviles, de diseño de interacción de recursos de m-learning. Estas aplicaciones se elaborarán a partir de la orientación provista por el marco y considerándolos como objetos de aprendizaje.

Las experiencias de uso de estos objetos de aprendizaje realimentarán el marco, enriqueciéndolo con cuestiones de desarrollo de aplicaciones móviles para el aprendizaje.

**Palabras clave:** m-learning, objetos de aprendizaje, educación superior, marco de análisis y evaluación del m-learning, aplicaciones móviles.

### 1 Contexto

Esta investigación se lleva a cabo en el marco del proyecto denominado "Optimización de la calidad de los

Sistemas Móviles mediante la implementación de nuevas arquitecturas, realidad aumentada, técnicas de visualización y redes móviles Ad-Hoc. Aplicaciones en m-learning y en gestión del conocimiento". A su vez, el proyecto forma parte de un Programa de Investigación en el área Informática cuyo objetivo consiste en realizar investigación aplicada sobre nuevas tecnologías que permitan mejorar la calidad tanto del proceso como del producto software.

El equipo de investigación del proyecto está conformado por docentes de diferentes universidades nacionales del NOA: del Instituto de Investigaciones en Informática y del Departamento de Matemática de la Facultad de Ciencias Exactas de la Universidad Nacional de Santiago del Estero (UNSE), del Departamento de Sistemas de la Facultad de Tecnologías y Ciencias Aplicadas de la Universidad Nacional de Catamarca (UNCA) y del Departamento de Computación de la Facultad de Ciencias Exactas de la Universidad Nacional de Salta (UNSA).

Además, el proyecto cuenta con el asesoramiento de investigadores de Institutos de Investigación de la Facultad de Informática de la Universidad Nacional de La Plata, en los temas referidos al aprendizaje mediado por tecnologías y al desarrollo de aplicaciones móviles.

Está financiado por el Consejo de Ciencia y Técnica de la Universidad Nacional de Santiago del Estero, durante el período Enero 2012-Diciembre 2015.

### 2 Introducción

Los dispositivos móviles constituyen una de las tecnologías más usadas y presentan ventajas en cuanto a su portabilidad y a su sensibilidad al contexto. El aprendizaje mediado por tecnologías móviles se conoce como m-learning (mobile-learning) [27] y es una de las modalidades que está tomando auge en la actualidad, ya que involucra dispositivos que son asequibles a toda la

población. Esta expansión no es ajena al nivel educativo superior de posgrado, donde los smartphones están permitidos durante las clases y son muy útiles para los aprendices que no disponen de tiempo dedicado para el estudio, debido a sus actividades laborales [8, 15, 24, 30, 33].

Si bien el m-learning se ha desarrollado y extendido mundialmente, presentando diversas estrategias, programas y herramientas [1, 2, 8, 9, 10, 11, 18]; se investiga y debate aún sobre las teorías que lo sustentan. Varios grupos de investigación están trabajando en esta temática y han resultado de antecedente para el trabajo que aquí se propone [20, 21, 28, 29], en particular se toman las fundamentaciones de m-learning basadas en conceptos de ecología socio-cultural [21, 22]. A nivel nacional existen pocas experiencias de m-learning realmente implementadas. Se presentan propuestas aisladas basadas en casos concretos [3, 4, 6, 26] o recomendaciones que no se basan en experiencias implementadas [3].

Esta investigación sobre m-learning comenzó en el año 2012; actualmente se está iniciando el segundo año del proyecto. Durante el primer año, se trabajó en la región del NOA de Argentina. Se presentó una herramienta metafórica que actúa a modo de marco de referencia para la implementación de propuestas de m-learning. Se abordó el concepto de ecosistema de m-learning y fue descrito detalladamente en [12, 13, 14]. A su vez dicho ecosistema permite introducir al estudio de diversas alternativas de m-learning (aplicaciones, programas, herramientas) que pueden ser clasificados en tres modos, siguiendo los trabajos de Woodill [33].

En base a todos los antecedentes mencionados, se está elaborando un marco referencial con el fin de analizar y evaluar propuestas de m-learning, a nivel de experiencias y programas [15]. Dicho marco incorpora no solo los antecedentes teóricos del m-learning a nivel mundial sino también las teorías y estrategias educativas mediadas por tecnología que caracterizan a la Argentina [7, 23, 25, 31].

Asimismo, retomando a Woodill, el m-learning se puede llevar a cabo a través de tres modos diferentes:

- Modo 1. Recuperación de información. Brindan comunicación en un único sentido, acceder a información.
- Modo 2. Recopilación y análisis de información. Brindan comunicación en ambos sentidos, acceso a información y respuestas o envíos de información a los compañeros o a la profesora.
- Modo 3. Comunicación, interacción y colaboración en redes. Brindan comunicación en las comunidades.

En el primer año de investigación se realizaron experiencias correspondientes al Modo 1 [14] y se prevé en el futuro inmediato desarrollar aplicaciones móviles que interactúen en modo 3: juego o ambiente simulado que permita el aprendizaje colaborativo utilizando smartphones.

Para ello, se diseñará la aplicación como un objeto de aprendizaje (OA) móvil que tome las cuestiones de diseño apropiadas para este tipo de recurso de m-learning [24].

Un OA es una colección de material usado para presentar y apoyar a un único objetivo, representado en un contenido, o una actividad de aprendizaje, dando soporte a diferentes ambientes.

Un OA, concepto que corresponde a un conjunto de recursos que podrían ser utilizados como unidades independientes y reutilizables en plataformas diferentes [32], puede ser cualquier entidad, digital o no-digital, usada para el aprendizaje, educación o entrenamiento [17]. Ese tipo de objetos, por su naturaleza, debe ser reutilizable (RLO: Reusable Learning Object) en el proceso de enseñanza.

Utilizando OA, es posible crear bibliotecas de conocimiento, permitiendo que cursos diferentes utilicen un mismo objeto. Otra ventaja está en la personalización del aprendizaje a partir de la selección y configuración de aquellos objetos que auxilien al aprendiz en la construcción y apropiación del propio saber.

Respecto al m-learning aplicado en la educación superior, es importante considerar las propuestas, tendencias y recursos sugeridos por el especialista en pedagogía estadounidense Clark Quinn [24]. Este autor sostiene que la educación superior debe incorporar los beneficios de la computación móvil, los cuales pueden sintetizarse en cuatro "C": contenido, captura, cálculo y comunicación. Ejemplifica diversas situaciones de aprendizaje en el nivel superior, que considera tanto el contenido como la interacción y la socialización. El contenido (conceptual, contextual y lingüístico) está relacionado con los archivos de diverso tipo que son importantes principalmente en la introducción, presentaciones conceptuales, ejemplos y síntesis. La captura es importante en la interacción; a su vez, la interacción y el cálculo (o capacidad de procesamiento) es relevante en el diseño de la práctica y de la evaluación. Por último, la comunicación en los móviles optimiza también la interacción aprendiz-aprendiz y esto ayuda a lograr un aprendizaje significativo basado en el constructivismo social. Esa comunicación puede darse a través de medios formales o informales (Facebook, Twitter), en forma síncrona o asíncrona.

Para el desarrollo del objeto de aprendizaje móvil se utilizará una metodología ágil de desarrollo apropiada para las aplicaciones móviles. En este sentido, se toman los avances de otra línea de investigación del mismo proyecto que estudia la optimización de la calidad utilizando estrategias de Ingeniería del Software y técnicas de visualización. Para el desarrollo se cuenta con el asesoramiento del instituto LIFIA quienes están estudiando aplicaciones móviles que apoyen el aprendizaje [19].

Las experiencias de uso de estos objetos de aprendizaje móviles realimentarán el marco, enriqueciéndolo con

cuestiones de desarrollo de aplicaciones móviles para el aprendizaje.

### 3 Líneas de investigación y desarrollo

La principal línea de investigación de esta propuesta consiste en el análisis y diseño de estrategias de m-learning apropiadas para la educación de posgrado.

Esta línea implica:

- Diseñar un marco para el análisis y evaluación de experiencias y programas de m-learning de posgrado;
- Desarrollar objetos de aprendizaje que apoyen estrategias de m-learning del modo 3 (siguiendo a Woodill).

### 4 Objetivos y resultados

El objetivo general que guía esta investigación aplicada es:

*Proponer un marco de análisis, diseño y evaluación de experiencias y programas de m-learning de posgrado que contenga:*

- *aspectos claves para clasificar, comparar, estudiar, elaborar y evaluar propuestas educativas de m-learning; y*
- *aspectos claves para desarrollar aplicaciones de m-learning.*

Los resultados que se obtuvieron en el primer año de desarrollo del proyecto son:

- Ecosistema móvil del NOA [12, 13].
- Estrategias de m-learning de interacción en una única dirección, modo 1 [14].
- Marco “inicial” para el análisis, diseño y evaluación de propuestas de m-learning de posgrado [15].

### 5 Formación de recursos humanos

La Directora del proyecto pertenece al Departamento de Sistemas de la UNCA. La Codirectora, al Departamento de Informática de la UNSE. La asesora en Sistemas Móviles pertenece al Laboratorio LIFIA de la UNLP. La asesora en Tecnologías Informáticas aplicadas a la Educación pertenece al Laboratorio LIDI, de la misma universidad. El resto de los integrantes son jóvenes que se inician en tareas de investigación, alumnos avanzados de carreras de Informática del NOA. Por lo tanto se requiere un proceso permanente de apoyo y formación de éstos.

Esta investigación contribuye al desarrollo de dos tesis doctorales en Ciencias Informáticas (UNLP) una referida a un marco para el análisis y evaluación de experiencias de m-learning y otra referida a objetos de aprendizaje. También en este marco se realizará un trabajo final de

Especialización en Enseñanza de las Ciencias Exactas (UNSE). Se tiene previsto incorporar trabajos finales de alumnos de la Licenciatura en Sistemas de Información y Licenciatura en Matemática.

### Referencias

1. Bannan, B., Peters, E., Martínez, P. *Mobile, Inquiry-based learning and geological observation: An exploratory study*. International Journal of Mobile and Blended Learning, 2(3), 13-29. 2010.
2. Blythe, M., Reid, J., Wright, P., Geelhoed, E. *Interdisciplinary criticism: analysis de experience of Riot! A location-sensitive digital narrative*. Behaviour and Information Technology, 25, 2, 127-139, 2006.
3. Cataldi, Z & Lage, F. *TICs en Educación: Nuevas herramientas y nuevos paradigmas. Entornos de Aprendizaje Personalizados en dispositivos móviles*. Congreso TE&ET. ISBN 978-987-28186-0-9. Pergamino, Buenos Aires, 2012.
4. Cukierman y Otros. *Informe Final del Proyecto “Integración de la Tecnología Móvil a los Entornos Virtuales de Enseñanza y de Aprendizaje”*. Comisión de Investigaciones Científicas del Gobierno de la Provincia de Buenos Aires. Buenos Aires, 2008.
5. Cukierman, U. y Otros. *Una experiencia de uso de celulares en un curso de articulación escuela media y universidad en modalidad a distancia*. Publicado en VirtualEduca, Brasil, 2007.
6. Cukierman, U. y Virgili, J. *La Tecnología educativa al servicio de la educación tecnológica*. UTN, Buenos Aires, 2010.
7. Ferreira Szpiniak, A., Sanz, C. *Hacia un modelo de evaluación de entornos virtuales de enseñanza y aprendizaje. La importancia de la usabilidad*. TE&ET, Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología. 2010.
8. Fox, S. *Mobile Learning: Is Your Business Ready for It?* Technology Evaluation Centers, disponible en: <<http://www.technologyevaluation.com/research/articles/mobile-learning-is-your-business-ready-for-it-21858/>>, Diciembre 2010, fecha consulta: 05/01/2011.
9. Gwee, S., Chee, Y. S., Tan, E. M. *The Role of Gender in Mobile Game-Based Learning*. International Journal of Mobile and Blended Learning, 3(4), 19-37. 2011.
10. Hansen, F., Kortbek, K., Gronbaek, K., Spierling, U., Szilas, N. *Mobile Urban Drama: Setting the Stage with Location Based Technologies*. ICIDS 2008, LNCS 5334, pp. 20–31, Springer-Verlag, Berlin Heidelberg, 2008.
11. Hansen, F., Kortbek, K., Gronbaek, K. *Mobile Urban Drama for Multimedia-Based Out-of-School Learning*. ACM. 978-1-4503-0424-5, 2010.
12. Herrera, S. I. & M. C. Fennema. *Tecnologías Móviles Aplicadas a la Educación Superior*. Congreso

- Argentino de Ciencias de la Computación. La Plata, 2011.
13. Herrera, S. I., J. L. Goñi & M. C. Fennema. *El m-learning en la educación universitaria de posgrado*. Jornadas de Ingeniería del NOA. Catamarca, 2011.
  14. Herrera, S. I., M. C. Fennema & C. V. Sanz. *Estrategias de m-learning para la formación de posgrado*. Congreso TE&ET. ISBN 978-987-28186-0-9. Pergamino, Buenos Aires, 2012.
  15. Herrera, S. I., Fennema, M. C., Carrizo, M. I., Ponce, G. *Computación móvil aplicada al aprendizaje. Un marco de análisis del m-learning*. VIII Jornadas de Ciencia y Tecnología de Facultades de Ingeniería del NOA. San Miguel de Tucumán, Septiembre 2012.
  16. Herrington, A. and Herrington, J. *Authentic mobile learning in higher education*. In: AARE 2007. International Educational Research Conference, Fremantle, Western Australia, 2007.
  17. IMS Global Learning Consortium. *Instructional Management Systems*. Disponible en : <http://www.imsproject.org/specifications.HTML>
  18. Kalloo, V., Mohan, P. *An Investigation Into Mobile Learning for High School Mathematics*. International Journal of Mobile and Blended Learning, 3(3), 59-76. 2011.
  19. Lliteras, A., Challiol, C., Gordillo S. *Juegos Educativos Móviles Basados en Posicionamiento: Una Guía para su Conceptualización*. 13th Argentine Symposium on Software Engineering, 41 JAIO, La Plata, 2012.
  20. Naismith, L., Lonsdale, P., Vavoula, G., and Sharples, M. *Literature Review in Mobile Technologies and Learning*. ISBN: 0-9548594-1-3. University of Birmingham, 2004.
  21. Pachler, N., B. Bachmair & J. Cook. *Mobile Learning. Structures, Agency, Practices*. Ed. Springer, ISBN 978-1-4419-0584-0. New York, USA, 2010.
  22. Pachler, N., Cook, J., Bachmair, B. *Appropriation of mobile Cultural resources for learning*. International Journal of Mobile and Blended Learning, 2(1), 1-21. 2010.
  23. Pozo, J. I. *Aprendices y Maestros. La psicología cognitiva del aprendizaje*. Alianza Editorial. 2º Edición. ISBN 978-84-206-8349-2. Madrid, 2011.
  24. Quinn, C.N. *The Mobile Academy. mLearning for Higher Education*. Ed. Jossy-Bass. USA. 2012.
  25. Sanz, C., A. Zangara, C. Manresa-Yee. *E-activities in teaching processes using ICTs collaborative activity as a case study*. EDULEARN12 Proceedings, ISBN 978-84-695-3491-5, pp. 2034-2041. Barcelona, 2012.
  26. Sanz, C. y Otros. *Integración de la tecnología móvil a los entornos virtuales de enseñanza y de aprendizaje*. Publicado en TE&T.
  27. Traxler, J. *Defining Mobile Learning*. IADIS International Conference Mobile Learning. 2005.
  28. Traxler, J. *Defining, Discussing and Evaluating Mobile Learning: the moving finger writes and having write*. The International Review of Research in Open and Distance Learning. Vol.8, Nº2. 2007.
  29. Traxler, J. *Mobile Learning: Starting in the Right Place, Going in the Right Direction?*, International Journal of Mobile and Blended Learning, 3(2), 57-67, April-June 2011.
  30. UNESCO. *Declaración mundial sobre la educación superior en el siglo XXI: visión y acción*.
  31. Vigotsky, L. *El desarrollo de los procesos psicológicos superiores*. Ed. Crítica. Barcelona. 1979.
  32. Wiley, D. A. *Connecting Learning Objects to Instructional Design Theory: A definition, a metaphor, and a taxonomy*, in Instructional Use of Learning Objects. Ed Association for Instructional Technology, pp. 3-23. 2002.
  33. Woodill, G. *The mobile learning edge*. Ed. Mc Graw Hill, 2011.



## ESPACIOS VIRTUALES PARA TRABAJO COLABORATIVO

Rodríguez, D., García-Martínez, R., Merlino, H., Charczuc, N., Lacabanne, Caracciolo, B., Iglesias, F.

Grupo Investigación en Sistemas de Información

Departamento de Desarrollo Productivo y Tecnológico. Universidad Nacional de Lanús  
29 de Septiembre 3901 (1826) Remedios de Escalada, Lanús. Argentina. Tel +54 11 6322-9200 Ext. 194  
rgarcia@unla.edu.ar

### RESUMEN

Los espacios virtuales de trabajo colaborativo permiten la integración de grupos de trabajo en la que sus miembros no están físicamente contiguos. Hay una amplia literatura vinculada al modelado de las arquitecturas software que soportan este tipo de ambientes. Sin embargo, los formalismos existentes atienden la interacción entre actores y sistema y entre componentes del sistema; descuidando los aspectos de interacción humana.

**Palabras clave:** espacios virtuales de trabajo colaborativo, formalismos, herramientas para el modelado y diseño, arquetipo patrón de arquitectura de espacio virtual, herramientas de medición de interacción humana en espacios virtuales.

### CONTEXTO

El teletrabajo es una forma flexible de organización del trabajo consistente en el desempeño de la actividad profesional sin la presencia del trabajador durante una parte importante de su horario laboral. Dichas actividades laborales pueden ser desarrolladas a tiempo parcial o completo (Salazar, 1999). La aparición de Internet (Leiner et al., 1999), hace más de dos décadas, ha generado en el campo laboral nuevos paradigmas de teletrabajo (Salazar, 1999).

Los ambientes virtuales se usan hace más de un lustro en Educación Superior. Las Universidades, basadas en el uso masivo de la tecnología de Internet, han incorporado los campus virtuales como un medio a través de los cuales ofrecen (sin necesidad de presencia de los alumnos): cursos de extensión, programas de posgrado de especialización y maestría; estando en la actualidad, comenzado a ofrecer asignaturas de grado.

Era impensable, antes de la aparición de Internet, que equipos de desarrollo de proyectos pudieran realizar sus actividades sin contar con un lugar físico en el que cada uno de sus integrantes desarrollase sus tareas o; se realizaran las reuniones de equipo para consolidar resultados, evaluar la marcha del trabajo o discutir posibles soluciones a problemas emergentes del proyecto.

El concepto de espacio virtual para trabajo colaborativo (EVTC), surge de la fusión de los conceptos de: teletrabajo, equipos de desarrollo y espacios virtuales. Un EVTC se puede definir como un espacio basado en tecnología de Internet que permite el trabajo colaborativo de grupos en los que sus miembros no se encuentran físicamente contiguos (Rodríguez et al., 2012b).

Algunas de las ventajas, entre otras, que ofrece el trabajo grupal basado en EVTCs son: [a] el soporte informático de todos los artefactos desarrollados por el equipo de trabajo permite la trazabilidad de los avances y en consecuencia mejorar el control y la gestión del proyecto; [b] los costos vinculados a conexión de internet y servidores requeridos para el trabajo sobre EVTCs son sensiblemente menores a los costos vinculados a infraestructura física de espacios para trabajos presenciales; [c] el tiempo dedicado a traslados hasta el lugar de trabajo es ganado por el individuo para ocio o descanso con el consecuente impacto positivo sobre su productividad en las horas de trabajo.

Si bien se ha avanzado en el desarrollo de arquitecturas de software que soportan EVTCs (Bibbo et al., 2008), hay poco trabajo realizado para sentar las bases de una ingeniería de este tipo de ambientes virtuales. En particular, no se cuenta con herramientas que contribuyan al análisis y diseño y que estén orientadas a modelar las interacciones entre los sujetos que usan el EVTC, la gestión de tareas grupales y de los artefactos que surgen del trabajo conjunto (Rodríguez, 2012). La ausencia de estas herramientas conlleva a no poder especificar satisfactoriamente las funcionalidades que el EVTC debe cumplir a efectos de asegurar que el equipo de trabajo disponga de todos los elementos que aseguren su máximo rendimiento.

### INTRODUCCIÓN

Carlsen [1997] presenta una teoría del conocimiento en el marco de su trabajo sobre modelado de flujos de trabajo en el que sostiene que los términos: datos, información y conocimiento, son utilizados en forma ambigua por lo que propone las siguientes definiciones:

*Conocimiento*: Es un conjunto relativamente estable y suficientemente consistente de conceptos sabidos por un grupo de personas.

*Datos*: Denotan algún conjunto de representaciones de conocimiento expresadas en un lenguaje.

*Información*: Es el incremento de los conocimientos producidos por la acción de recibir un mensaje, es decir, es la diferencia entre las concepciones interpretadas a partir de un mensaje recibido y el conocimiento antes de la acción de recepción.

Drucker [1988] en sus trabajos sobre la información y sociedad del conocimiento, y sobre la transformación de las organizaciones basadas en la información y la organización de los especialistas científicos; propone la siguiente definición: "La información son datos dotados de relevancia y propósito; convertir datos en información requiere de conocimientos; el conocimiento, es por definición, especializado".

Nonaka [1991; 1994] define al conocimiento como una "creencia verdadera y justificada", sosteniendo que la información es un flujo de mensajes, y que el conocimiento "es creado y organizado por el flujo mismo de la información, basándose en el compromiso y las creencias de su poseedor"; de esta manera liga estrechamente la creación del conocimiento a la acción humana.

Carlsen [1997] establece que un punto central a las teorías de Drucker y de Nonaka es que el conocimiento dentro de una organización o grupo es creado a través de un continuo dialogo entre el conocimiento tácito y explícito desarrollado por los distintos actores del grupo, contribuyendo esta interacción a la amplificación y desarrollo de nuevo conocimiento.

La distinción entre conocimiento tácito y explícito se encuentra establecida por la ingeniería de conocimiento [García-Martínez y Britos, 2004] en la que se define al conocimiento explícito (conocimiento público o conocimiento codificado) como transmisible en lenguaje formal y sistemático, mientras que el conocimiento tácito tiene una cualidad personal que hace que sea difícil de articular, formalizar y comunicar.

Nonaka [2007] identifica cuatro patrones de interacción entre el conocimiento implícito y el conocimiento explícito, a los cuales llama modos de conversión de conocimiento como se presenta en la figura 1.

Carlsen [1997] sostiene que el modo de internalizar y externalizar la creación de conocimientos en la creación colaborativa de soluciones para problemas, se encuentra estrechamente relacionado con el proceso de "aprender haciendo", por lo tanto, la

acción está relacionada con el proceso de internalización.

Nonaka [1994] argumenta que las teorías tradicionales sobre el aprendizaje grupal, descuidan el abordaje de la noción de la externalización de lo aprendido y que prestan poca atención a la importancia de la socialización del conocimiento. Propone que las capacidades de aprendizaje son implícitamente mejoradas (o desarrolladas) durante el proceso de creación del modelo de conocimiento, ya que los grupos crean continuamente nuevos conocimientos mediante la reconstrucción de las perspectivas existentes del modelo de conocimiento desarrollado por ellos. Lo que hace única a esta concepción es la visión dinámica del conocimiento, que está en permanente creación, refinamiento y reformulación a partir de la información aportada por los miembros del grupo.

		Conocimiento Tácito	Conocimiento Explícito
Desde	Conocimiento Tácito	<b>SOCIALIZACIÓN</b> Creación del conocimiento tácito a partir de compartir experiencias	<b>EXTERNALIZACIÓN</b> Conversión del conocimiento tácito en conocimiento explícito
	Conocimiento Explícito	<b>INTERNALIZACIÓN</b> Conversión del conocimiento explícito en conocimiento tácito	<b>COMBINACION</b> Creación de nuevo conocimiento a partir de conocimiento explícito

Fig. 1. Modos de conversión de conocimiento según Nonaka

En los grupos de trabajo, el conocimiento explícito está normalmente representado por un prototipo o modelo que puede ser un representativo de un concepto. La innovación surge cuando se produce la interacción entre el conocimiento tácito y el conocimiento explícito. Nonaka [2007] establece que la interacción está determinada por los cambios entre los modos de conversión del conocimiento, inducida por varios factores desencadenantes, como se muestra en la Figura 1. En la figura 2, se muestra el modo de socialización de partida con la construcción de un espacio de interacción para facilitar el intercambio de experiencias y modelos mentales.

El enfoque tradicional de la gestión de flujo de trabajo se centra en el flujo de control dentro de la definición de un proceso [Jablonski y Bussler, 1996]. Las perspectivas que son relevantes para el modelado de flujo de trabajo y su ejecución son: (a) perspectiva desde el flujo de control o proceso, (b) perspectiva desde los recursos u organización, (c) perspectiva desde los datos o información, (d)

perspectiva desde la tarea o función y (e) perspectiva desde la operación o aplicación.

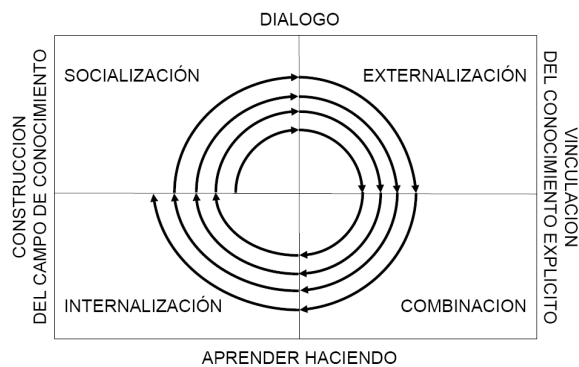


Fig. 2. Cambios entre los modos de conversión del conocimiento según Nonaka

Garrido [2003] propone para el modelado de flujo de trabajo, un marco conceptual basado en un modelo cooperativo representado por cuatro vistas realizadas bajo diferentes niveles de abstracción [Isla *et al*, 2004; 2007; Noguera, 2009]:

*Vista organizacional*: Refiere a la estructura estática y dinámica del grupo. Los estados representan los diferentes roles que pueden desempeñar los miembros en el grupo y las transiciones reflejan los posibles cambios de rol en virtud del cumplimiento de ciertas restricciones. Estas restricciones pueden ser capacidades (restricciones cognitivas impuestas a un actor para participar bajo un rol determinado) o leyes (restricciones impuestas por la propia organización que identifican las reglas sociales que deben ser preservadas en el grupo).

*Vista cognitiva*: Representa las tareas que puede llevar a cabo cada miembro del grupo en el escenario colaborativo. Por un lado se define la interfaz del rol, el cual incluye las características más relevantes del conjunto de tareas a realizar, y por otro lado se describen las tareas. En esta vista pueden aparecer elementos de las vistas de información (documentos, datos, recursos) y de interacción (protocolos).

*Vista de interacción*: Se analiza la forma de comunicación entre participantes y los recursos usados mediante protocolos de interacción de alto nivel.

*Vista de información*: Refleja la información que es compartida en el escenario o que se utiliza para la comunicación (documentos, eventos, recursos).

Estas vistas son modeladas a partir de una serie de componentes relativos al grupo y complementarios entre sí, y contribuyen a la comprensión dimensión del grupo como entidad organizativa [Fields *et al.*, 1997].

*Estructura*: Un aspecto fundamental de todo sistema es analizar y comprender su composición. Permite analizar la evolución que se produce en la organización (y por tanto en su propia estructura) mediante relaciones con el contexto.

*Comportamiento*: El grupo se organiza para realizar una finalidad. Este objetivo condiciona la manera de llevar esta labor y la división del trabajo. Permite abordar la realización de actividades por parte del grupo. Las tareas a realizar no se asignan directamente a actores, sino que se delegan a roles, condicionados por las estrategias del grupo. Los procesos cognitivos necesarios para realizar las tareas están distribuidos en la comunidad, y estos procesos se usan para reaccionar ante los nuevos eventos que se producen.

*Entorno*: Constituye el espacio de trabajo donde se desenvuelven los grupos.

*Dinámica*: Los grupos involucrados en una organización de tareas están sujetos a una dinámica cambiante en un proceso evolutivo. Los factores que pueden condicionar este cambio son alteraciones del entorno (nuevos objetivos), cambios estructurales (modificación de los miembros del grupo) o formas de llevarlo a cabo (nuevos métodos de interacción, dispositivos, entre otros). Para ello, habrá que identificar los aspectos más relevantes que influyen a un grupo bajo un modelo dinámico.

García Peñalbo y García Carrasco (2002) sostienen que un espacio virtual educativo debe ofrecer un conjunto de servicios educativos funcionales a los participantes del proceso formativo. Éstos pueden soportar una interacción síncrona, cuando los participantes están presentes "en línea" al mismo tiempo mientras se lleva a cabo el servicio, o asíncrona, cuando la presencia de todos los participantes no es requerida para desarrollar la actividad.

Para García Carrasco y su equipo de colaboradores (1999) los servicios provistos por el espacio virtual educativo pueden clasificarse diversos grupos no disjuntos entre tales como:

*Servicios de comunicación*: Facilitan la comunicación entre los protagonistas del proceso formativo (estudiantes y profesores). En este grupo se incluyen servicios tan populares como el correo electrónico, foros de discusión (síncronos como el IRC, o asíncronos como los grupos de noticias), seminarios virtuales, videoconferencias o publicación de documentos en formato digital.

*Servicios de información*: Ofrecen información genérica estructurada y dispuesta de forma



eficiente para un uso específico. Ejemplo de este servicio son las páginas web.

*Grupos de trabajo colaborativo:* Ofrecen la posibilidad de que varias personas trabajen juntas utilizando ordenadores y tecnología informática, facilitando el trabajo en equipo y un intercambio eficiente de información.

*Servicios de información:* Ofrecen información genérica estructurada y dispuesta de forma eficiente para un uso específico. Ejemplo de este servicio son las páginas web.

*Grupos de trabajo cooperativo:* Ofrecen la posibilidad de que varias personas trabajen juntas utilizando ordenadores y tecnología informática, facilitando el trabajo en equipo y un intercambio eficiente de información.

*Servicios de administración:* Permiten la gestión administrativa de las diversas entidades que conforman el dominio del problema del ámbito educativo, esto es, profesores, alumnos, cursos, informes estadísticos.

*Servicios de entretenimiento:* Son servicios, educativos o no, diseñados en su mayor medida para el ocio, como juegos en línea o tableros de noticias.

*Servicios y herramientas de autor:* mediante las cuales los formadores pueden producir unidades de actividad que, al tiempo que recuperan los modos escritos de oferta de conocimiento, pueden incorporar el modo oral, el icónico, y el audiovisual, dotados de reticularidad, organización topológicas y navegables en función de los intereses particulares del usuario.

Estos servicios quedan establecidos en el espacio virtual educativo dentro de un conjunto de componentes software de carácter pedagógico, junto a un repositorio de información, donde quedarán almacenados los diferentes activos de información que se intercambian en el proceso educativo. La interacción de los participantes en dicho proceso educativo se hace a través de dichos componentes software, en sus versiones cliente y servidor, donde normalmente el cliente manejado es un clásico navegador web, que da acceso al resto de los componentes (García Carrasco *et al.*, 1999).

El proyecto que se presenta articula una línea de trabajo que el grupo de investigación viene explorando desde el año 2009.

En esta exploración se han formulado consideraciones sobre el uso de espacios virtuales en la formación de investigadores (Rodríguez *et al.*, 2009), se ha propuesto diseño conceptual de espacios virtuales para ese fin (Rodríguez, D. 2010); se han realizado experiencias sobre entrenamiento mediado por espacios virtuales (Rodríguez *et al.*,

2010a); se han propuesto elementos para el análisis y diseño conceptual de espacios virtuales de trabajo colaborativo (Rodríguez *et al.*, 2010b); se han realizado pruebas de concepto y la correspondiente fundamentación sobre formación mediada por espacios virtuales (Rodríguez *et al.*, 2010c); se ha propuesto trabajar sobre la identificación de usos educativos de espacios de encuentro virtual (Charczuk, 2011); se han validado elementos de análisis y diseño para espacios virtuales centrados en formación (Rodríguez, 2012); se ha comenzado a trabajar sobre el diseño conceptual de espacios virtuales para el desarrollo de proyectos en materias de carrera de grado (Garbarini, 2012); se ha consolidado un modelo colaborativo de formación basado en espacios virtuales (Rodríguez *et al.*, 2012a); y se han propuesto líneas de investigación en Trabajo Colaborativo basado en Espacios Virtuales (Rodríguez *et al.*, 2012a).

## OBJETIVOS E HIPOTESIS DE INVESTIGACION

Los espacios virtuales de trabajo colaborativo permiten la integración de grupos de trabajo en la que sus miembros no están físicamente contiguos. El proyecto articula una línea de trabajo que el grupo de investigación viene desarrollando sobre esta temática desde el año 2009.

La pregunta problema que anima la investigación es: ¿Se puede cubrir la vacancia de disponer de un conjunto unificado de herramientas para el modelado y diseño de espacios virtuales de trabajo colaborativo?

*Hipótesis I:* Los espacios virtuales de trabajo colaborativo permiten la integración de grupos de trabajo en la que sus miembros están físicamente no contiguos. Hay una amplia literatura vinculada al modelado de las arquitecturas software que soportan este tipo de ambientes. Sin embargo, los formalismos existentes atienden la interacción entre actores y sistema y entre componentes del sistema.

*Hipótesis II:* Apoyados en las nuevas tecnologías de la información y de las comunicaciones, los ambientes virtuales de trabajo colaborativo abren la posibilidad de disponer un espacio donde el encuentro virtual del equipo de trabajo permita mejorar su productividad. Sin embargo, no todas las arquitecturas propuestas para estos espacios satisfacen las funcionalidades de: gestión de tareas grupales, gestión de los artefactos conceptuales derivados del trabajo grupal, gestión de las interacciones grupales; y documentación de la trazabilidad de las interacciones de los miembros.



Tampoco se dispone de las herramientas de modelado correspondientes.

*Objetivo General:* El objetivo de este proyecto es sistematizar el conocimiento existente sobre espacios virtuales de trabajo colaborativo y formular una propuesta unificada de herramientas para su modelado y diseño.

*Objetivo específico vinculados a la Hipótesis I:*

1.- Desarrollar herramientas para el modelado y diseño de espacios virtuales para trabajo colaborativo, con énfasis en las interacciones humanas que deben soportar.

*Objetivo específico vinculados a la Hipótesis II:*

2.- Desarrollar un arquetipo patrón de arquitectura de espacio virtual que guíe el diseño de espacios virtuales dedicados al desarrollo de proyectos grupales.

3.- Desarrollar herramientas para realizar mediciones de interacción en grupos que realicen trabajo colaborativo basado en espacios virtuales.

## METODOLOGÍA DE TRABAJO

Se prevé: realizar investigación documental identificando casos de estudio, desarrollar mediante la metodología de prototipado evolutivo las versiones iniciales de los instrumentos combinado con mejora incremental, probar la versión de producción de los prototipos en casos seleccionados.

## RESULTADOS OBTENIDOS/ESPERADOS

Como resultado de este proyecto, se esperan contar al fin de su desarrollo con los siguientes instrumentos: (a) herramientas para el modelado y diseño de espacios virtuales para trabajo colaborativo con énfasis en las interacciones humanas que deben soportar, (b) un arquetipo patrón de arquitectura de espacio virtual dedicados al desarrollo de proyectos grupales, y (c) herramientas de medición de interacción humana en grupos que realicen trabajo colaborativo basado en espacios virtuales.

## FORMACIÓN DE RECURSOS HUMANOS

El grupo de trabajo se encuentra formado por dos investigadores formados, dos investigadores en formación y tres becarios alumnos de la carrera Licenciatura en Sistemas de la UNLa. En su marco se desarrolla una Tesis de Doctorado en Ciencias Informáticas, una Tesis de Maestría en Tecnología Informática y tres Trabajos Finales de Licenciatura en Sistemas.

## REFERENCIAS

- Bibbo, L., García, D., Pons, C. 2008. A Domain Specific Language for the Development of Collaborative Systems. Proceedings International Conference of the Chilean Computer Science Society (SCCC '08). Pág. 3-12. ISBN 978-0-7695-3403-9.
- Carlsen, S. 1997. Conceptual Modeling and Composition of Flexible Work Flow Models. PhD Thesis on Engineering. Information Systems Group. Department of Computer and Information Science. Norwegian University of Science and Technology. <http://www.idi.ntnu.no/~sif8060/pensum/A15-thesis-sca.pdf>. Página vigente al 21/12/10.
- Charczuk, N. 2011. Identificación de Usos Educativos de Espacios de Encuentro Virtual. 2014en preparación. Propuesta Técnica de Tesis de Maestría en Tecnología Informática Aplicada a la Educación. Facultad de Informática. Universidad Nacional de La Plata. Codirector: Mg. Darío Rodríguez.
- Drucker, P. 1988. The Coming of the New Organization. Harvard Business Review, Nber. January-February. Pág. 45-53. ISSN 0017-8012.
- Fields, B., Merriam, N., Dearden, A. 1997. DMVIS: Design, Modelling and Validation of Interactive Systems. En Design, Specification and Verification of Interactive Systems. Springer-Verlag.
- Garbarini, R. 201420122014. Diseño Conceptual de Espacios Virtuales para el desarrollo de proyectos en materias de carrera de grado 2014en preparación. Propuesta Técnica de Tesis de Maestría en Tecnología Informática Aplicada a la Educación. Facultad de Informática. Universidad Nacional de La Plata. Mg. Darío Rodríguez.
- García Carrasco, J., García del Dujo, A., López Fernández, R. 1999. Nuevas tecnologías y formación. PCWEEK. Editorial America Iberica.
- García Martínez, R. y Britos, P. 2004. Ingeniería de Sistemas Expertos. Editorial Nueva Librería. ISBN 987-1104-15-4.
- García Peñalvo, F., García Carrasco, J. 2002. Los espacios virtuales educativos en el ámbito de internet un refuerzo a la formación tradicional. Teoría de la Educación: Educación y Cultura en la Sociedad de la Información, N° 3. ISSN 1138-9737.
- Garrido, J. 2003. AMENITIES: Una Metodología para el Desarrollo de Sistemas Cooperativos Basada en Modelos de Comportamiento y Tareas. Tesis Doctoral del Departamento de Lenguajes y Sistemas Informáticos. Universidad de Granada. España.
- Isla, J., Gutiérrez, F., Paderewski, P. 2007. Una Aproximación Basada en Patrones para el Modelado Conceptual de Sistemas Cooperativos. IEEE Latin America Transactions, 5201442014: 204-210.
- Isla, J., Gutiérrez, F., Gea, M., Garrido, J. 2004. Descripción de Patrones de Organización y su Modelado con AMENITIES. Proceedings 4<sup>th</sup> Jornadas Iberoamericanas de Ingeniería del Software e Ingeniería del Conocimiento. Pág. 3-14. ISBN 978-987-1437-47-6.
- Jablonski, S., Bussler, C. 1996. Workflow Management: Modeling Concepts, Architecture and Implementation. International Thomson Computer Press ISBN 185-0322-22-8.
- Leiner, B., Cerf, V., Clark, D., Kahn, R., Kleinrock, L., Lynch, D. Postel, J., Roberts, L., Wolf, S. 1999. Brief History of the Internet. CERN Document Server. Report Number cs.NI/9901011.
- Noguera, M. 2009. Modelado y Análisis de Sistemas CSCW Siguiendo un Enfoque de Ingeniería dirigido por Ontologías. Tesis Doctoral en Informática. Departamento de Lenguajes y Sistemas Informáticos. Universidad de Granada. <http://hera.ugr.es/tesisugr/18014094.pdf>. Página vigente al 21/12/10.
- Nonaka, I. 1991. The Knowledge-Creating Company. Harvard Business Review, Nber. November-December. Pág. 96-104. ISSN 0017-8012.
- Nonaka, I. 1994. A Dynamic Theory of Organizational Knowledge Creation. Organizational Science, 5201412014: 14-37. ISSN 1526-5455.
- Nonaka, I. 2007. The Knowledge-Creating Company. Harvard Business Review, Nber. July-August. Pág. 162-171. ISSN 0017-8012.
- Rodríguez, D. 2010. Diseño Conceptual de Espacios Virtuales para la Formación de Investigadores. Propuesta Técnica de Tesis de Maestría en Tecnología Informática Aplicada a la Educación. Facultad de Informática. Universidad Nacional de La Plata.
- Rodríguez, D. 2012. Espacios Virtuales para la Formación de Investigadores. Elementos de Analisis y Diseño. Tesis de Maestría en Tecnología Informática Aplicada a la Educación. Facultad de Informática. Universidad Nacional de La Plata.
- Rodríguez, D., Bertone, R., García-Martínez, R. 2010c. Formación de Investigadores Mediada por Espacios Virtuales. Fundamentación y Prueba de Concepto. Proceedings del V Congreso de Tecnología en Educación y Educación en Tecnología. Pág. 512-421. ISBN 978-987-1242-42-9.
- Rodríguez, D., Bertone, R., García-Martínez, R. 2009. Consideraciones sobre el Uso de Espacios Virtuales en la Formación de Investigadores. Revista de Informática Educativa y Medios Audiovisuales, 6: 35-42. ISSN 1667-8338
- Rodríguez, D., Bertone, R., García-Martínez, R. 2010a. Collaborative Research Training Based on Virtual Spaces. En Key Competencies in the Knowledge Society 2014Eds. Reynolds, N. & Turcsányi-Szabó, M. 2014. IFIP Advances in Information and Communication Technology, 324: 344-353. ISBN 978-3-642-15377-8.
- Rodríguez, D., Bertone, R., Pollo-Cattaneo, F., García-Martínez, R. 2012a. Modelo Colaborativo de Formación de Investigadores. Proceedings II Jornadas de Enseñanza de la Ingeniería 2014en prensa 2014. Programa de Tecnología Educativa y Enseñanza de la Ingeniería. Secretaría de Ciencia, Tecnología y Posgrado. Universidad Tecnológica Nacional.
- Rodríguez, D., Charczuk, N., Garbarini, R., García-Martínez, R. 2012b. Trabajo Colaborativo basado en Espacios Virtuales. Proceedings II Jornadas de Enseñanza de la Ingeniería 2014en prensa 2014. Programa de Tecnología Educativa y Enseñanza de la Ingeniería. Secretaría de Ciencia, Tecnología y Posgrado. Universidad Tecnológica Nacional.
- Rodríguez, D., Pollo-Cattaneo, F., Bertone, R., García-Martínez, R. 2010b. Elementos para el Análisis y Diseño Conceptual de Espacios Virtuales de Trabajo Colaborativo Orientados a la Formación de Investigadores. Anales del XVI Congreso Argentino de Ciencias de la Computación. Pág. 364-373. ISBN 978-950-9474-49-9.
- Salazar, C. 1999. Teletabajo. Ingeniería informática, 4. ISSN 0717-4195.

## Laboratorio Virtual y Remoto para la Enseñanza de Diseño y Administración de Redes de Computadoras

Daniel Britos, Laura Vargas, Silvia Arias, Nicolás Giraudo, Guillermo Veneranda  
 dbritos@efn.uncor.edu, lvargas@efn.uncor.edu, sil\_var@hotmail.com, nicolasgiraudo@hotmail.com, guillevene@gmail.com

Laboratorio de Redes y Comunicaciones de Datos, Departamento de Computación, Facultad de Ciencias Exactas, Físicas y Naturales, Universidad Nacional de Córdoba

### RESUMEN

En las carreras de ingeniería, las asignaturas relacionadas con redes de computadoras buscan que el alumno alcance competencias en diseño, implementación y administración de redes. Esto requiere abundante práctica interrelacionada con la teoría. Dado el alto costo que tiene trabajar en laboratorios con equipos físicos y considerando que en las redes la administración se suele realizar en forma remota, se decidió implementar un laboratorio virtual de redes de computadoras. Para esto se analizaron distintas herramientas gratuitas, una vez hecha la selección de las mismas, se sometió el conjunto a pruebas a fin de evaluar su rendimiento. De este modo, se buscó complementar prácticas en equipos físicos y virtuales simultáneamente, así como permitir la interacción entre los mismos. La implementación seleccionada se plantea como alternativa a la utilización de las herramientas que simulan componentes, ya que estas solo ofrecen los comandos que el desarrollador decidió incluir. En este caso, se hacen las prácticas con sistemas operativos de internetworking originales, accediendo así a todos los comandos y parámetros posibles. Por otra parte, el acceso en forma remota a los equipos reproduce una situación real y es un modo flexible respecto a necesidades de alumnos y docentes.

**Palabras clave:** emulación, virtualización, IOS, GNS3, redes de computadoras.

### CONTEXTO

El presente trabajo se realiza en el ámbito del Laboratorio de Redes y Comunicaciones de Datos que funciona en el Departamento de Computación perteneciente a la Facultad de Ciencias Exactas, Físicas y Naturales de la Universidad Nacional de Córdoba, cuyo director es el Mag. Ing. José Daniel Britos, quien es profesor ordinario titular de las asignaturas Redes de Computadoras y Sistemas de Computación. Es continuación del presentado en el año 2011, durante las Jornadas de Enseñanza de la Ingeniería (JEIN) que tuvieron lugar en la Facultad Regional Buenos Aires de la Universidad Tecnológica Nacional "Las TICs en la enseñanza de TICs", y publicado en el Libro de Artículos Completos (e-book), ISBN 978-950-42-0138-0. Es una de las líneas de investigación planteadas para el Proyecto Laboratorios Remotos para la Enseñanza de Redes de Computadoras, aprobado por SECyT, UNC, período 2010-

2011, código 05/M147.

En el año 2012, Nicolás Giraudo y Guillermo Veneranda realizaron la tesis de grado que les permitió alcanzar el título de Ingenieros en Computación: "Implementación de un laboratorio virtual con acceso remoto para la enseñanza de Redes de Computadoras".

Los resultados obtenidos se implementarán este año en el dictado de las asignaturas del área Redes.

### INTRODUCCIÓN

Las asignaturas relacionadas con redes de computadoras poseen una extensa base teórica de la que se abreva según la orientación de la carrera en la que se dictan. En particular, en las carreras de ingeniería se espera que el estudiante adquiera competencias en diseño, implementación y administración de redes por lo que tiene especial importancia la interrelación de la práctica. Se plantea la dificultad de mantener actualizados los laboratorios en cuanto a equipos, ya que rápidamente se vuelven obsoletos y cada vez son más costosos sumada a la necesidad de sostener personal suficiente para mantener las prácticas.

Por otro lado, a medida que pasa el tiempo, el término "virtualización" va teniendo más repercusión en todos los ámbitos universitarios y empresariales dada su gran evolución y constante mejora.

Así, la posibilidad de implementar una herramienta para la simulación de redes de computadoras con dispositivos de red emulados dio el impulso inicial para esta investigación. Los laboratorios de redes virtuales le acercan a los ingenieros y/o administradores de redes una nueva alternativa para poner en práctica los conocimientos teóricos adquiridos y experimentar todas las funcionalidades de los Sistemas Operativos de Internetworking.

En lo que sigue se definirán algunos conceptos fundamentales para la implementación de un laboratorio virtual.

### VIRTUALIZACIÓN

La *virtualización* es la creación de una versión virtual de algún recurso tecnológico [1]. La misma oculta las características físicas de los recursos tecnológicos a las aplicaciones haciendo que un solo recurso de hardware parezca como múltiples recursos virtuales y viceversa, que

múltiples recursos físicos aparezcan un único recurso virtual.

La *virtualización de plataforma o de hardware* es la creación de una Máquina Virtual que actúa como una computadora real con un Sistema Operativo (SO). El Software ejecutado en estas Máquinas Virtuales está separado de los recursos de hardware subyacentes.

Una *Máquina Virtual* (VM) es una representación lógica en software de una computadora que ejecuta programas como una máquina real. Al desacoplar el hardware físico del SO, la virtualización de plataformas permite compartir los recursos tecnológicos entre múltiples VMs, cada una corriendo su propio SO, aumentando así la tasa de utilización del hardware subyacente [2].

Una *máquina host* es la máquina física en la cual se lleva a cabo la virtualización, y una *máquina guest* es la VM. Las palabras host y guest se utilizan también para distinguir el software que se ejecuta en la máquina física del software que se ejecuta en la VM.

Un *hypervisor o Virtual Machine Monitor (VMM)* es el software o firmware que crea una capa de abstracción entre el hardware del host y el SO del guest, dividiéndose el recurso tecnológico en uno o más entornos de ejecución. El VMM maneja, gestiona y arbitra los cuatro recursos principales de una computadora (CPU, Memoria, Almacenamiento y Conexiones de Red) distribuyendo los mismos en forma estática o dinámica entre todas las VMs definidas en el computador central. Los hypervisors se pueden clasificar en dos tipos [3]:

- Hypervisor de tipo 1 (nativo o bare-metal): se ejecuta directamente sobre el hardware del host. (Citrix XenServer, VMware ESX/ESXi y Microsoft Hyper-V).
- Hypervisor de tipo 2 (sobre host): se ejecuta dentro de un entorno de SO convencional. (KVM y VirtualBox).

Un *emulador* es un hypervisor que imita las funciones de una computadora (guest) en otra computadora diferente (host), independientemente de cómo sea el hardware físico, de modo que el comportamiento imitado se asemeja mucho al comportamiento de la computadora real. Se destaca por hacer énfasis en la reproducción del comportamiento característico y de aspectos funcionales de otro dispositivo de hardware, dejando de lado aspectos no funcionales como el tiempo de respuesta o el rendimiento.

Una *red virtual* es una red de computadoras que se compone, al menos en parte, de enlaces de red virtuales. Un enlace de red virtual no se implementa mediante una conexión física (por cable o inalámbrica) entre dos dispositivos de computación, sino que se lleva a cabo utilizando métodos de virtualización. En particular, el interés se ha centrado en redes virtuales basadas en dispositivos que conectan VMs en un hypervisor y también con el medio exterior y no en las redes virtuales basadas en protocolo como VLANs (Red de Área Local Virtual) y VPN (Red Privada Virtual), entre otras.

Las principales *ventajas de la virtualización de*

*plataforma* son:

- Permite la coexistencia de múltiples SO en el mismo computador, aislados unos de otros.
- La VM puede proveer un ISA (Instruction Set Architecture) diferente al de la máquina física.

**DESARROLLO DEL LABORATORIO VIRTUAL**

De ahora en adelante, cuando se hable de redes virtuales se considerarán exclusivamente las basadas en dispositivos virtuales, y no en protocolos, cuyas componentes son:

- Switch virtual (vSwitch)
- Adaptador Ethernet Virtual (VEA)
- Servidor DHCP y DNS virtual
- Enlace virtual

Los componentes clave son los VEAs utilizados por las VMs y los vSwitches que conectan las VMs entre sí. Los vSwitches permiten a las VMs, comunicarse entre sí utilizando los mismos protocolos usados por los switches físicos. Un vSwitch es una simple construcción de software que generalmente está dentro del hypervisor pero a veces se presenta al SO host como una interfaz de red. Una VM puede ser configurada con uno o varios VEAs, cada uno de los cuales tiene su propia dirección IP y MAC. Como resultado, desde el punto de vista de las redes, las VMs tienen las mismas propiedades que las máquinas físicas [4].

Para cada VM, el hypervisor puede crear varios VEAs, seleccionando en cada uno tanto el tipo de hardware que será virtualizado como el modo de virtualización en cual estará operando en relación con la red física del host.

**SELECCIÓN DE HERRAMIENTAS PARA LA VIRTUALIZACIÓN**

En una primera etapa, se analizaron varias herramientas para simular redes de computadoras.

Herramienta	Packet Tracer	Boson NetSim	VNX	GNS3	Netkit	Marionet	CORE
Licencia	Privativo	Privativo	GPL	GPLv2	GPL	GPLv2	BSD
Última versión	5.3.3 Feb. 2012	8.0 Nov. 2011	2.0 Mar. 2012	0.8.2 Mar. 2012	2.8 May. 2011	0.6 Feb. 2011	4.3 Mar. 2012
Sistemas Operativos soportados	Windows Linux	Windows	Linux	Windows Linux Mac OS X	Linux	Linux	Linux FreeBSD
Soporta los TP	Si	Si	No	Si	Si	Si	Si
Fácil de utilizar	Si	Si	No	Si	No	Si	Si
IOS Cisco GUI	No	No	No	Si	No	No	No
Conexión con equipos reales	No	No	Si	Si	Si	Si	Si

Tabla 1 - Herramientas para virtualización de redes de computadoras

Luego del análisis realizado sobre las herramientas disponibles (Tabla 1), se decide utilizar a GNS3 como GUI y Dynamips como emulador ya que cumplen con los siguientes puntos:

- Ambos son de distribución gratuita.
- GNS3 posee una interfaz gráfica simple y fácil de utilizar.
- Dynamips emula IOS Cisco reales. Puede ejecutarse en hosts independientes y permitir balanceo de carga.



- Pueden ejecutarse en un SO potente y robusto como lo es Linux.
- Permiten la conexión con dispositivos reales.
- Ambos se encuentran en pleno desarrollo y con constantes mejoras.

Respecto a la virtualización de computadoras, existe una amplia oferta de software. VirtualBox es la herramienta que mejor se adapta a los requerimientos planteados. Cabe destacar que VirtualBox puede ser instalado en un servidor dedicado, tiene la capacidad de clonar VMs, está integrado a GNS3 y mediante phpVirtualBox se puede acceder a las distintas VMs a través de un navegador web. Se elige la herramienta VirtualBox, aun cuando no se trata de un software de distribución libre y phpVirtualBox no maneja el concepto de propiedad de una VM. VMware Server fue en algún momento el gran candidato pero como el proyecto se ha discontinuado se lo descartó.

Herramienta	VirtualBox	Vmware Server	ESXi	XenServer	QEMU
<b>Hypervisor</b>	Tipo 2	Tipo 2	Tipo 1	Tipo 1	Tipo 2
<b>Entorno subyacente en el servidor</b>	Windows Linux	Windows Linux	Hardware	Hardware	Windows Linux Mac
<b>Licencia</b>	VirtualBox PUEL	Gratuita	Gratuita	Gratuita	GNU GPL
<b>Estado del proyecto</b>	Activo	Discont.	Activo	Activo	Activo
<b>Soporte de software</b>	Si	No	Si	Si	Si
<b>Admin. remota</b>	CLI WUI	WUI	GUI	CLI GUI	CLI WUI
<b>Entorno subyacente en el cliente</b>	Windows Linux Browser	Browser	Windows	Windows Linux	Windows Linux Browser
<b>Integración con GNS3</b>	Si	No	Si	Si	Si
<b>Interfaz de red virtual accesible</b>	Si	Si	No	No	Si
<b>Permisos y propiedad de las VMs</b>	No	No	Si	Si	No
<b>Comp. Con hardware</b>	Alta	Alta	Baja	Baja	Alta
<b>Manipulación de imágenes de disco</b>	Alta	Alta	Baja	Baja	Alta
<b>Asistente para clonación</b>	Si	No	No	No	No
<b>Snapshots</b>	Si	Si	Si	Si	Si

Tabla 2 - Herramientas para virtualización de computadoras

## INTEGRACIÓN DE LAS HERRAMIENTAS

En un paso posterior, se integraron las herramientas. La emulación de routers Cisco se realiza con el programa Dynamips mientras que para virtualizar los dispositivos finales se utiliza VirtualBox. Además, VirtualBox, hypervisor tipo 2, puede virtualizar routers para luego ser utilizados por GNS3 [6] como componentes de una topología virtual. Como interfaz web para la administración remota de las VMs se usa phpVirtualBox.

VBoxWrapper es un módulo servidor escrito en Python y usado para controlar VMs. En una topología de red creada con GNS3, una VM debe tener configurado su adaptador de red virtual en modo adaptador genérico para que pueda comunicarse con otro dispositivo [7].

PhpVirtualBox un proyecto independiente que consiste en una WUI (Web User Interface) para administrar y monitorear las VMs por medio del navegador web [8].

Cuando un usuario accede a phpVirtualBox, obtiene control total sobre todas las VMs dentro del entorno de VirtualBox. PhpVirtualBox realiza todas las acciones administrativas a través de vboxwebsrv, un servidor SOAP (Simple Object Access Protocol) distribuido junto con VirtualBox [9].

El servicio web es brindado por un archivo independiente llamado vboxwebsrv que actúa como un servidor HTTP. Vboxwebsrv es un programa de consola en modo texto que viene incluido con el paquete binario estándar de VirtualBox. Además, acepta conexiones SOAP y las procesa de manera remota o desde la misma máquina. Una vez que el servicio web se ha iniciado, actúa como un front end para el VirtualBox del usuario que está ejecutando vboxwebsrv [10].

## RESULTADOS

### ARQUITECTURA PROPUESTA

Respecto al funcionamiento conjunto, la forma más fácil de trabajar con estas herramientas es ejecutándolas en la misma computadora. Previamente a agregar un componente en GNS3 debe configurar su imagen IOS. Desde el momento en que se agrega el primer router a la topología, GNS3 ejecuta Dynamips automáticamente en segundo plano y queda a la escucha en el puerto correspondiente. Luego de que un router ha sido encendido es posible abrir su consola. De igual forma ocurre con el programa VBoxWrapper cuando lo que se agrega es una VM. Para que una VM pueda ser añadida hay que pre-configurarla en GNS3. Al encender una VM se abre su consola de VirtualBox.

La emulación de routers necesita gran cantidad de recursos de procesamiento y memoria. Muchas veces esos requerimientos no son alcanzados con el hardware de una computadora. GNS3 ofrece la posibilidad de configurar hypervisors externos para delegar esta carga de trabajo. Un hypervisor externo es una instancia de Dynamips ejecutándose en otra computadora. Al momento de crear la topología se puede seleccionar explícitamente en cual hypervisor es emulado cada router. Pero también se puede configurar a GNS3 para que realice automáticamente un balanceo de carga entre los hypervisors.

La virtualización de computadoras también requiere una gran cantidad de recursos computacionales. Distribuir esta carga de trabajo requiere que GNS3 se pueda comunicar con una instancia de VirtualBox corriendo en otra computadora. La comunicación se logra ejecutando VBoxWrapper en el servidor de virtualización y configurando GNS3 con la dirección IP del mismo. Una vez realizado esto, será posible prender y apagar las VMs remotamente desde GNS3. Para poder iniciar una sesión remota en estos guest es necesario instalar y configurar phpVirtualBox. Este último requiere que VBoxWebSrv esté ejecutándose para establecer una comunicación con VirtualBox. También se pueden configurar varios servidores de virtualización y dividir la carga de trabajo manualmente entre ellos.



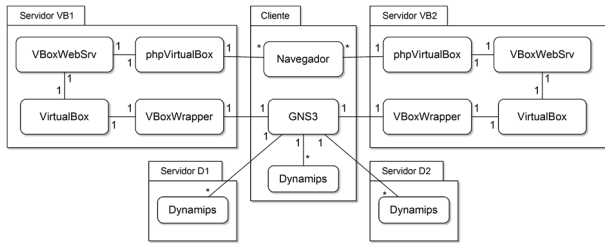


Figura 1 - Dos servidores VirtualBox y dos servidores externos

En la Figura 1 se observa la arquitectura planteada.

**PARTICULARIDADES DEL MODELO ELEGIDO**

GNS3 permite conectar dispositivos virtuales con equipos reales. La conexión se realiza utilizando nodos tipo nube que pueden comunicar redes virtuales con redes físicas. Es necesario ejecutar GNS3 como root para configurar correctamente la nube, además se debe emular en forma local al dispositivo.

Se puede concluir que la utilización conjunta de las herramientas elegidas permite virtualizar redes complejas de computadoras. Además, la arquitectura formada por estas herramientas es altamente escalable, por lo tanto, los clientes pueden utilizar varios servidores externos como se observa en la Figura 2.

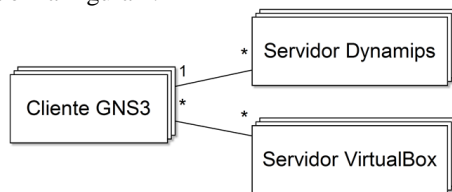


Figura 2 - Clientes y Servidores.

GNS3 no tiene las funcionalidades necesarias para permitir a varios clientes trabajar sobre una misma topología de red. Sin embargo, si la topología se divide en partes y cada una es simulada localmente en un cliente GNS3 distinto, entonces, es posible conectar las partes y trabajar de manera distribuida sobre la misma topología. WinPcap en Windows y libpcap en Linux son las herramientas que permiten a GNS3 capturar, inyectar, analizar y construir paquetes en la red física del host, como se muestra en la Figura 3.

Con VirtualBox es posible crear VMs con SOs Linux adaptados para funcionar como routers virtuales.

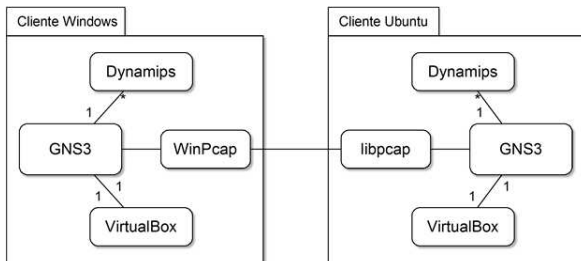


Figura 3 - Comunicación entre clientes GNS3

**PRUEBAS DE RENDIMIENTO**

En una segunda etapa, se realizaron pruebas para evaluar

el rendimiento. El Prototipo es el servidor encargado de emular los routers. Debido a que este servicio es escalable horizontalmente, se configuraron dos hosts para que lleven a cabo las tareas de emulación. En las Tablas 3 y 4 se describen respectivamente las características de hardware del Servidor 1 y Servidor 2.

Componente	Cantidad	Modelo
Procesador	1	AMD Phenom II X4 955 3200 MHz
Memoria	2	G.Skill RipjawsX DDR3-1600 PC3-12800 2 GB
Motherboard	1	GA-880GM-UD2H
Disco Rígido	1	250 GB

Tabla 3 - Hardware Servidor 1

Componente	Cantidad	Modelo
Procesador	1	Intel Core 2 Duo E6750 2660 MHz
Memoria	2	G.Skill PI Black Edition DDR2-800 PC2-6400 4 GB
Motherboard	1	Intel DP35DP
Disco Rígido	1	500 GB

Tabla 4 - Hardware Servidor 2

Se resalta la escalabilidad horizontal que tiene la herramienta porque permite proyectar el conjunto de servidores necesarios para emular los 90 routers requeridos. La memoria RAM es más fácil de escalar verticalmente en cambio el poder de procesamiento es más fácil de escalar horizontalmente.

Se planteó una prueba de carga consistente en someter al prototipo Servidor 1 a una carga de trabajo equivalente a una clase práctica en el Laboratorio de Redes y Comunicaciones de Datos para 7 grupos de alumnos.

Como resultado de todas las pruebas de estrés realizadas se pudo concluir que la terminación del proceso Dynamips, por motivo de una falla interna o por la llegada de una señal externa, detiene la simulación en curso y no puede restablecerse, aunque la topología puede volver a usarse. Además, un aumento considerable del tráfico de red que atraviesa los routers, detiene la simulación y corrompe la topología de manera que no puede volver a utilizarse. Por otra parte, la pérdida de conectividad entre el servidor y los clientes no detiene la simulación, una vez restablecida dicha conectividad, los dispositivos pueden volver a comunicarse entre sí.

En base a las pruebas realizadas sobre el prototipo se determinó que con 12 GB de RAM se puede suplir la necesidad de memoria que implica la emulación de 90 routers. Las pruebas también reflejan que el Servidor 1 puede soportar la carga de procesamiento de emular 30 routers. Por lo tanto, para emular 90 routers se necesita un poder de cómputo equivalente al de 3 servidores similares. GNS3 y VirtualBox permiten crear topologías complejas de redes de computadoras sin necesidad de contar con equipos de interconexión ni gran cantidad de recursos computacionales.

Respecto a las aplicaciones, se pudo observar el alto grado de escalabilidad que tienen las herramientas seleccionadas. Montando 3 servidores similares al

prototipo, cada uno con 4GB de RAM, es posible emular 90 routers.

No obstante, no es necesario comprar 3 computadoras, en una primera fase pueden utilizarse equipos en existencia de menor potencia hasta alcanzar los requerimientos. Luego de evaluar el grado de aceptación de la herramienta, en una etapa más avanzada o cuando se considere conveniente, se recomienda adquirir un equipo servidor de características similares a las mostradas en la Tabla 5.

Componente	Cantidad	Modelo	Precio
Procesador	2	AMD Opteron 4280 Valencia 2.8 GHz	\$ 2.475
Memoria	4	4 GB ECC Registered DDR3 1333 Mhz	\$ 693
Motherboard	1	Supermicro H8DCT-HLN4F-B two Socket C32	\$ 1.742
Disco Rígido	1	Western Digital 1 TB 7200 RPM 64 MB Cache SATA 6.0 Gb/s	\$ 500
<b>Total</b>			<b>\$ 5.410</b>

Tabla 5 - Hardware del servidor recomendado

## CONCLUSIONES

Las herramientas GNS3 y VirtualBox permiten crear topologías complejas de redes de computadoras sin necesidad de contar con equipos de interconexión ni gran cantidad de recursos computacionales. A continuación se listan las conclusiones detalladas con respecto a las aplicaciones utilizadas:

- GNS3 es una herramienta que está en continuo desarrollo, agregando nuevas funcionalidades, arreglando bugs e integrándose con nuevas herramientas. Esto obligó a cambiar el rumbo del Proyecto Integrador en varias oportunidades, pero también sirvió para evitar los problemas de integración encontrados con determinados hypervisors.
- GNS3 permite a varios usuarios trabajar sobre la misma topología.
- VirtualBox es una herramienta potente que permite experimentar con los SOs actuales más utilizados. Además, esta integrado a GNS3.
- PhpVirtualBox no está diseñado para funcionar en un entorno de hosting donde se necesita el concepto de propiedad de una VM.
- Los routers emulados con Dynamips no soportan un alto tráfico de red. En un ambiente educativo posibilita experimentar las funcionalidades de las IOS permitiendo a los estudiantes afianzar sus conocimientos.

## OBJETIVOS Y LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

### A. Objetivo General

Implementar un Laboratorio Virtual para hacer prácticas de redes de computadoras con acceso al mismo en forma remota, que posibilite la interacción directa entre los ingenieros y/o administradores de redes y los IOS de los equipos de interconexión de redes.

### B. Objetivos Específicos

- Brindar una solución al problema del costo creciente de laboratorios de redes de computadoras.
- Crear, configurar y dar acceso a Máquinas Virtuales con Sistemas Operativos de escritorio.

- Configurar un servicio web que permita el acceso a las Máquinas Virtuales a través de un navegador.
- Diseñar topologías de laboratorios de redes virtuales de computadoras.

### C. Trabajos Futuros

A partir del 2013, los alumnos podrán utilizar la arquitectura de laboratorios remotos planteada. Se evaluará in situ el rendimiento y se comparará con los resultados obtenidos en la primera etapa.

En cuanto a la investigación, se espera agregar a phpvirtualbox los conceptos de: propiedad de una VM, privilegios, roles de usuarios y grupos de usuarios y modificar el código de GNS3 para permitir a varios usuarios trabajar sobre la misma topología.

Se comparará una herramienta con fines educativos contra otra, diseñada para trabajar en un entorno de producción .

## FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo está formado por tres investigadores categorizados por el Programa Nacional de Incentivos (categorías III y IV), dos de ellos Especialistas en Docencia Universitaria y dos Magister en Ciencias de la Ingeniería, mención Comunicaciones y Redes, respectivamente. También integran el equipo, dos alumnos que realizaron su tesis final de grado en la carrera de Ingeniería en Computación. Además se incorporará al equipo este año un Ingeniero en Computación, encargado de las prácticas en laboratorio.

Esta línea de trabajo en el Laboratorio de Redes y Comunicaciones de Datos permitirá la capacitación de los alumnos de las carreras de Ingeniería en Computación e Ingeniería Electrónica en forma flexible y sobre equipos actuales. Esto tendrá un efecto multiplicador, permitiendo egresados más capacitados que brinden mejores servicios a la industria y a la sociedad en un área en creciente desarrollo.

## REFERENCIAS

- [1] Turban, E; King, D; Lee, J; Viehland, D. *Chapter 19: Building E-Commerce Applications and Infrastructure*. Electronic Commerce A Managerial Perspective, 5ta Edición. Prentice-Hall, pp. 27. 2008.
- [2] *Virtualization in education*. IBM Global Education. <http://www.07.ibm.com/solutions/in/education/download/Virtualization%20in%20Education.pdf>, 2007.
- [3] Goldberg, R.P. *Architectural Principles for Virtual Computer Systems*. Harvard University [http:// flint.cs.yale.edu/cs428/doc/goldberg.pdf](http://flint.cs.yale.edu/cs428/doc/goldberg.pdf). pp. 22-26, 1973.
- [4] *VMware Virtual Networking Concepts*. [http://www.vmware.com/files/pdf/virtual\\_networking\\_concepts.pdf](http://www.vmware.com/files/pdf/virtual_networking_concepts.pdf), 2007.
- [5] *Chapter 6. Virtual networking*. Oracle VM VirtualBox User Manual. Virtualbox.org. <http://www.virtualbox.org/manual/ch06.html>, 2012.
- [6] Graphical Network Simulator 3 <http://www.gns3.net/>, 2012.
- [7] *Welcome to VirtualBox.org!*. <https://www.virtualbox.org/>, 2012.
- [8] *VirtualBox Web Console*. <http://code.google.com/p/vboxweb/>, 2012.
- [9] *phpVirtualBox* <http://code.google.com/p/phpvirtualbox/>, 2012.
- [10] Oracle VM VirtualBox: Programming Guide and Reference. Version 4.1.18, <http://download.virtualbox.org/virtualbox/SDKRef.pdf>, 2012.

# Análisis del Aprendizaje usando el Nivel de la Incertidumbre en la Adquisición de Conocimiento

Constanza R. Huapaya, Francisco A. Lizarralde, Graciela M. Arona, Jorge R. Vivas

Grupo de Investigación en Inteligencia Artificial (integrante del CIMEPB)

Departamento de Matemática-Facultad Ingeniería

Universidad Nacional de Mar del Plata

Juan B. Justo 4302. 7600. Mar del Plata

223-4816600 int 259

[constanza.huapaya@gmail.com](mailto:constanza.huapaya@gmail.com)

[francisco.lizarralde@gmail.com](mailto:francisco.lizarralde@gmail.com)

[graron@fi.mdp.edu.ar](mailto:graron@fi.mdp.edu.ar)

[jvivas@mdp.edu.ar](mailto:jvivas@mdp.edu.ar)

## Resumen

Como continuación de la investigación sobre el uso de la Minería de Datos Educacional (MDE) para descubrir nuevo conocimiento basado en datos de la actividad del estudiante se ha elegido estudiar la **incertidumbre** en la medición de la adquisición de conocimiento a fin de mejorar la estimación de perfiles cognitivos del estudiante sobre los cuales se está investigando desde hace dos años.

El conocimiento adquirido por los estudiantes es, en general, imperfecto. El grado de profundidad en la adquisición de habilidades cognitivas varía durante el aprendizaje. Se ha desarrollado un modelo difuso para medir la incertidumbre en la adquisición del conocimiento de los estudiantes.

En base a los seis objetivos educacionales de la taxonomía de Bloom revisada (que han sido llevados a tres niveles) se han definido conjuntos difusos asociados a cada uno de ellos. Se estudia la posibilidad de los perfiles de cada

estudiante en cada uno de los tres niveles propuestos. En base a esa información se puede calcular la incertidumbre posibilística total a fin de analizar el nivel adquisición de conocimiento de grupos de estudiantes.

La plataforma que usamos es un Ambiente Virtual de Aprendizaje. Con esos datos podemos monitorear características de los cursos de la plataforma.

**Palabras clave:** incertidumbre, modelos difusos, minería de datos educativos.

## Contexto

La investigación descrita en esta presentación se desarrolla en el marco del Proyecto “Ambientes virtuales de aprendizaje para la enseñanza de la ingeniería” perteneciente al Departamento de Matemática de la Facultad de Ingeniería (integrante del Centro de Investigación en Procesos Básicos, Metodologías y Educación (CIMEPB, Facultad

de Psicología UNMDP) y aprobado por la Universidad Nacional de Mar del Plata.

## Introducción

Generalmente las representaciones del mundo real son imperfectas y esa imperfección se traslada a los sistemas de información. Podemos comprobar que cuando se trabaja con modelos del mundo real, generalmente lo hacemos con aproximaciones. En contraposición, puede considerarse que la información es perfecta cuando es consistente, precisa y cierta. El término **imperfección de los datos** se usa aquí como el concepto más general, que incluye a los demás (inconsistencia, imprecisión e incertidumbre)[1] [2].

Información imperfecta	Imprecisión	Varios mundos satisfacen la sentencia	
	Inconsistencia	Conclusión incoherente	
	Incertidumbre	vaguedad (=borrosidad)	
		ambigüedad	no-especificidad conflicto

Figura 1: taxonomía de la información imperfecta

En la figura 1 se aprecia una clasificación de la información imperfecta. La **imprecisión** y la **inconsistencia** son propiedades asociadas al contenido de una sentencia (sin importar si algún mundo real es compatible con la información) mientras que la **incertidumbre** nace de la falta de información del mundo real para decidir si la sentencia que constituye la información es verdadera o falsa. La incertidumbre es una propiedad de la relación entre la información y nuestro conocimiento de la realidad. En una visión restringida puede verse a la incertidumbre como una deficiencia de la información, mientras que la información se percibe como la capacidad de reducir la incertidumbre.

Las sentencias pueden verdaderas o falsas, pero la **incertidumbre** aparece cuando el conocimiento que se posee acerca del mundo real no nos permite decidir si la sentencia es

verdadera o falsa. Si la certidumbre es el conocimiento completo sobre el valor de verdad de los datos, la incertidumbre es el conocimiento parcial sobre el valor de verdad de la información

La incertidumbre se refiere, entonces, a un estado de creencia parcial, modelado mediante el agregado a cada proposición (o evento o conjunto de posibles estados) de un peso que caracteriza nuestra confianza en la verdad de tal proposición (u ocurrencia del evento, o la afirmación de que el conjunto de estados posibles contiene al estado real de casos). La incertidumbre, además de originarse en la imprecisión, puede deberse a la inconsistencia o a la aleatoriedad.

Según Klir [3][4][5] existen dos grandes categorías de incertidumbre: la **vaguedad** y la **ambigüedad**. En general, la **vaguedad** emerge cuando es difícil encontrar límites precisos entre los mundos posibles, motivo por el cual se dificulta la elección del usuario sobre cuál de ellos elegir (*borrosidad, nebulosidad, difuso* aparecen como conceptos asociados a vaguedad). La **ambigüedad** se manifiesta cuando el usuario tiene dos o más posibilidades para elegir una opción, esto es, no puede distinguir una opción de otra. La ambigüedad está conectada a conceptos como *no-especificidad y relación uno-a-muchos*.

La **medida** de la vaguedad o borrosidad ha sido uno de temas asociados con el desarrollo de la teoría de **conjuntos difusos**. En este contexto existen en la literatura diversas medidas de borrosidad. En general, una medida de la borrosidad es una función

$$f : \mathcal{P}(X) \rightarrow \mathfrak{R}$$

Donde  $\mathcal{P}(X)$  es el conjunto de todos los subconjuntos de  $X$  y  $f$  asigna  $f(A)$  a cada subconjunto difuso  $A$  de  $X$  caracterizando el grado de borrosidad de  $A$ .



La **medida** de la ambigüedad [6] vista como no-especificidad, disonancia y confusión ha sido estudiada en el contexto de las medidas de la plausibilidad y las medidas de creencia. En nuestro estudio nos interesa tratar con el principio de invariancia de la incertidumbre. En particular trabajamos con la incertidumbre posibilística total vista como suma de la no-especificidad y el conflicto.

Volviendo a nuestra investigación, se ha considerado muy pertinente el análisis de la **incertidumbre** asociada al proceso de enseñanza/ aprendizaje. En particular investigamos la relación entre la incertidumbre y el diagnóstico cognitivo [7,8, 9,10, 11].

En primer lugar, consideremos la evaluación subjetiva que llevan a cabo los docentes/evaluadores: se puede comprobar que existe el fenómeno de la **vaguedad** cuando se trata de analizar el nivel de adquisición del conocimiento por parte de los alumnos (es difícil precisar ese nivel de adquisición). Ha sido desarrollado [12,13] un modelo de diagnóstico basado en Lógica Difusa [14,15, 16] cuyo esquema se aprecia en la figura 2.

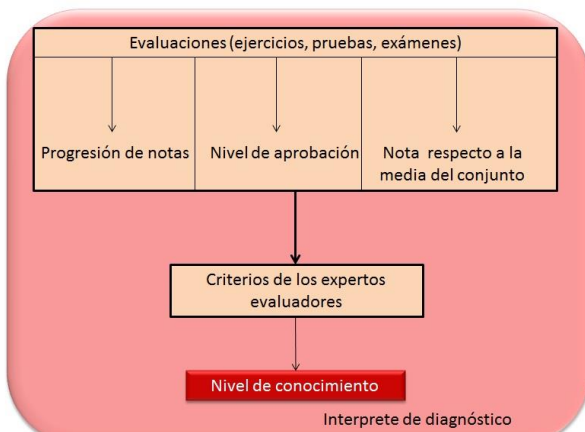


Figura 2: Modelo base del diagnóstico

En segundo lugar, examinemos características cognitivas comunes de los alumnos. A tal fin se han definido perfiles de aprendizaje de grupos de estudiantes usando el intérprete de diagnóstico de la figura 2. En este caso se

presenta el fenómeno de la **ambigüedad** (no es específico a que perfil pertenece cada alumno). Los perfiles están basados en la taxonomía revisada de Bloom [17,18], ver figura 3.

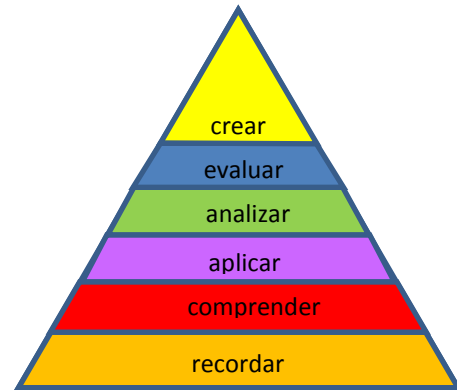


Figura 3: Taxonomía revisada de Bloom

A partir de la taxonomía se han construido tres niveles de complejidad creciente: A (recordar y comprender), B (aplicar y analizar) y C (evaluar y crear). Cada uno de estos niveles son valorados usando 5 variables lingüísticas para cada uno de los niveles (muy bajo (MB), bajo (B), medio (M), alto (Al) y muy alto (MA)). Ejemplo de un perfil es la terna (MA,M,B) indicando que el alumno ha alcanzado el nivel muy alto en el nivel A, medio en el nivel B y bajo en el nivel C. En la figura 4 se aprecia un esquema del cálculo de la incertidumbre en el aprendizaje de un grupo de estudiantes.

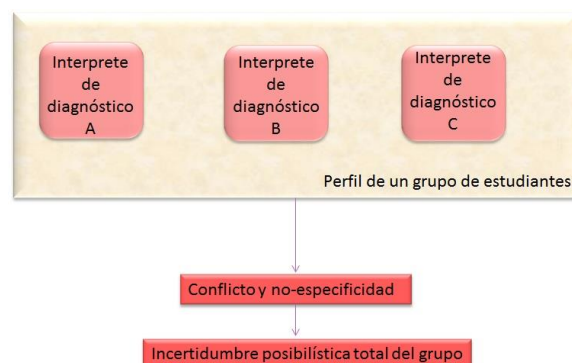


Figura 4: Modelo de medición de incertidumbre

Para construir, automáticamente, los perfiles se usa el Intérprete del Diagnóstico aplicado a cada uno de los niveles para cada de los

alumnos del grupo. Con esta información se estudia la posibilidad de todos los posibles perfiles para finalmente calcular la **incertidumbre posibilística total**. Cuanto más bajo sea este valor, más alto es el nivel de adquisición de conocimiento del grupo de estudiantes.

La presente investigación está enmarcada dentro de los métodos que estamos desarrollando para nuestra biblioteca de algoritmos propios de la Minería de Datos Educacional [19,20,21,22]. Ya están desarrolladas técnicas de clustering difuso y modelos difusos para el análisis de adquisición de conocimiento.

## Líneas de investigación y desarrollo

- *Uso de un modelo difuso para analizar la incertidumbre en el proceso de aprendizaje de un grupo de estudiante.*
- *Análisis de la incertidumbre en la construcción de perfiles de los estudiantes.*
- *Construcción de un perfil cognitivo del estudiante basado en los datos filtrados por técnicas de Minería de Datos Educacional.*

## Resultados y Objetivos

- Se ha desarrollado un modelo de perfiles cognitivos considerando una variable lingüística de salida (nivel de conocimiento) y tres variables lingüísticas de entrada (progresión de notas, nivel de aprobación de las pruebas y nota final respecto a la media del curso). El modelo se completa con 27 reglas difusas que capturan la experticia de los profesores. Este modelo ha sido encapsulado como un intérprete de diagnóstico.

- Se han desarrollado representaciones matemáticas del proceso de aprendizaje basadas en la taxonomía de Bloom a fin de afianzar el perfil cognitivo individual.
- Se ha desarrollado un espacio tridimensional que se genera a partir de la evaluación de tres habilidades cognitivas de cada estudiante: comprensión del problema, planteo general de la solución y eficiencia de la solución. Posteriormente, se aplicaron diversas técnicas de clustering y se analizaron los resultados a fin de encontrar los perfiles del grupo.
- Se ha desarrollado una técnica de medida de la incertidumbre para un grupo de estudiantes.

## Formación de Recursos Humanos

El equipo de trabajo está integrado por especialistas en Informática, en Psicología y estudiantes de Ingeniería y Psicología. Actualmente se están desarrollando tres tesis de la Maestría en Tecnología Informática Aplicada a Educación de la Facultad de Informática de la Universidad Nacional de La Plata y una tesis de doctorado de la misma Facultad.

## Referencias

- [1] Smets Ph.(1997) Imperfect information : Imprecision - Uncertainty. Uncertainty Management in Information Systems. From Needs to Solutions. A. Motro and Ph. Smets (eds.), Kluwer Academic Publishers. 225-254.
- [2] Smets Ph.(1998) Probability, Possibility, Belief: Which and Where ? Handbook of Defeasible Reasoning and Uncertainty Management Systems. Gabbay D. and Smets Ph. (Series Eds). Ph. Smets (Vol. eds.), Vol. 1: Quantified Representation of Uncertainty & Imprecision, Kluwer, Dordrecht 1-24.
- [3] Klir G. J. (1995 ) Principles of uncertainty: What are they? Why do we need them? Fuzzy Sets and Systems. (74)15-31.
- [4] Klir G.J. (1987) Where do we stand on measures of uncertainty, ambiguity, fuzziness and the like?. Fuzzy Stes and Systems . (24),141-160.

- [5] Klir G.J. (2006) *Uncertainty and Information. Foundations of Generalized Information Theory* . John Wiley & Sons.
- [6] Shafer G. (1976). *A Mathematical Theory of Evidence*. Princeton Univ. Press. Princeton.
- [7] Ma.J. y Zhou D. (2000) Fuzzy Sets Approach to the Assessment of Student-centered Learning. *IEEE Transactions on Education*, 43(2), 237-241.
- [8] Jeong H. y Biswas G. (2008). Mining Student Behavior Models in Learning-by-Teaching Environments. *The 1st International Conference on Educational Data Mining*. 127-136.
- [9] Grigoriadou M., Kornilakis H., Papanikolaou K. y Magoulas G. (2002) Fuzzy Inference for Student Diagnosis in Adaptive Educational Hypermedia. *Methods and Applications of Artificial Intelligence. Lecture Notes in Computer Science*, , Volume 2308/2002.
- [10] Stathacopoulou R., Grigoriadou M., Samarakou M., Mitropoulos D.(2007) Monitoring students' actions and using teachers' expertise in implementing and evaluating the neural network-based fuzzy diagnostic model. *Expert Systems with Applications* 32 , 955–975.
- [11]Chenn-Jung Huang, Ming-Chou Liu, San-Shine Chu, Chih-Lun Cheng.(2007). An intelligent learning diagnosis system for Web-based thematic learning platform. *Computers & Education* 48 , 658–679
- [12] Huapaya C.R.,Lizarralde F.A.Arona G.M. (2012) Modelo basado en lógica difusa para el diagnóstico cognitivo del estudiante. *Formación universitaria*. 5(1), 13-20.
- [13] Huapaya C.R. (2012) Proposal of Fuzzy Logic-based Students' learning assessment model. *CACIC 2012 WTIAE*. Bahía Blanca.
- [14] Zadeh L.A (1965). Fuzzy sets. *Information and control* 8. p 338-353
- [15] Zadeh L.A.. The concept of a linguistic variable and its application to approximate reasoning . I-III *Information Sci* 8, , 199-250 y 301-357. 1975
- [16] Dubois D. y Prade H. What are fuzzy rules and how to use them. *Fuzzy sets and systems*, 84:169-185. 1996.
- [17] Pohl, M. (2000). Bloom's (1956) Revised Taxonomy. Retrieved April 14, 2008, from the QSITE Higher Order Thinking Skills Online Course Web site: [http://www.odu.edu/educ/roverbau/bloom/blooms\\_taxonomy.htm](http://www.odu.edu/educ/roverbau/bloom/blooms_taxonomy.htm). Consulta: 8/3/2013.
- [18 ] Anderson, Et Al. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives (Complete ed.)*. New York: Longman.
- [19] Hammouda, K., Kamel, M. (2006). *Data Mining in e-learning. E-Learning Networked Environments and Architectures: A Knowledge Processing Perspective*, Samuel Pierre (Ed.), Springer Book Series: Advanced Information and Knowledge Processing (pp. 1-28).
- [20] Romero C, Ventura S. y García E. (2008) *Data Mining in course management systems: Moodle case study and tutorial*. *Computers & Education*. Volume: 51, (1), 368-384
- [21] Romero, C., Ventura, S., & Bra, P. D. (2004). Knowledge discovery with genetic programming for providing feedback to courseware author. *User Modeling and User-Adapted Interaction: The Journal of Personalization Research*, 14(5), 425–464.
- [22 ]Romero, C., & Ventura, S. (2006). *Data mining in e-learning*, Southampton, UK: Wit Press.

## Integrando Redes Sociales y Agentes de Software en Entornos Educativos

Antonieta Kuz<sup>(2,3)</sup>, Roxana Giandini<sup>(1,2)</sup>, Leopoldo Nahuel<sup>(1,2)</sup>

<sup>(1)</sup> LIFIA - Facultad de Informática - UNLP  
{giandini, lnahuel}@lifia.info.unlp.edu.ar

<sup>(2)</sup> LINSI - Departamento de Sistemas de Información - FRLP - UTN

<sup>(3)</sup> Facultad de Ciencias Exactas - UNICEN  
akuz@linsi.edu.ar

### RESUMEN

El objetivo de esta propuesta es realizar actividades de investigación y desarrollo en temas relacionados con Análisis de Redes Sociales (ARS) en conjunción con Agentes de Software inteligentes, aplicados a Entornos Educativos virtuales.

Específicamente se pretende proveer una metodología con soporte tecnológico (basada en ARS y Agentes de Software) que ayude al profesor a evaluar cualitativamente los distintos lazos internos del alumnado y tomar decisiones pedagógicas más precisas, a través de un ambiente educativo virtual. El agente permitirá que el profesor obtenga en forma dinámica una descripción de la estructura interna de los grupos, su organización, cohesión y quiénes son los miembros más significativos dentro del aula.

**Palabras clave:** *Redes Sociales, Análisis de Redes sociales, Agentes de Software, Entornos Educativos Web, Software Educativo para Docencia.*

### CONTEXTO

Actualmente el uso de redes sociales como Facebook, Twitter, otras redes de profesionales como LinkedIn, así como foros, listas de correo electrónico y salas de chat, se ha convertido en un factor determinante para la comunicación y relación entre las personas [1]. La revolución en las Tecnologías de la Información y las Comunicaciones es un proceso que avanza y de innovación constante, que ha cambiado los escenarios en los cuales se mueven personas y organizaciones [2].

El interés en el campo de investigación de las redes sociales se fue incrementando y evolucionando a través de una variedad de investigaciones. El enfoque metodológico del estudio de interacciones humanas representadas

mediante las redes sociales se denomina Análisis de Redes Sociales (ARS) [3]. El ARS se enfoca en la idea de profundizar en el conocimiento que se puede obtener a partir de los fenómenos sociales. Intentando llevar estos conceptos y desarrollar modelos para aquellos procesos sociales susceptibles de tratamiento matemático, se introdujo el concepto de red social [4] [5].

Por otro lado, en el Área de Inteligencia Artificial (IA) [9] [10], los agentes han sido un tópico de investigación intensiva durante muchos años. Este tópico aborda a los agentes como entidades que puedan actuar de forma autónoma, razonada y del cual se desprenden varios enfoques de investigación [11].

El objetivo general de nuestra propuesta es proveer una metodología con soporte tecnológico (basada en ARS y Agentes de Software) que ayude al profesor en un ambiente educativo [18] [19]. El agente permitirá que el profesor obtenga en forma dinámica una descripción de la estructura interna de los grupos, su organización, cohesión y quiénes son los miembros más significativos dentro del aula. La Figura 1 muestra un modelo preliminar UML de dicho Agente.

Actualmente, esta propuesta se enmarca en un proyecto PID-UTN que será presentado para su evaluación en el corriente año.

### 1. INTRODUCCION

Una red social es un conjunto de personas (organizaciones u otra entidad social) conectadas por un conjunto de vínculos, como amistad, laboral o solo de intercambio de información, como se observa en la Figura 2 creada con el software NodeXL [6] [7]. Formalmente, una red se define como una serie de nodos o actores, vinculados por una serie de relaciones que cumplen determinadas propiedades como se observa en la Figura 3. Un nodo en una red está vinculado con otro



mediante un vértice que presenta la dirección y el sentido del vínculo, estos vértices que conectan los diferentes nodos tienen un valor que puede ser o no numérico. Los actores pueden ser personas, organizaciones, entre otros y los vínculos entre los actores cumplen con una serie de propiedades, que pueden ser la intensidad de la relación, la posición del actor, la accesibilidad de un actor respecto a los demás, entre otros [8]. Estas propiedades definen la función o funciones de la red social y repercuten sobre los diferentes aspectos de las relaciones sociales. Con lo cual, una red social puede reflejar una variedad de relaciones sociales en las que un actor podría estar involucrado [4].

Por otro lado, como mencionamos en la sección anterior, en el Área de Inteligencia Artificial, los agentes de software tomaron relevancia en los últimos años. Existen diversas definiciones de agentes; una definición generalmente aceptada, considera a un agente como un sistema de computación que se encuentra situado en un ambiente y es capaz tener un comportamiento independiente con la misión de cumplir sus objetivos de diseño [12]. Un agente cumple con algunas características que mencionaremos a continuación y que lo distinguen de un simple programa [13] [14]:

- Autonomía: percepción del entorno.
- Sociabilidad: este atributo permite a un agente comunicar con otros agentes incluso con otras entidades.
- Racionalidad: el agente siempre realiza lo correcto a partir de los datos que percibe del entorno.
- Reactividad: un agente actúa como resultado de cambios en su entorno. En este caso, un agente percibe el entorno y esos cambios dirigen el comportamiento del agente.
- Proactividad: un agente es pro-activo cuando es capaz de controlar sus propios objetivos a pesar de cambios en el entorno.
- Adaptabilidad: está relacionado con el aprendizaje que un agente es capaz de realizar y si puede cambiar su comportamiento basándose en ese aprendizaje.

La estructura de un agente queda definida por:

- El Programa: es la función que implementaría la transformación

(mapping) de secuencias de percepciones en acciones

- La Arquitectura: se descompone en un conjunto de módulos que interactúan entre sí para lograr una funcionalidad requerida.

Como puede observarse dadas sus características un agente puede ser empleado en varias áreas, como aplicaciones industriales, comerciales, control de procesos, mediador, comercio electrónico, entre otras. Con lo cual pueden clasificarse según la función que realicen [15] [16].

Las más comunes son:

- Asistentes: Trabajan realizando tareas típicas como el manejo de la agenda. Estos agentes ayudan al usuario a planificar las reuniones. Sus acciones incluyen negociar, aceptar o rechazar reuniones.
- Filtros: Su tarea principal es la de analizar información según un conjunto de reglas dadas por el usuario. La aplicación típica es el filtrado de mensajes de correo electrónico.
- Guías: Asisten a los usuarios en el uso de una aplicación. Estos agentes monitorean las acciones de los usuarios e intentan sugerir qué pasos realizar para alcanzar el objetivo.

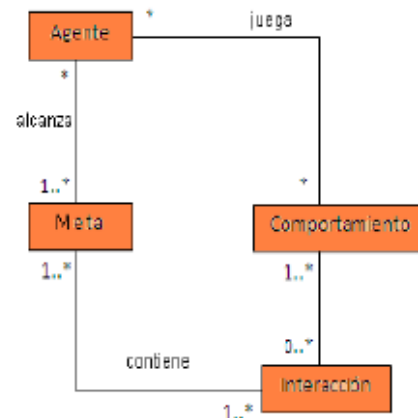


Figura 1: Modelo UML de Agente

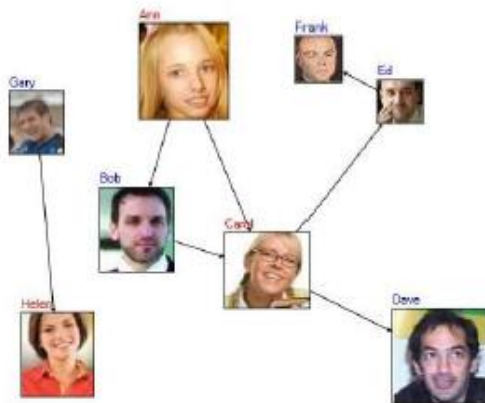


Figura 2: Ejemplo de un modelo de Red Social, modelada usando Node XL

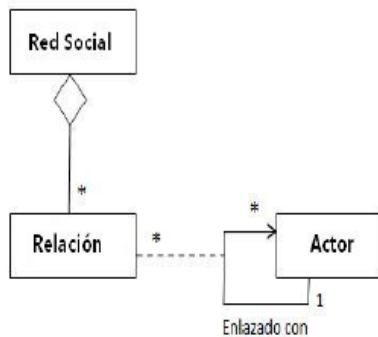


Figura 3: Modelo de Red Social en UML

## 2. LINEAS DE INVESTIGACION y DESARROLLO

Será necesario para el desarrollo de la propuesta, profundizar en las siguientes líneas de trabajo:

- Relevamiento bibliográfico pertinente de estudio.
- Análisis comparativo de técnicas de ARS (Análisis de Redes Sociales)
- Estudio y aplicación de Métricas en ARS
- Diseño del componente de aprendizaje de los agentes basado en las etapas de KDD con diferentes técnicas
- Integración de las técnicas desarrolladas para agentes inteligentes en diferentes dominios
- Experimentación y análisis de resultados
- Investigación de herramientas existentes para ARS, como NodeXL

## 3. RESULTADOS Y OBJETIVOS

El objetivo general de nuestra propuesta es proveer una metodología con soporte tecnológico (basada en ARS y Agentes de Software) que ayude al profesor en un ambiente educativo. El agente permitirá que el profesor obtenga en forma dinámica una descripción de la estructura interna de los grupos, su organización, cohesión y quiénes son los miembros más significativos dentro del aula. Para alcanzar específicamente este objetivo, se tendrán en cuenta tanto las interacciones entre los estudiantes en el aula como la interacción de los estudiantes con el profesor. A partir de dicha interacción se obtendrá un modelo de red social o sociograma. En este caso, el sociograma obtenido permitirá determinar la naturaleza e intensidad de las relaciones interactivas que se dan dentro de un grupo; el grado de cohesión y la estructura de un grupo, así como la posición que ocupa cada miembro dentro de él. La Figura 3 muestra un modelo preliminar de la red en UML.

Estas técnicas de ARS [22], sumadas al agente definido, serán las herramientas para que profesores de cursos on-line e investigadores evalúen la participación e interacción entre los alumnos. Esto también conducirá y permitirá generar a los profesores nuevas estrategias pedagógicas. El esquema general de nuestro Modelo de Solución puede verse en la Figura 4. Este modelo de solución, tendrá la versatilidad de permitirle al profesor además, intervenir intencionalmente y oportunamente para prevenir y/o modificar situaciones conflictivas que se puedan presentar en el aula.

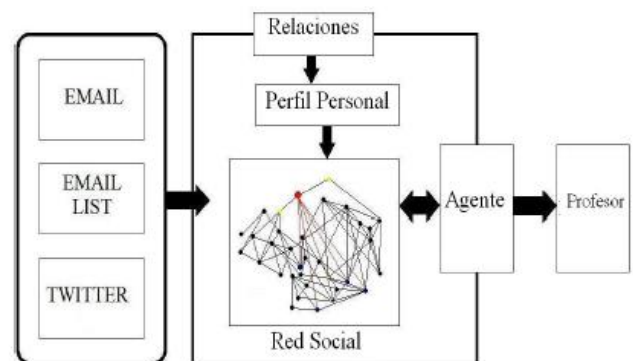


Figura 4: Modelo de Solución

Se espera por otro lado, la publicación de artículos en ámbitos científicos, de los resultados obtenidos.

#### 4. FORMACION DE RECURSOS HUMANOS

El Proyecto I/D en el que se enmarcarán estas líneas de trabajo, es parte de un proceso de incentivación para el desarrollo de actividades I/D, que se originan en el Laboratorio LINSI del Departamento de Sistemas de Información de la FRLP-UTN. Consecuentemente, se está poniendo acento y esfuerzo en las siguientes actividades:

- Desarrollo de seminarios abiertos de formación general en relación a temas, técnicas y tecnologías incluidos en esta línea de investigación, para alumnos avanzados en la carrera de Ingeniería en Sistemas y para becarios de este equipo de trabajo, realizados en UTN-Facultad Regional La Plata y UNLP-Facultad de Informática.

- A cargo de integrantes de este equipo de trabajo, se encuentran en curso 3 Tesis de Postgrado (Magíster en Ingeniería de Software y Magister en Tecnología aplicada a la Educación, UNLP y de Doctorado en UNICEN) relacionadas con el campo de investigación de este proyecto.

- Finalmente, los docentes integrantes de esta línea de investigación participan en el dictado de asignaturas con contenidos relacionados a este campo I/D:

En grado: UTN, Facultad Regional La Plata, Ingeniería en Sistemas de Información-Facultad de Ingeniería, UCASAL

En postgrado: UNLP, Maestría en Tecnología Informática aplicada a Educación – UNICEN, Facultad de Ciencias Exactas.

#### 5. REFERENCIAS

[1] Bernardo A. Huberman, Daniel M. Romero, and Fang Wu. Social networks that matter: Twitter under the microscope. CoRR, abs/0812.1045, 2008.

[2] Viktor Lee and Viktor Lee. Discussion. In How Firms Can Strategically Influence Open Source Communities. Gabler Verlag, 2012.

[3] J. Scott. Social Network Analysis: A Handbook. Sage Publications, 2000.

[4] Robert A. Hanneman and Mark Riddle. Introduction to social network methods. University of California, Riverside, Riverside, CA, 2005.

[5] A. Abbasi and J. Altmann. On the Correlation between Research Performance and Social Network Analysis Measures Applied to Research Collaboration Networks. In System Sciences (HICSS), 2011 44th Hawaii International Conference on, pages 1\_10, jan. 2011.

[6] Sorin Matei. Analyzing Social Media Networks with NodeXL: Insights from a Connected World by Derek Hansen, Ben Shneiderman, and Marc A. Smith. International Journal of Human-Computer Interaction, 27(4):405\_408, 2011.

[7] Marc Smith, Ben Shneiderman, Natasa Milic-Frayling, Eduarda Mendes

Rodrigues, Vladimir Barash, Cody Dunne, Tony Capone, Adam Perer, and

Eric Gleave. Analyzing (social media) networks with NodeXL. In C&T '09: Proc. fourth international conference on Communities and Technologies, pages 255\_264. ACM, 2009.

[8] Reinhard Diestel. Graph Theory (Graduate Texts in Mathematics). Springer, August 2005.

[9] Bull;j. The Caring Personal Agent. International Journal of Artificial Intelligence in Education, 13:14, 2003.10

[10] Mccalla. The Caring Personal Agent. International Journal of Artificial Intelligence in Education, 13:21\_34, 2003.

[11] Yoav Shoham. Agent oriented programming: An overview of the framework and summary of recent research. pages 123\_129. 1994.

[12] S.S. Priya, R. Subhashini, and J. Akilandeswari. Learning agent based knowledge management in Intelligent Tutoring System. In Computer Communication and Informatics (ICCCI), 2012 International Conference on, pages 1\_5, jan. 2012.

[13] Nicholas R. Jennings. An agent-based approach for building complex software systems. Commun. ACM, 44(4):35\_41, apr 2001.

[14] Andrew B. Williams and Zijian Ren. Agents teaching agents to share meaning. In Proceedings of the \_fth international conference on Autonomous

agents, AGENTS '01, pages 465\_472, New York, NY, USA, 2001. ACM.

[15] Silvia Schia\_no and Analia Amandi. Building an expert travel agent as a

software agent. *Expert Syst. Appl.*, 36(2):1291\_1299, mar 2009.

[16] Randy W. Kamphaus. *Clinical Assessment of Child and Adolescent Intelligence* (2nd Edition). Pearson Education, 2 edition, nov 2000.

[17] Xuedong Sun and Lianhao Zhang. Basic Factors Design of Pedagogical Agent System in an Ecology View. In *Hybrid Intelligent Systems*, 2009. HIS '09. Ninth International Conference on, volume 3, pages 183 \_186, aug. 2009.

[18] Ye Xiu-min and Fan Zhu-Qing. The Research of Link Relationships Based on SNA. In *Service Systems and Service Management*, 2007 International Conference on, pages 1 \_6, june 2007.

[19] Ajith Abraham, Aboul-Ella Hassanien, and Vaclav Snasel. *Computational Social Network Analysis Trends, Tools and Research Advances* Aboul-Ella Hassanien. 2010.

[20] L. Tang and H. Liu. GRAPH MINING APPLICATIONS TO SOCIAL NETWORK ANALYSIS.

[21] Duncan Watts. *Six Degrees: The Science of a Connected Age*. W. W.Norton & Company, 2003.

[22] N P Hummon and P Doreian. *Computational Social Network Analysis*. *Social Networks*, 12(4):2\_25, 2010.



# Tópicos Avanzados en la Programación de Computadoras

Lidia López, Silvia Amaro

{lidia.lopez, silvia.amaro}@fai.uncoma.edu.ar

Ingrid Godoy, Ana C.Alonso de Armiño, Andrea P.Alonso de Armiño

{ingridgodoy, anacarolinaalonsode, andrea.paola.alonso}@gmail.com

Marcela Leiva - marce\_leiva76@yahoo.com.ar

Departamento de Programación - Facultad de Informática

Universidad Nacional del Comahue

Buenos Aires 1400, Q8300IBX Neuquén, Argentina

## 1. Resumen

Existe la preocupación en la comunidad académica por los cursos de enseñanza de la programación de computadoras debido a la alta tasa de desgranamiento y al bajo nivel de motivación de los estudiantes. Se han propuesto herramientas y enfoques pero es difícil proponer una respuesta integral a la problemática. Son muchos los aspectos que influyen en el proceso de enseñanza-aprendizaje de la programación, algunos son intrínsecos, pero hay muchos otros que son ajenos al proceso de programación y que influyen fuertemente, que están ligados al cambio tecnológico, a los recursos didácticos utilizados, e incluso, a la situación social de los estudiantes.

Este proyecto pretende brindar aportes novedosos sobre la enseñanza de la programación así como el desarrollo de aplicaciones que mejoren la transferencia de conocimientos en programación de computadoras.

*Palabras clave:* programación de computadoras, evolución de la programación, enseñanza de la programación.

## 2. Contexto

A primera vista, la enseñanza de la programación de computadoras parece un problema sencillo; sin embargo, no lo es. Muchos enfo-

ques y herramientas han sido propuestos en los últimos diez años, no obstante, no parece existir una solución completamente satisfactoria.

En los últimos años se vienen desarrollando, en la Facultad de Informática, acciones tendientes a mejorar las condiciones de permanencia de los estudiantes, No obstante, se mantiene un alto grado de deserción en los cursos de enseñanza de programación de computadoras. Debido a este factor, sumado al bajo nivel de motivación de los estudiantes, es necesario analizar detenida y constantemente cómo poder delinear propuestas que integren los conocimientos del campo de la programación con los de la pedagogía y didáctica, acordes con las tecnologías y tendencias actuales.

La presente investigación corresponde al proyecto "Tópicos Avanzados en Programación de Computadoras" en evaluación por la Secretaría de Investigación de la Universidad Nacional del Comahue.

Se pretende hacer un análisis crítico acerca de las técnicas de enseñanza-aprendizaje de la disciplina programación de computadoras con vistas a proponer un enfoque metodológico para

facilitar la resolución de problemas, la construcción algorítmica y la programación de alto nivel que responda a las exigencias actuales y en constante evolución.

### 3. Introducción

La programación de computadoras ha evolucionado rápidamente: en menos de 60 años ha pasado de ser un arte, conocido por unos pocos expertos, a ser una disciplina de estudio incluida en muchas currículas de nivel de enseñanza secundaria y en planes de estudios universitarios. Y su evolución continua.

En la actualidad se cuenta con equipos mucho más poderosos, portables y económicos, provistos con múltiples entornos y herramientas de programación. Igualmente, se cuenta con abundante documentación sobre ingeniería de software, paradigmas, técnicas y herramientas.

No obstante, muchos cursos de programación impartidos en las universidades no responden a las exigencias que el momento histórico les impone, ya sea por el paradigma, el enfoque, las herramientas o la estrategia metodológica con que se desarrollan. Como consecuencia muchos estudiantes, después de aprobar las asignaturas respectivas, no se sienten preparados para programar[5].

En la actualidad las personas tienen nuevas maneras de expresarse, de comunicarse e informarse. Por esta razón, es necesario que se incorpore el uso de las TIC en el aula. Esto requiere que se busquen nuevas formas de generar el conocimiento, que se modifiquen ciertos hábitos y conductas, y se piense en una nueva forma de enseñar [1].

La Educación a Distancia [2] es una metodología de enseñanza que permite considerar las características y necesidades de

cada persona. El apoyo en las tecnologías para implementarla es una opción prometedora. La investigación debe centrarse en la evaluación de las metodologías de enseñanza haciendo uso de diferentes recursos y considerando los requisitos para una implementación exitosa. Este aprendizaje servirá también para mejorar la calidad de la enseñanza en la modalidad presencial.

En la Facultad de Informática hemos iniciado nuestra investigación en relación con las Plataformas de Educación a Distancia, tales como Moodle [3], hace ya varios años. En el año 2004 se ha implementado la plataforma PEDCO [4] y ha sido usada desde entonces como soporte de contenido y comunicación.

Algunas cuestiones a las que pretendemos encontrar respuestas son:

¿Cómo enseñar a programar? ¿Cuáles herramientas son las más indicadas? ¿Cómo usaremos cada una de estas herramientas?

Por otro lado para lograr la eficiencia en la resolución de problemas y la programación de las soluciones se requiere considerar aspectos avanzados de la programación.[13]

La evolución de las metodologías de programación apunta en general a optimizar los tiempos de desarrollo de software por lo tanto debemos estudiar y/o incorporar todas aquellas técnicas y herramientas que brinden la adquisición de las habilidades de las nuevas herramientas (o sus nuevas versiones) de manera rápida y eficaz.[7]

En este contexto identificamos como áreas de la programación a desarrollar a las siguientes:

- *Programación Multiparadigma y Programación Políglota:*

Se refiere a la integración de conceptos y construcciones de dos ó más paradigmas,

de forma que el programador pueda describir problemas y sus soluciones en un estilo mixto utilizando una vista apropiada en cada momento. Esto permite eficiencia en el modelado, programación y resolución de problemas. En este escenario los desarrolladores requieren mejorar su productividad, y para ello una opción interesante es un desarrollo basado en programación multiparadigma, o programación políglota, de múltiples lenguajes.[6]

- *Programación Extrema y Refactorización:*

Se basan en desarrollos dirigidos por testing y refactorización continua. Se trata de la evolución del código que está en funcionamiento de manera eficiente.[9]

- *Lenguajes Específicos de Dominio:*

Se refiere a un lenguaje de programación que simula los términos, modismos, y expresiones utilizadas entre expertos en un dominio optimizando el esfuerzo requerido para escribir o modificar código para las características de la aplicación. [8][10][11]

Pretendemos trabajar sobre alguna de estas líneas como aportes novedosos sobre las metodologías de programación así como el desarrollo de aplicaciones que mejoren la transferencia de conocimientos en programación de computadoras.

## 4. Líneas de Investigación y Desarrollo

Pretendemos trabajar sobre alguna de estas líneas como aportes novedosos sobre las metodologías de programación así como el desarrollo de aplicaciones que mejoren la transferencia de conocimientos en programación de computadoras.

Se establecen dos líneas de investigación conectadas transversalmente: los procesos de

enseñanza-aprendizaje de la programación se asisten de aplicaciones sobre dominios específicos, y los desarrollos de aplicaciones utilizan conceptos que se transmiten a través de metodologías de la enseñanza. Se identifican estas líneas de la siguiente manera:

*Enseñanza-aprendizaje de la Programación:* orientada al uso de TICs como técnicas no convencionales que permitan al alumno adquirir el conocimiento a partir de espacios virtuales de comunicación interactiva.

*Desarrollo de aplicaciones específicas:* análisis, diseño y construcción de algoritmos y herramientas que permitan desarrollar aplicaciones que asistan en la enseñanza de la programación. Para ello se considerarán enfoques avanzados de programación.

## 5. Resultados esperados

Este proyecto puede considerarse un desprendimiento del proyecto “Técnicas Avanzadas y Análisis para el desarrollo multiparadigma” finalizado. Se pretende continuar en esa dirección avanzando sobre el estudio de estas técnicas y la aplicación de las mismas en el proceso de enseñanza-aprendizaje de la programación en todos sus niveles.

Se define el objetivo general de la siguiente manera: Establecer modelos, procesos y técnicas de programación de computadoras cuyo alcance abarca la enseñanza de la programación a partir de las TICs, hasta aplicaciones específicas que complementen la enseñanza y la construcción de software.

## 6. Formación de Recursos Humanos

El mayor impacto del presente proyecto se centra en la formación de recursos humanos, consolidación de grupos de investigación e interacción entre grupos interdisciplinarios.

El equipo de investigación está conformado por 2 Profesoras, 3 Asistentes de Docencia y 1 Ayudante de Primera. Además participan 2 estudiantes.

Relacionados al proyecto se cuenta con 2 tesis de posgrado y 4 tesinas de grado en curso. Una de las tesinas de grado corresponde a una beca TICs de la convocatoria 2012 desde la Agencia Nacional de Promoción Científica y Tecnológica, a través del FONSOFT, en el marco del Programa para promover la innovación productiva a través del fortalecimiento y consolidación de Capital Humano aplicado a la industria de las Tecnologías de la Información y las Comunicaciones.

## Referencias

- [1] *Tutorías virtuales: acortando distancias a través de la comunicación electrónica*. J.Fernández Pinto. Quaderns Digitals N°29.
- [2] *Educación a distancia, educación presencial y usos de la tecnología: una tríada para el progreso educativo*. Albert Sangra Morer. Edutec. N° 15. Mayo 2002.
- [3] Moodle. <http://moodle.org.ar>
- [4] *PEDCO (Plataforma de Educación a Distancia Universidad Nacional del Comahue)*. C. Fracchia, A. Alonso de Armiño. Workshop de Tecnología Informática aplicada en Educación (WTIE). Congreso Argentino de Ciencias de la Computación. Universidad Nacional de La Matanza, San Justo, Buenos Aires, Octubre 2004.
- [5] *Reflections on the Teaching of Programming: Methods and Implementations*, Bennedsen, Jens. , 2008, Springer.
- [6] *Multiparadigm Constraint Programming Languages*, Hofstedt, Petra., 2011, Springer-Verlag.
- [7] *Guide to Teaching Computer Science: An Activity Based Approach*, Orit Hazzan, Tami Lapidat, Noa Ragonis. 2011, Springer. ISBN: 978-0-857294-42-5.
- [8] *DSLs In Action*, Ghost, Debasish. 2011, Manning Publications Co. ISBN: 978-1-935182-45-0
- [9] *Refactoring to Patterns*, Kerievsky, Joshua. 2011, Addison-Wesley. ISBN: 0-321-21335-1
- [10] *Metaprogramming Ruby*, Perrotta, Paolo. 2010, Jill Steimberg. ISBN: 978-1-934356-47-0
- [11] *Ruby On Rails 3 -Tutorial*, Hartl, Michael. 2011, Addison-Wesley.
- [12] *Proceedings of the Fourth Workshop on Refactoring tools, WRT 2011*, Dig, Danny and Batory, Don. 2011. ISBN: 978-1-4503-0579-2.
- [13] *Aprenda a Pensar como un programador*, Downey Allan, Elkner Jeffrey, Meyers Chris. 2002 . ISBN: 0-9716775-0-6



# Software educativo para la resolución numérica y gráfica de integrales y de ecuaciones diferenciales ordinarias

Ascheri M., Pizarro R., Astudillo G., García P., Culla E.

Facultad de Ciencias Exactas y Naturales. Universidad Nacional de la Pampa

Avda. Uruguay 151 - Santa Rosa - La Pampa

02954-425166

mavacheri@exactas.unlpam.edu.ar

## Resumen

En un Proyecto previo, desarrollamos un software educativo para la *resolución numérica de ecuaciones no lineales, interpolación y aproximación polinomial*, y *ajuste de curvas por mínimos cuadrados* utilizando herramientas gratuitas disponibles en la Web, tales como PHP, HTML, CSS, la librería JGraph y GIMP para el diseño y edición de imágenes. Usando las mismas herramientas, nos proponemos incorporar al software educativo métodos numéricos para la *resolución numérica y gráfica de integrales y de ecuaciones diferenciales ordinarias*.

Este recurso pedagógico se utilizará, básicamente, en el curso de Cálculo Numérico que se dicta para las carreras de Profesorado en Matemática, Licenciatura en Física e Ingeniería Civil.

Nos proponemos lograr una integración curricular de los distintos contenidos temáticos de Cálculo Numérico y poder alcanzar los siguientes objetivos:

- Generar el contexto educativo adecuado a los contenidos a desarrollar y a los objetivos propuestos en la asignatura.
- Proporcionar una herramienta tecnológica para afianzar los conceptos teóricos y la aplicación de los métodos numéricos a través de la visualización

gráfica del funcionamiento de los mismos.

- Poner a disposición de los estudiantes una herramienta de acceso libre y disponible en la Web.

**Palabras clave:** software educativo, cálculo numérico, herramientas gratuitas.

## Contexto

El presente Proyecto de Investigación, acreditado y financiado por la Facultad de Ciencias Exactas y Naturales de la Universidad Nacional de La Pampa, aborda líneas de investigación estrechamente vinculadas con varios de los Proyectos que se desarrollan en el Departamento de Matemática del cual formamos parte. Algunos de ellos se relacionan con matemática y su enseñanza y la inclusión de tecnologías en el proceso de enseñanza aprendizaje. El desarrollo de este Proyecto demanda una continua interacción con los investigadores de los demás Proyectos de la Institución, actuando en beneficio de sus avances.

## Introducción

La experiencia obtenida a partir de la estrategia metodológica empleada en el Proyecto de Investigación anterior, nos

induce a que continuemos con el desarrollo de software educativo que incluya otras temáticas de Cálculo Numérico, para lograr una integración curricular y complementar los resultados obtenidos hasta el momento. Esta herramienta tiene como funcionalidad pretendida, guiar el aprendizaje como apoyatura a la explicación del profesor. Se aplicará, fundamentalmente, para la enseñanza y el aprendizaje de los temas: *Resolución numérica y gráfica de integrales y de ecuaciones diferenciales ordinarias*, incluidos en el curso de Cálculo Numérico que se dicta para las carreras de Profesorado en Matemática, Licenciatura en Física e Ingeniería Civil.

En este Proyecto centramos nuestra atención en la influencia que tiene el uso de la tecnología computacional en procesos de enseñanza y de aprendizaje, es decir, en el valor de dicha tecnología desde lo pedagógico y lo didáctico.

Según Rivera Porto (1997), el diseño de material de este tipo es más que un “software” y debe presentar ventajas respecto a otros medios instruccionales, y esto debe quedar muy claro al autor y a los sujetos del curso que, de esta manera, estarán más motivados a estudiar bajo esta modalidad que por aquella que sólo utiliza los medios tradicionales. Por ello es que nos proponemos ampliar nuestro software educativo incorporándole nuevos contenidos temáticos de la asignatura “Cálculo Numérico” (Ascheri y Pizarro, 2007; Chapra y Canale, 2007; Edwards y Penney, 1994; Mathews y Fink, 2000; Nakamura, 1997), teniendo como base la experiencia recabada hasta el momento.

La variedad de carreras a la que está dirigida esta asignatura, requiere que se muestren aplicaciones específicas de los temas que se desarrollan. Cada estudiante tiene sus propias necesidades, motivaciones, deseos, aspiraciones, las cuales dependen de su estructura cognitiva y varían por medio del aprendizaje. Se logrará una mayor

receptividad por parte de los estudiantes al presentarles los temas a través de situaciones que les resulten cotidianas, con la incorporación de un software educativo como una herramienta que respalde el proceso de enseñanza y de aprendizaje. No sólo pretendemos que sea un apoyo didáctico para el desarrollo de los contenidos teóricos de la materia, de la guía de trabajos prácticos y de una serie de problemas y de actividades complementarias, sino que además pretendemos que en el futuro les brinde a los estudiantes herramientas que les permitan palear algunas de las dificultades que se enfrentarán en la actividad como profesionales.

La realización de ejercicios y prácticas es una de las modalidades más aplicadas en Matemática, debido a la naturaleza misma de la materia. Con el uso de este software el estudiante podrá desarrollar las actividades prácticas en una forma mucho más dinámica, interactuando fácilmente y teniendo una rápida respuesta a sus inquietudes. Para que esta modalidad realmente sea efectiva, previo al uso de un software de este tipo, el estudiante ha debido adquirir los conocimientos de conceptos y destrezas que va a practicar.

También el rol docente se verá afectado con su implementación. Con la inclusión de herramientas informáticas en nuestras clases, la actividad del docente cambiará del tradicional rol de informante a la del facilitador o guía (Cataldi, 2000).

Se modificará además, la planificación de las clases teóricas, prácticas y evaluaciones parciales, las que deberán incluir herramientas tecnológicas en las actividades a efectuar, tales como el uso del software a desarrollar.

De la búsqueda y análisis de elementos ya existentes (Dunn, 2000; Mora, 2010), de nuestra experiencia lograda a partir del desarrollo del Proyecto de Investigación anterior y de la priorización de temas de

más difícil comprensión, surgió que en esta segunda instancia decidiéramos incluir los temas: *Resolución numérica y gráfica de integrales y de ecuaciones diferenciales ordinarias*.

En la actualidad, en el curso de Cálculo Numérico se desarrollan una amplia variedad de problemas y de actividades afines a cada una de las carreras antes citadas, con el complemento de la informática.

Entre las actividades planificadas en esta asignatura, se encuentran aquellas en donde los estudiantes deben elaborar los programas correspondientes a los diferentes métodos numéricos desarrollados en las clases teóricas, complementando, de esta forma, las actividades de cálculo manual por medio de las cuales se obtienen los resultados una vez que se aplican dichos métodos. Los programas que diseñan e implementan los estudiantes en Octave, les permiten, en cada ejecución, ingresar los datos y obtener la solución de la situación problemática planteada.

Consideramos que la elaboración de estos programas es muy útil para los estudiantes, ya que les brinda la posibilidad de interpretar el método numérico y, eventualmente, adaptarlo a ejemplos que se puedan encontrar en su futuro profesional, debido a que han afianzando suficientemente sus conocimientos de programación. Pero, a menudo, pierden de vista cuándo y cómo "funciona" el método aplicado, no realizan ningún otro tipo de análisis a posteriori y sólo toman en consideración el resultado que arroja el programa. Además, hacer que realicen una prueba lo más completa posible sobre si el programa por ellos desarrollados "trabaja bien" independientemente de la elección de las diferentes funciones y de los valores iniciales, que analicen las condiciones de convergencia de algunos métodos, que corrijan los errores que pueden cometer al cargar los datos en la ejecución del

programa, constituye una tarea realmente dificultosa.

La implementación del software que hemos desarrollado hasta la actualidad ha sido un aporte relevante para subsanar, gradualmente, estos inconvenientes. La aplicación se encuentra disponible en el sitio <http://online2.exactas.unlpam.edu.ar/numerico/> dependiente de la Facultad de Ciencias Exactas y Naturales de la UNLPam.

La cátedra podrá abreviar el tiempo de dictado al apoyarse en esta herramienta educativa, en donde estará concentrada toda la información temática requerida. Esta disminución de los tiempos de trabajos de estudiantes y docentes redundará en la posibilidad de profundizar y extender el tratamiento de ciertos temas de difícil comprensión y/o dedicar tiempo a tareas remediales.

En la actualidad, se puede disponer de software libre y que además, cumpla con las necesidades de la asignatura en lo que se refiere al proceso de enseñanza y aprendizaje de los contenidos temáticos involucrados en la misma. Tal es el caso del software Octave (Eaton, 2002) que se apega a la filosofía GNU. Octave es un lenguaje de alto nivel para realizar cálculos numéricos en la computadora, y también es un programa de características similares a MATLAB. Los estudiantes, elaborarán sus propios programas en Octave.

En lo que se refiere al software educativo, trabajaremos con el lenguaje PHP, un preprocesador del hipertexto (Achour et al, 2003-2004).

PHP es un lenguaje de "código abierto" interpretado, de alto nivel, contenido en las páginas HTML y ejecutado en el servidor. PHP es un lenguaje de programación de estilo clásico, es decir, con variables, sentencias condicionales, bucles, funciones. No es un lenguaje de marcas como podría ser HTML, XML o WML. Está más cercano a JavaScript o a Java. Pero a diferencia de estos que se ejecutan en el navegador, PHP

se ejecuta en el servidor, por eso nos permite acceder a los recursos que tenga el servidor como, por ejemplo, podría ser una base de datos. El lenguaje PHP, es ejecutado en el servidor y el resultado enviado al navegador.

En síntesis, consideramos que la combinación de estas dos herramientas tecnológicas nos ha permitido desarrollar un software educativo que se encuentra disponible para todos los usuarios de la Web. Actualmente, nos planteamos ampliarlo teniendo en cuenta los objetivos propuestos para la enseñanza y el aprendizaje de temas de Cálculo Numérico.

### Líneas de investigación y desarrollo

Las líneas de investigación y desarrollo que seguimos son:

- Búsqueda y análisis de herramientas libres que permitan desarrollar software educativo con acceso Web.
- Desarrollo de software educativo para la enseñanza-aprendizaje de algunos de los métodos de cálculo numérico.
- Elaboración de una planificación adecuada que permita la eficiente inserción del software en el desarrollo de las clases, optimizando de esta forma los resultados esperados.

### Resultados y Objetivos

Hasta el momento, dado la etapa inicial del proyecto y considerando el proyecto anterior, se está recopilando bibliografía sobre los temas que nos ocupan y sobre los software existentes. Se han definido las herramientas a utilizar para el desarrollo del software.

Con este Proyecto esperamos brindar una metodología adecuada para la generación de herramientas educativas y proporcionar un nuevo software educativo para la cátedra de

Cálculo Numérico, para otras cátedras afines y para otros niveles educativos.

### Formación de Recursos Humanos

Se espera continuar con un grupo de trabajo considerando:

- La motivación de los integrantes a participar de jornadas y eventos científicos.
- La contribución al desarrollo de capacidades y habilidades intelectuales en jóvenes estudiantes que participen o se relacionen con el Proyecto.
- La posibilidad de interesar a otros docentes y estudiantes en participar en actividades de investigación.

A partir de la presentación de los resultados de proyectos relacionados con el actual se han adquirido compromisos tendientes a realizar intercambios de las producciones que surjan, logrando así ampliar la población a la que estará dirigida la implementación del software educativo y que no se restrinja sólo a la cátedra de Cálculo Numérico.

### Referencias

**Achour, M. - Betz, F. - Dovgal, A. - Lopes, N. - Olson, P. - Richter, G. - Seguy, D. - Vrana, J.,** *Manual de PHP*, Grupo de Documentación de PHP, 2003-2004. En <http://es.php.net/manual/es/>

**Ascheri, María E. - Pizarro, Rubén A.,** Libro de Texto para Estudiantes Universitarios: *CÁLCULO NUMÉRICO*. 1a ed.-Santa Rosa: Universidad Nacional de la Pampa, Fecha de catalogación 26/02/2008, Impreso en Argentina, EdUNLPam, ISBN N° 978-950-863-100-8, 500 páginas, 2007.

**Cataldi, Z.,** *Una Metodología para el Diseño, Desarrollo y Evaluación de Software Educativo*, Universidad Nacional de La Plata, Facultad de Informática, Tesis de Maestría, 2000.



- Chapra, S. - Canale, R.**, *Applied Numerical Methods with MATLAB*, 2ª Ed., Mc Graw - Hill / Interamericana de España, S. A. U., España, 2007. En línea: <http://www.geocities.com/eriverap/libros/Aprend-comp/apen1.html>
- Dunn, P. K.**, Autor de *Generalized Linear Models in MATLAB, glmLab version 2.5*, 2000. En línea: <http://www.sci.usq.edu.au/staff/dunn/glmlab/glmlab.html>
- Eaton, J. W.**, *GNU Octave Manual. A high-level Interactive Language for Numerical Computations*, Publisher: Network Theory Ltd., Free License: GNU General Public License, Edition 3 for Octave version 2.0.13. 2002.
- Edwards, C. H., JR - Penney, D. E.**, *Ecuaciones diferenciales elementales y problemas con condiciones en la frontera*. México: Prentice Hall Hispanoamericana, 1994.
- Marquès, P.**, *El software educativo*, Universidad Autónoma de Barcelona, 1996. En [http://www.lmi.ub.es/te/any96/marques\\_software](http://www.lmi.ub.es/te/any96/marques_software)
- Mathews, J. - Fink, K.**, *Métodos Numéricos con MATLAB*, Prentice - Hall, 2000. (Trad. de Numerical Methods using MATLAB, Prentice-Hall, 1999).
- Merril, P. et al**, *Computers in Education*, Allyn & Bacon, 1996.
- Mora, W.**, *Introducción a los métodos numéricos. Implementaciones Basic-Calc de LibreOffice y wxMaxima*, Revista digital Matemática Educación e Internet, Primera Edición, 2010. En [www.tec-digital.itcr.ac.cr/revistamatematica/](http://www.tec-digital.itcr.ac.cr/revistamatematica/)
- Nakamura, S.**, *Análisis Numérico y Visualización Gráfica con MATLAB*, Pearson Educación, 1997. (Trad. de Numerical Analysis and Graphic Visualization with MATLAB, 1a Ed., Prentice - Hall, 1996).
- Rivera Porto, E.**, *Aprendizaje Asistido por Computadora. Diseño y Realización*, 1997.

## GESTIÓN DE CONTENIDOS EN LA ERA DE LAS TIC.

Claudia Russo, Marina Calderone, Laura Duran  
 Pedro Ochipinti, Mariana Saenz, Mónica Sarobe, Mariano Briolotti, Cecilia De Vito,  
 Héctor Becerra, Rosana Piergalini, Bárbara Squillace, Germán Osella Masa

Instituto de Investigación y Desarrollo de Transferencia y Tecnología  
 Escuela de Tecnología  
 Universidad Nacional del Noroeste de la Provincia de Buenos Aires (UNNOBA)

crusso@unnoba.edu.ar, marina.calderone@nexo.unnoba.edu.ar, lauraduran@unnoba.edu.ar,  
 pocchipinti@unnoba.edu.ar, mariana.saenz@nexo.unnoba.edu.ar,  
 monicasarobe@unnoba.edu.ar, mbriolotti@hotmail.com, cecilia.devito@nexo.unnoba.edu.ar,  
 info@estudiocats.com.ar, mrosanapierg@yahoo.com.ar, barbarasuillace@hotmail.com,  
 german.osella@nexo.unnoba.edu.ar,

### Resumen

*La línea de investigación presentada propone desde una mirada interdisciplinaria determinar y sistematizar aspectos determinantes en el desarrollo de materiales didácticos digitales a fin de constituir un Manual de Buenas Prácticas para su producción, administración y almacenamiento.*

*Se consideran materiales digitales aquellos desarrollados para su utilización en los EVA (Entorno Virtual de Enseñanza y Aprendizaje), PLE (Entornos personales de aprendizaje), televisión digital, etc.*

*A la hora de desarrollar materiales digitales se cuenta con la posibilidad de utilizar diferentes medios y sistemas simbólicos para presentar, concretar y estructurar la información. Ello nos permitirá desde poder ofrecer una redundancia de la información más significativa, hasta concretarla o especificarla de manera más clara posible. Los medios con los que podemos contar, fundamentalmente, para la realización del contenido son la ilustración, el audio, el vídeo, la animación, la tecnología web, los hipertextos, entre otros.*

*Es importante, respecto de los diferentes medios, no perder de vista los comentarios que realiza Kilian (2001)[1], cuando nos llama la atención respecto al diseño de la información para la red”... exige una clase*

*de escritura muy diferente a las de otros medios, ni mejor ni peor, sólo diferente.” Es por ello que a la hora de producir materiales digitales debemos tener en cuenta aspectos que van desde el tipo de utilización educativa que realizaremos de la red, hasta la dimensión conceptual en la cual nos apoyemos para la realización de los materiales que transmitiremos a través de ella.*

*Desde este proyecto se plantea el desarrollo de un Manual de Buenas Prácticas para la producción, administración y almacenamiento de contenido digital. El cual contendrá entre otras cosas la sistematización de aspectos que facilitan la producción de materiales didácticos digitales, aspectos que favorezcan a la gestión y administración de dichos materiales.*

**Palabras clave:** *materiales didácticos, gestión de contenidos, entorno virtual de enseñanza – aprendizaje*

### Contexto

Esta línea de investigación fue aprobada por la Secretaría de Investigación, Desarrollo y Transferencia de la UNNOBA en el marco de la convocatoria a Subsidios de Investigación Bianuales (SIB2013), se desarrolla en el Instituto de Investigación en Tecnologías y Transferencia (IITT) dependiente de la mencionada Secretaría. El

IIT trabaja en conjunto con la Escuela de Tecnología. El equipo está constituido por docentes e investigadores pertenecientes a los departamentos de Informática y Tecnología, Humanidades y Afines y Complementarias y alumnos de las carreras Licenciatura en Sistemas, Ingeniería en Informática y Diseño Gráfico de la UNNOBA.

Asimismo se coordina con otros dos proyectos presentados por el Instituto de Investigación y Transferencia en Tecnología (ITT), uno relacionado con la Gestión de la Tecnología Aplicada en contextos Educativos y otro relacionado a los Sistemas Ubicuos: desarrollo y aplicaciones. Además se relaciona con el Proyecto de Certificación de Calidad y digitalización de Procesos en Organizaciones presentado por el Instituto de Política y Gobierno (IPG). Esto muestra una clara la interdisciplinariedad y transversalidad del proyecto presentado.

### Introducción

Los rápidos avances del conocimiento y de la tecnología han sido capaces de cambiar los perfiles de la actividad humana en muy pocos años. La sociedad del conocimiento se promueve desde las organizaciones internacionales como fórmula de respuesta a los nuevos retos planteados por la tercera revolución industrial, la globalización de la economía y la mundialización de la cultura.

“Cuando nos referimos a la economía y la técnica, nos encontramos ante procesos que reproducen sus mecanismos, de modo igual, en todos los rincones del planeta... Sin embargo, es difícil sustentar el mismo argumento respecto de los universos culturales. Por ese motivo, prefiero utilizar el término “globalización” al referirme a la economía y la tecnología; son dimensiones que nos reenvían a una cierta unicidad de la vida social. Y reservo entonces el término “mundialización” para el dominio específico de la cultura.” (Ortiz, R. 1996) [2]

Los ciudadanos de esta sociedad, aprendices permanentes a lo largo de su vida, deben contar con instrumentos cognitivos para adquirir nueva y cambiante

información, nuevos y diferentes roles profesionales, destrezas y habilidades diversas, más sutiles y tecnificadas, y en la esfera personal, actitudes y valores capaces de permitirles adaptarse a todos estos cambios.

La sociedad del conocimiento demanda un saber aprender que debe reactivarse de manera continua y que se presenta como la única herramienta eficaz para afrontar un entorno semejante.

Uno de los desafíos más difíciles será el de modificar nuestro pensamiento de manera que enfrente la complejidad creciente, la rapidez de los cambios y lo imprevisible que caracterizan nuestro mundo.

... (en) una sociedad y economía basadas en el conocimiento, las estrategias del aprendizaje durante toda la vida son necesarias para encarar los desafíos de la competitividad y el uso de nuevas tecnologías y para mejorar la cohesión social, la igualdad de oportunidades y la calidad de vida. (Declaración de Praga, 2001, en Macías Gómez, p. 80) [3]

En este contexto, la educación mediada por tecnologías esta llamada a desarrollar nuevas posibilidades hasta este momento inéditas o escasamente atendidas. Por su capacidad de fomentar el aprendizaje autónomo, convierte al alumno en el protagonista del proceso, permitiéndole establecer su propio ritmo, estilo y método de aprendizaje, organizar sus espacios y tiempos de estudio según sus necesidades, desarrollando su habilidad de aprender a aprender, requisito imprescindible en la sociedad del conocimiento. Y, particularmente en su modalidad a distancia o semipresencial, por superar barreras geográficas, temporales, sociales, económicas, étnicas y culturales, democratiza el acceso a la educación, camino hacia la reducción significativa de la actual y acumulada inequidad social y económica.

No es que esta modalidad vaya a sustituir - ni siquiera en el futuro- a los niveles educativos considerados formales, tampoco su principal función es competir con ellos, sino sumarse y complementarlos en

aquellos aspectos en que éstos no son suficientes o no son competitivos.

“La historia de la comunicación es acumulativa; cada nuevo medio, cada nuevo lenguaje, se suma a los ya existentes sin eclipsarlos”. Jean Cloutier (1972) [4]

Es precisamente esta capacidad acumulativa y de complementariedad lo que ha permitido a la educación mediada por tecnologías en todas sus formas el crecimiento exponencial que experimenta.

Por otro lado es propicio destacar las experiencias previas llevadas a cabo en el marco del proyecto “UNNOBA VIRTUAL. Una plataforma para la integración de sistemas, metodologías y herramientas de enseñanza y aprendizaje.”, acreditado en el marco de la convocatoria a Subsidios de Investigación BIANUALES (SIB) 2010, llevado a cabo en el Instituto de Investigación y Transferencia en Tecnología (ITT). En este proyecto se propuso la definición de un modelo que permitiera gestionar las actividades formación con el uso de las TIC, planteando una serie de áreas que abordaban diferentes aspectos considerados necesarios y/o deseables para brindar una experiencia de educación a distancia con requerimientos de calidad. Dichas áreas llevarían a cabo distintas funciones, estarían constituidas por recursos humanos especializados y cooperarían entre sí. Uno de los componentes del modelo propuesto hace referencia a la generación de materiales digitales para los procesos de Enseñanza y Aprendizaje en el entorno virtual es por ello que la implementación del presente Proyecto tendrá un alto impacto en el modelo de enseñanza y aprendizaje planteado para la Universidad y en los Organismos públicos y privados.

### **Líneas de investigación y Desarrollo**

Se plantean dos etapas para el desarrollo de la presente investigación. Una primera etapa que abordaría la instancia de validación conceptual, empírica y operativa a desarrollarse durante el primer año,

donde se focalizará en el conocimiento de situaciones, costumbres y actitudes predominantes a través de la descripción exacta de las actividades, objetos, procesos y personas; y una segunda etapa de validación expositiva, dedicada a desarrollar un Manual de Buenas Prácticas en la producción y gestión de material didáctico digital.

### **Objetivos Generales**

- Analizar las variables que generan el cambio de paradigma en las condiciones de acceso al conocimiento.
- Realizar un estudio de las diferentes modalidades que la aparición de las nuevas tecnologías introducen en el ámbito de la educación formal.
- Sistematizar un corpus de buenas prácticas en la producción, gestión y almacenamiento de materiales didácticos digitales.
- Realizar un manual de buenas prácticas que aporte a la construcción y gestión de materiales didácticos digitales para la coordinación del equipo interdisciplinario encargado de su implementación y establezca una sistematización para su administración, gestión y almacenamiento.

### **Objetivos Específicos**

- Indagar en las diferentes herramientas de la actualidad que suponen un cambio sustantivo en la relación sujeto-objeto del conocimiento
- Analizar las conductas que se generan tanto en el espacio aúlico como en el extráulico a partir del uso de las nuevas tecnologías
- Relevar y reconocer constantes y variables

### **Resultados e impacto esperados**

El manual de buenas prácticas constituirá información básica para los actores intervinientes en el proceso de producción y gestión de los materiales digitales permitiendo una sistematización que facilite la producción, reutilización y administración de los mismos.

Se espera que los aspectos investigados contribuyan al fortalecimiento de las



experiencias de enseñanza y aprendizaje llevadas a cabo por nuestra universidad.

Por otra parte, es objetivo del presente proyecto promover en el marco de desarrollo del trabajo la publicación de artículos en revistas especializadas, en congresos nacionales e internacionales para difusión y transferencia, así como también la transferencia hacia la educación formal del ámbito educativo a través de aportes al diseño curricular vinculados a las carreras afines. También se promueve el desarrollo de seminarios y conferencias en el marco local y regional, y la participación en encuentros, jornadas y congresos a través de ponencias y comunicaciones, etc.

### Formación de Recursos Humanos

Se pretende propiciar un marco formal para el establecimiento de líneas de investigación en estas temáticas, atendiendo a la formación de recursos humanos, en su rol de investigadores o partícipes activos en equipos de investigación; fomentando la culminación de sus estudios superiores, promoviendo la redacción, exposición y defensa de Trabajos Finales de Grado y Postgrado.

En relación a este tema, para los próximos dos años se espera contribuir al inicio y concreción de dos Tesinas en el área de informática y diseño y dos Tesis de Magister.

### Referencias

- [1] **Kilian**, Crawford. 2001. Escribir para la Web. Ed. Deusto, Barcelona
- [2] **Ortiz**, Renato. 1996. Otro territorio. Ensayos sobre el mundo contemporáneo. UNQ. Bs As.
- [3] **Macías Gómez**, Escolástica. 2011. La práctica de la tutoría en la universidad, en Nuevos contextos de enseñanza aprendizaje en el espacio europeo de enseñanza superior. Capítulo 2. Colección Ideas en Debate. Serie Educación. Miño y Dávila Editores.
- [4] **Cloutier**, Jean (1972) La era de EMEREC. París

### Bibliografía

**Almenara, J.; Graván P.** E-actividades. Un referente básico para la formación en Internet. Primera edición. Editorial Eduforma. España, 2006.

**Alonso**, Rodrigo. 2009. Tu computadora es un campo de batalla. Tensiones tecnológico político-culturales en la era de las TIC, en ¿Desea guardar los cambios? Centro Cultural España-Córdoba. Argentina

**Casamayor, G.; Alós, M.; y otros.** La Formación On-line, Una mirada integral sobre el e-learning, b-learning. Primera edición. Editorial Graó. Barcelona, España. 2008

**Barberà**, E. La incógnita de la Educación a Distancia. Primera edición. Editorial Horsori. Barcelona, España. 2001.

**Levis, D.; Gutierrez Ferrer, M.** ¿Hacia la herramienta educativa universal?, Enseñar y aprender en tiempos de Internet. Editorial Ciccus. Buenos Aires, Argentina. 2000.

**Litwin**, E. Tecnologías educativas en tiempos de Internet. Primera edición. Editorial Amorrortu. Buenos Aires, Argentina. 2009.

**Litwin, E.; Maggio, M.; Lipsman, M.** Tecnologías en las aulas, Las nuevas tecnologías en las prácticas de la enseñanza. Casos para el análisis. Primera edición. Editorial Amorrortu. Buenos Aires, Argentina. 2005.

**Logan**, Robert K. 2008. ¿Que es un libro? ¡Pasado, Presente y Futuro! De la tabla de arcilla al smartbook, en Carlón / Scolari, El fin de los medios masivos. El comienzo de un debate. La Crujía. Buenos Aires.

**Macías Gómez**, Escolástica. 2011. La práctica de la tutoría en la universidad, en Nuevos contextos de enseñanza aprendizaje en el espacio europeo de enseñanza superior. Capítulo 2. Colección Ideas en Debate. Serie Educación. Miño y Dávila Editores.

**Mitchel, William J.** E-topía. Vida Urbana, Jim; pero no la que nosotros conocemos. GG.

**Ortiz**, Renato. 1996. Otro territorio. Ensayos sobre el mundo contemporáneo. UNQ. Bs As.

## Algoritmos. Un Análisis comparado de su enseñanza teórica y práctica en universidades del NOA y extranjeras.

Carlos H. SAVIO<sup>1</sup>, Hugo DIP<sup>1</sup>, María V. POLICHE<sup>1</sup>, Juan C. GHIRALDI<sup>1</sup>, Walter BIVANCO<sup>1</sup>

1) Departamento de Sistemas, Facultad de Tecnología y Ciencias Aplicadas, Universidad Nacional de Catamarca, Maximio Victoria 35, (4700), Catamarca, Argentina.

Tel: +54 (383) 4435112

e-mails: {csavio, hrdip, vpoliche, junca, wbivanco}@tecno.unca.edu.ar

### RESUMEN

Desde la implantación del CCA -Ciclo Común de Articulación- en las carreras de ingeniería de las cinco universidades nacionales del NOA, se han consensado los contenidos mínimos que las asignaturas correspondientes al primer año deben contener. Está incluida en este grupo la asignatura Fundamentos de Informática, cuyos principales contenidos giran alrededor de la enseñanza de los algoritmos como metodología de aplicación para la solución de problemas, la que se aplica para satisfacer los requerimientos competenciales aprobados desde el CONFEDI para las carreras de ingeniería. Comparar las diferentes metodologías de enseñanza de algoritmos utilizadas entre las universidades miembros del CCA y relevar cómo se enseña a diseñar algoritmos en universidades extranjeras elegidas, permitirá al equipo de cátedra de Fundamentos de Informática de la UNCa, contar con elementos suficientes para mejorar metodologías y prácticas aplicadas actualmente.

Simultáneamente se iniciará un proceso exploratorio en cuanto a la posibilidad de incorporar parte o todo el dictado de la asignatura en un método de enseñanza a distancia o semipresencial, relevando las necesidades y requerimientos físicos, tecnológicos y de capacitación de recursos humanos necesarios para una eventual implementación.

Se espera que la incorporación de esta información, estrategias y métodos, contribuyan a disminuir la deserción de los estudiantes de primer año y agregue valor a la aplicación de la algoritmia.

Palabras Claves: algoritmos, metodologías aplicadas, fundamentos de informática, CCA.

### CONTEXTO

Este proyecto de investigación fue aprobado y subvencionado por la Secretaria de Ciencia y Tecnología (SCyT) de la Universidad Nacional de Catamarca (UNCa), Argentina. Este proyecto se inició en Enero de 2012 y su fecha de finalización es en Diciembre de 2013. Se desarrolla en el Departamento de Formación Básica y el Departamento de Informática de la Facultad de Tecnología y Ciencias Aplicadas de la Universidad Nacional de Catamarca.

Este proyecto se encuadra en el área de la Ingeniería del Software

### INTRODUCCIÓN

La implantación del CCA -Ciclo Común de Articulación- es el resultado de un proyecto desarrollado entre SEP02 y MAR03, que involucró a una red de cinco universidades nacionales del noroeste argentino: Tucumán, Salta, Jujuy, Santiago del Estero y Catamarca con el objetivo de diseñar un Ciclo General Básico para la Familia de Carreras de

Ingeniería a fin de establecer mecanismos de articulación y acreditación simultánea. Para ello y producto de algunas reuniones de trabajo, se consensuaron contenidos mínimos y comunes de cada una de las asignaturas desarrolladas en el 1er año de las Ingenierías. Entre ellas, la asignatura Fundamentos de Informática (en adelante FDI), tiene establecido como uno de sus contenidos mínimos, la enseñanza y uso de algoritmos para la resolución de los problemas. Sin embargo, y no obstante ser la algoritmia la principal metodología al momento de identificar, formular y resolver problemas de ingeniería, los estudiantes no logran en un cuatrimestre internalizar su mecanismo de aplicación y mucho menos desarrollar habilidades para la resolución de los problemas pasibles de ser modelizados. Este problema se constituye en una fuerte restricción para las asignaturas siguientes en el plan de estudios, que basan su desarrollo en los algoritmos, utilizando los conocimientos que supuestamente fueron aprehendidos por los estudiantes en la asignatura FDI.

El CONFEDI – Consejo Federal de Decanos de Ingeniería- en el XL Plenario de OCT06, aprobó un documento mediante el cual se definen las competencias genéricas de las carreras de Ingeniería involucradas que prevén procesos de planificación, formulación, gestión, desarrollo y ejecución de proyectos de ingeniería. Éstos conllevan implícita y explícitamente la aplicación de los algoritmos: por un lado como metodología para el análisis y resolución de problemas y por otro lado, como herramienta de diseño de programas computacionales, producto de la modelización de aspectos de la realidad.

Por otro lado la posibilidad de la incorporación de métodos de enseñanza semi presenciales, atento a la naturaleza de la asignatura, será motivo de análisis a fin de de estimular a los estudiantes y lograr un permanente contacto de la cátedra con ellos.

### **LINEAS DE INVESTIGACION Y DESARROLLO**

Siendo la asignatura FDI de la Facultad de

Tecnología y Cs. Aplicadas –FTyCA- de la UNCa, la responsable de introducir y capacitar a los estudiantes de las Ingenierías en el uso de los algoritmos, se considera necesario realizar investigaciones en dos ámbitos diferentes, uno subregional y otro internacional:

**a)** En las asignaturas FDI pertenecientes a las 7 facultades de ingeniería signatarias del CCA: realizar un análisis comparado sobre ingreso / regularización / aprobación de los estudiantes; formas de evaluación; metodologías de enseñanza y aplicación de algoritmos; bibliografía utilizada; estructuras de las cátedras; opinión de docentes de FDI acerca de la asignatura en general y de la enseñanza de algoritmos en particular; recabar información de docentes a cargo de cátedras subsiguientes en el diseño curricular que utilicen algoritmos, acerca del desempeño de sus estudiantes en materia de comprensión y resolución de problemas.

**b)** En dos universidades extranjeras: Investigar en 2 universidades extranjeras de reconocida trayectoria en su desarrollo tecnológico, qué se enseña cuando se enseñan algoritmos, y qué metodologías se aplican.

### **OBJETIVOS**

A los efectos de dar cumplimiento con las dos líneas de investigación planteadas con anterioridad se fijaron los siguientes objetivos:

- Contar con información actualizada sobre datos estadísticos de estudiantes en las asignaturas FDI pertenecientes al CCA de las UUNN del NOA
- Evaluar la percepción que tienen los docentes acerca de la problemática del aprendizaje de los algoritmos
- Conectar a todas las cátedras de FDI del CCA estimulando un permanente intercambio de información
- Lograr una articulación horizontal entre asignaturas FDI del CCA y una articulación vertical con las asignaturas subsiguientes en los respectivos diseños curriculares que necesiten de los conocimientos de algoritmos

- Comparar metodologías de enseñanza de algoritmos entre las universidades relevadas (argentinas y extranjeras).
- Establecer vínculos con universidades extranjeras con vistas a trabajos conjuntos de investigación y/o desarrollo.
- Estudiar y analizar las normas estandarizadas para el diseño de algoritmos.
- Proponer metodologías y métodos superadores de enseñanza para la asignatura FDI de la UNCa, explorando especialmente la potencialidad de los sistemas semi presenciales.

### RESULTADOS OBTENIDOS

- Se elaboró un cuadro con datos referidos a composición e integración de las cátedras, antecedentes de trabajos de investigación similares, métodos de enseñanza de los algoritmos, bibliografía utilizada, contactos establecidos con otras instituciones de ES para los fines del presente proyecto, etc
- Se implementó una base de datos responsables informantes de las asignaturas FDI de las 5 universidades miembros del CCA
- Se realizó una encuesta a los docentes de las asignaturas usuarias de la cátedra FDI de las 4 carreras de ingeniería de la Facultad de Tecnología y Ciencias Aplicadas, a fin de conocer en qué condiciones encuentran estos docentes a sus alumnos en materia de, entre otras cosas, análisis, entendimiento y resolución de problemas, manejo de algoritmos, lógica algorítmica, etc.
- Se realizó un análisis comparado del rendimiento de los estudiantes en esta asignatura y las restantes que integran el primer año del CCA durante el período 2008-2011.
- Se diseñó una encuesta los miembros de los equipos de cátedra de las 4 universidades del NOA a los efectos de sobre que recoja datos sobre contenidos teóricos y prácticos de la asignatura FI y

también sus opiniones en cuanto a las calificaciones, actitudes y aptitudes de sus alumnos ante determinadas situaciones características de los procesos algorítmicos.

### TRABAJOS FUTUROS

- Se elegirán 2 universidades extranjeras de reconocida trayectoria. Una pertenecerá a un país europeo y la otra a un país latinoamericano. Se establecerán contactos electrónicos con los pares de aquéllas a fin de incorporarlos a la base de contactos.
- Se pondrá un especial énfasis en mantener contactos bilaterales con miras a la integración de investigaciones de carácter transnacional
- Se elaborará una encuesta destinada a una muestra de estudiantes de las asignaturas FDI de las 5 facultades miembros del CCA que recoja principalmente dificultades en el aprendizaje de algoritmos y donde puedan proponer mejoras.
- Se contrastarán resultados numéricos correspondientes a estudiantes de las 5 asignaturas FDI del CCA.
- Se cruzarán las variables recogidas utilizando el método multivariado
- Se propondrá la realización de pasantías de los auxiliares docentes de FDI de la UNCa en las otras universidades del NOA, aprovechando el convenio existente firmado por los Decanos, a los efectos del intercambio de experiencias.
- Se capacitará al equipo en manejo del SPSS y conceptos de educación a distancia y elaboración de materiales
- El equipo de trabajo mantendrá reuniones de trabajo periódica con fines de evaluación de material leído y relevado
- Se contrastarán los resultados obtenidos en encuestas y entrevistas acerca de aplicaciones metodológicas con las actividades que se realizan en FDI de la UNCa.



## FORMACION DE RECURSOS HUMANOS

El grupo de trabajo está conformado por: un director, un co-director y por el equipo de docentes y auxiliares de la cátedra Fundamentos de Informática correspondiente al 1er año del CCA de la FTyCA de la UNCa. También participan del equipo el profesor a cargo de la asignatura Programación II de la carrera Ing. en Informática, el Profesor de la asignatura Álgebra del CCA, quien viene participando de diversos análisis estadísticos evaluando el rendimiento de los estudiantes del 1er año de las carreras de ingeniería de la FTyCA.

En este proyecto también participan dos participan estudiantes de la Carrera Ingeniería en Informática de la Facultad de Tecnología y Ciencias Aplicadas.

Se pretende que durante el presente año dos integrantes del equipo realizarán pasantías en cátedras FDI de otras Universidades del NOA miembros del CCA.

5. Revista CONFEDI. Competencias Genéricas de Ingeniería, Argentina (2007).

## REFERENCIAS

1. Fundamentos de informática en el primer año de las ingenierías de la Universidad Nacional de Catamarca en el marco de un proyecto articulado”. IV Jornadas de Ciencia y Tecnología de Facultades de Ingeniería del NOA - Universidad Nacional de Santiago del Estero - 4 y 5 de Diciembre de 2008. ISBN N° 978-987-1341-38-2
2. Convenio de Articulación UU.NN- Ciclo Común de Articulación CCA-NOA. Universidad Norte Grande (2007).
3. Medina, Liliana: La articulación como estrategia de calidad en la Facultad de Tecnología y Ciencias. Aplicadas de la UNCa. Universidad Nacional de Catamarca, Argentina (2006).
4. Programas Analíticos de la asignatura Fundamentos de Informática de la Facultad de Tecnología y Ciencias. Aplicadas de la UNCa, Argentina (2008-2011).

# Televisión Digital como plataforma educativa

Maximiliano Abrutsky, Federico Bobbio, Ignacio Giagante, Ricardo Medel,  
Néstor Navarro, María Alejandra Odetti

Laboratorio de Investigación de Software  
Departamento de Ingeniería en Sistemas de Información  
Universidad Tecnológica Nacional – Facultad Regional Córdoba

Maestro M. Lopez esq. Cruz Roja Argentina  
Ciudad Universitaria – Córdoba  
{mxtrula,federicobobbio,igiagante,ricardo.h.medel,nestornav,maaodetti}@gmail.com

## Resumen

En esta línea de I+D investigamos las características, potencial, limitaciones y herramientas que condicionan o facilitan el desarrollo de contenidos educativos para la plataforma Televisión Digital Terrestre, la cual tiene un potencial de transformarse en una plataforma educativa de alcance masivo.

Nuestros objetivos formar recursos humanos con habilidades para imaginar, diseñar, desarrollar y desplegar aplicaciones educativas en TVD, para lo cual vamos a producir al menos dos aplicaciones educativas y generar un espacio de discusión sobre las mejores prácticas de Ingeniería de Software para el desarrollo de contenidos educativos en TVD.

**Palabras clave:** Televisión Digital Terrestre, Educación, Ginga, NCL, Lua.

## Contexto

La línea de I+D presentada aquí se desarrolla actualmente en el marco del proyecto de investigación “*Educación Multimedia utilizando la Televisión Digital como plataforma*”, dirigido por el Dr. Ricardo Medel y desarrollado en el Laboratorio de Investigación de Software

del Departamento de Ingeniería en Sistemas de Información de la Facultad Regional Córdoba de la Universidad Tecnológica Nacional. El proyecto ha sido aprobado para ser desarrollado entre los meses de Enero de 2011 y Diciembre de 2013, y es financiado por la Secretaría de Ciencia y Tecnología de la UTN bajo el código UTN1402.

## Introducción

El objetivo de este proyecto es establecer las condiciones sociales y tecnológicas que permitirían aprovechar las posibilidades que brinda la Televisión Digital (TVD) como plataforma de educación de alcance masivo.

La TVD es una moderna tecnología de transmisión digital de contenidos televisivos. A diferencia de lo que sucede en la televisión analógica tradicional, en la TVD la información se transmite codificada en forma binaria, permitiendo una óptima calidad de imagen y sonido, dado que distorsiones e interferencias pueden ser corregidos por software [1].

Además, la TVD no sólo permite transmitir video y sonido, sino que se pueden transmitir datos (binarios) al mismo tiempo, lo que permite transmitir y ejecutar software, abriendo infinitas posibilidades al

medio televisivo.

Por ejemplo, se pueden ofrecer servicios interactivos que permiten al usuario ser un eslabón activo de la cadena de transmisión de información, dejando su usual lugar de receptor pasivo. Esto permite la creación de nuevos negocios basados en la provisión no sólo de contenidos, sino de software para el desarrollo y provisión de material relacionado a diversas temáticas, tales como la educación, los derechos, la cultura, la religión, el entretenimiento, etc.

Técnicamente, la TVD puede dividirse en TVD por Satélite, TVD por Cable y TVD Terrestre. En nuestro trabajo nos enfocamos particularmente en la TVD Terrestre, dado su alcance masivo, potenciado por varias políticas de estado en la región, tales como la creación del Sistema Argentino de Televisión Digital Terrestre (SATVD-T) [2], el Sistema Brasileiro de TV Digital Terrestre [3] y el programa de TVD de CANTV (Venezuela) [4].

Debido a sus ventajas técnicas y operativas, nuestro país y otros 11 países han seleccionado para la TVD Terrestre la norma del Sistema Brasileño de Televisión Digital Terrestre (SBTVD-TB), basada a su vez en el sistema Japonés ISDB-T [5].

A fin de recibir la señal de TVD e interactuar con el usuario, el televisor debe conectarse a un equipo receptor de TVD, el cual puede estar integrado en el televisor o bien estar en un dispositivo aparte o *set-top box* (STB). Este equipo receptor es programable, lo que permite el desarrollo de software que explote las posibilidades brindadas por la gran cantidad de información posible de ser transmitida por el sistema de TVD.

La norma brasileña incluye el *middleware* Ginga [6], que habilita el desarrollo de software bajo dos paradigmas de programación diferentes: la programación

procedural con Ginga-J (basada en Java) y la programación declarativa con Ginga-NCL [7].

Aunque en el contexto actual la TVD en nuestra región tiene un fuerte apoyo político y gran potencial económico, enfrenta una falta casi completa de contenidos significativos y escaso desarrollo de las herramientas que permitan desarrollar dichos contenidos y aprovechar la plataforma en todo su potencial.

De particular importancia es la potencialidad de uso de la TVD Terrestre como herramienta educativa no formal, dado el gran alcance que la TVD ha logrado en nuestro país tanto en cobertura (86% de la población del país en Diciembre de 2012, según [8]) como los niveles sociales a los que llega. Por ejemplo, a través del Plan Operativo de Acceso al Equipamiento para la Recepción de la TVD Terrestre se entregan STB gratuitos a organismos estatales donde se realizan actividades sociales, culturales, y educativas, a organizaciones sociales, a jubilados y a hogares en situación de vulnerabilidad social y económica [9].

## Líneas de investigación y desarrollo

En esta línea de I+D investigamos las características, potencial, limitaciones y herramientas que condicionan o facilitan el desarrollo de contenidos educativos para la TVD.

El proyecto se ha estructurado en etapas que permitan una apropiación incremental del conocimiento y el desarrollo de ideas innovadoras.

**1ª etapa:** Se propuso estudiar las tecnologías y estándares de TVD y analizar las herramientas de desarrollo disponibles.

**2ª etapa:** Se propuso estudiar los conceptos de Aprendizaje y Enseñanza en

general y en el marco de la TVD. A partir de dicho estudio deben plantearse propuestas de desarrollo de aplicaciones educativas para TVD y desarrollo de herramientas para simplificar la creación de contenidos.

**3ª etapa:** Se planea desarrollar los contenidos educativos propuestos en la etapa anterior, realizando además un estudio sobre las prácticas de Ingeniería de Software aplicadas, a fin de definir el ciclo de vida de una aplicación educativa para TVD.

#### **Publicidad Interactiva en TVD**

Una segunda línea de investigación fue agregada durante el año 2012, a fin de desarrollar una arquitectura que permita la incorporación de publicidad interactiva y, posiblemente, *t-commerce* (oferta, compra y venta de bienes y servicios a través de TVD) en cualquier sistema de TVD que siga la norma brasilera mencionada.

### **Resultados y Objetivos**

Hasta el momento de este informe se han completado las dos primeras etapas, obteniendo como resultado un Reporte Técnico sobre TVD y propuestas de desarrollo de aplicaciones educativas.

Al finalizar la tercera etapa, a fines del corriente año, se habrán completado al menos dos aplicaciones educativas en TVD y se publicarán las mejores prácticas de Ingeniería de Software para el desarrollo de contenidos educativos multimediales. Asimismo, se definirán líneas de desarrollo de herramientas que simplifiquen y potencien la creación de dichos contenidos, luego de haber evaluado las herramientas disponibles, tales como CreaTVDigital [10].

En cuanto a la segunda línea de I+D, esto es, la publicidad o *t-commerce* en TVD, se está completando el trabajo final para una Especialidad (tal como se verá en la

siguiente sección) con la propuesta de la arquitectura del software requerido.

### **Formación de Recursos Humanos**

#### **• Becas y Prácticas Supervisadas realizadas en el marco del proyecto:**

- Un alumno realizó la Práctica Supervisada (PS, requisito para obtener el título de Ingeniero en Sistemas de Información) en el marco de este proyecto y actualmente, ya como ingeniero recién egresado, se desempeña como líder de un grupo de desarrollo.

- Un alumno obtuvo una beca de iniciación a la investigación por dos años, financiada por la Secretaría de Asuntos Estudiantiles de la UTN. El alumno se desempeña ahora como líder de un grupo de desarrollos. Actualmente se ha realizado la convocatoria a postulantes a becas para el corriente año.

- Un ingeniero recién recibido realizó una Beca de INiciación a la Investigación y Desarrollo – BINID durante 2012, financiada por la Secretaría de Ciencia y Tecnología de la UTN. Actualmente el ingeniero es líder de un grupo de desarrollo.

#### **• Estudios de posgrado:**

- Dos alumnos de posgrado (un Ing. en Sistemas de Información y un Lic. en Cs. de la Computación) están realizando su Trabajo Final Integrador para la Especialización en Sistemas y Servicios Distribuidos (en FaMAF, Universidad Nacional de Córdoba), como parte de nuestra segunda línea de I+D: **Publicidad Interactiva en TVD**. Se planea defender el Trabajo Final durante el ciclo lectivo 2013.

#### **• Otra formación de RR.HH.**



○ A partir de noviembre de 2012 un grupo de 10 estudiantes de los primeros años de la carrera de Ingeniería en Sistemas de Información han recibido entrenamiento en TVD y su metodología de programación, a fin de formar grupos de desarrollo que realizarán durante 2013 las aplicaciones mencionadas previamente.

## Referencias

- [1] "Televisión digital", Ministerio de Planificación Federal, Inversión Pública y Servicios, 2010.  
[http://institucional.minplan.gov.ar/html/tv\\_digital/](http://institucional.minplan.gov.ar/html/tv_digital/) (visitado 9/3/13).
- [2] Televisión Digital Abierta,  
<http://www.tda.gob.ar> (visitado el 9/3/13).
- [3] "Dados sobre a TV digital no Brasil, atualizados em Setembro de 2012",  
<http://www.dtv.org.br/theoffice/wp-content/uploads/2012/04/Resumo-dos-dados-sobre-TV-digital.pdf> (visitado el 9/3/13).
- [4] <http://www.cantv.com.ve/seccion.asp?pid=1&sid=2353> (visitado el 9/3/13).
- [5] "Televisão digital terrestre - Codificação de dados e especificações de transmissão

para radiodifusão digital. Parte 1: Codificação de dados", Associação Brasileira de Normas Técnicas, 2007.

[6] "Ginga, el middleware para las aplicaciones de TV digital", Alejandro Alvarez, LIFIA, Abril 2010.

[7] "La Universidad Nacional de La Plata libera su adaptación de Ginga", Noticias de SOLAR (Software Libre Argentina),  
<http://www.solar.org.ar/spip.php?article743> (visitado el 9/3/13).

[8] "La Presidenta inaugura 17 nuevas estaciones de TV Digital Abierta", Sala de Prensa, Presidencia de la Nación Argentina,  
<http://www.prensa.argentina.ar/2012/12/13/36920-la-presidenta-inaugura-17-nuevas-estaciones-de-tv-digital-abierta.php>

[9] "Mi TV Digital"  
<http://www.mitvdigital.gob.ar/contenidos/beneficiarios.html> (visitado el 10/3/13).

[10] Arroyo, Schwartz, Cardozo, Tardivo, "CreaTVDigital: Composición Visual de Aplicaciones Interactivas para TV Digital", 10° Simposio sobre la Sociedad de la Información, SSI 2012.

# Resultados del estudio de requerimientos de la Didáctica para una aplicación de Mundos Virtuales en educación

Perez Cota, Manuel

*Área de Lenguajes y Sistemas Informáticos  
Departamento de Informática  
Escuela Universitaria de Ingeniería técnica Industrial  
Universidad de Vigo - Vigo - España  
mpcota@uvigo.es*

Maldonado, Calixto

Etcheverry, Patricia

*Decanato de Investigación.  
Universidad Empresarial Siglo 21 UES21  
Córdoba, Argentina  
cmaldonado@uesiglo21.edu.ar  
patriciaetcheverryperelmuter@gmail.com*

Marcelo Marciszack, Mario Groppo

*Laboratorio de Investigación en Software  
Departamento de Ingeniería en Sistemas de Información  
Facultad Regional Córdoba – UTN  
Córdoba Argentina  
marciszac@gmail.com; groppo@groppo.com.ar*

**Resumen:** Se presentan los resultados del proyecto sobre Mundos Virtuales desarrollado por dos años en la UES21 enumerando los requerimientos didácticos detectados y que deben ser atendidos por el software a desarrollar de mundos virtuales para favorecer el aprendizaje en los estudios medios y universitarios.

**Palabras Claves:** mundos virtuales; moodle; sloodle; didáctica; educación a distancia, aprendizaje colaborativo, MMORPG las palabras clave no coinciden...

## Contexto

El presente trabajo es el resultado del proyecto “Uso de Mundos Virtuales para Educación”, desarrollado en el contexto del Decanato de Investigación de la *Universidad Empresarial Siglo21*, quienes financian la totalidad del mismo.

El proyecto soporta la elaboración de una tesis de doctorado asociada al programa “Ingeniería de Software basado en componentes reusables con aplicación en interfaz hombre máquina” del Ing. Maldonado en la Universidad de Vigo. El tema de tesis específico es el estudio y uso de nuevas tecnologías para su aplicación en educación.

Está lograda la documentación requerida a fin de realizar la presentación en el Fonsoft 2013 para obtener fondos de

financiamiento, y de ese modo lograr una dedicación full time de ingenieros y programadores de software para implementar el desarrollo. Estos resultados serán utilizados como insumo de trabajo en un proyecto a homologar en la Sec. de Ciencia y Técnica de la FRC de la UTN, para desarrollar los aplicativos de servidor y de cliente de mundos virtuales en conjunto entre las dos universidades.

## Introducción

Este trabajo presenta los avances alcanzados en sus respectivas líneas de investigación según el siguiente detalle.

### **Líneas de investigación y desarrollo**

#### **A. Requerimientos para la creación de estrategias didácticas actuales**

Se ha avanzado en el estudio de los fundamentos de Didáctica buscando definir los requerimientos que esta disciplina exige cumplir para lograr un aprendizaje significativo y duradero

La Didáctica es una disciplina en cuyo marco se construye nuevo conocimiento en constante desarrollo y en relación a diferentes contextos de aplicación, ubicando al profesor en rol de docente-investigador quien al mismo tiempo que interviene también aporta hacia la construcción de su corpus o marco

conceptual. Sin olvidar la intencionalidad del dispositivo didáctico en función de provocar cambios en el sentido de aprendizajes significativos en los jóvenes, que les permitirán incorporarse activamente en la cultura y la sociedad.

En el Estudio se ha llegado en forma preliminar a las siguientes conclusiones:

- Se debe seleccionar un contenido curricular del prescripto a fin de valorar su metodología más adecuada.

- Se deben plantear situaciones conflictivas. Desde el formato “juego interactivo” o simil, exigir distintas actividades a ser realizadas en pos de un objetivo dado y así posibilitar distintas estrategias y proveer herramientas para su posible resolución.

- Se deberá introducir, previo estudio de las características socio culturales y etarias de la población de los asistentes, a los mismos mediante un ejemplo corto, sea éste narrado o animado a fin de introducirlos en el efecto sugestivo y predisponerlos del mejor modo para lo que vendrá, es decir, el efecto motivacional.

- Es imperativo considerar las consecuencias didácticas y metodológicas del trabajo cooperativo colaborativo, entre grupo de pares. Es decir, es esencial plantear actividades colaborativas desde las herramientas y actividades a desarrollar.

- Debe existir un registro de la elaboración discursiva en los distintos niveles de los contenidos trabajados. Por ejemplo, si se planteara un juego con postas, o desafíos a modo de diálogo es esencial, llegar a construir colaborativamente la síntesis del trabajo. Así se irá chequeando en espiral dialéctico y se ajustarán las variaciones surgidas. Además el juego también podría avanzar dialécticamente, complejizándose cada vez más a partir de contradicciones y nuevas síntesis a medida que se va recreando el contenido.

- Considerar la creación de un espacio dentro de la plataforma donde se plantee un “recreo”, donde se puedan volcar comentarios, pareceres y opiniones casuales. Ambiente que en general no se considera en la educación a distancia, sin embargo, éste sera un espacio donde podrán surgir cuestiones de importancia, conocido desde la dinámica de grupos, como lo instituyente y adquiriendo valor de verdad frente a lo establecido.

- Finalmente, todo este despliegue desarrollado en MV, Moodle y demas, debe estar contenido, como un componente de una actividad integral que, se asume, debe estar planificada y articulada con acceso a recursos educativos como textos académicos, debates y todo aquello que forme parte de la planificación en su totalidad.

#### B. Adopción de herramientas para lograr el Aprendizaje Colaborativo.

- Elegir inicialmente Second Life [linden lab, 2007] y en forma paralela avanzar con un equipo de desarrollo y las Fuentes de Proyecto RedDwarf para construir y alojar en un servidor, toda la infraestructura para soportar un mundo virtual propio, y así evitar la dependencia del proveedor existente con Second Life.

- Elegir como Software de Administración de Aprendizaje a la plataforma Moodle [Poveda, 2010]

- Instalar como Interfaz a Sloodle [Torres Vallejo, 2012], Es un proyecto de código abierto que integra los ambientes virtuales multiusuario como Second Life junto con la plataforma de gestión de aprendizaje MOODLE. Sloodle provee un amplio rango de herramientas que permite a los mundos virtuales el aprendizaje y la enseñanza inmersiva.

#### C. Estado del Arte del Software de mundos Virtuales

Second Life es un servidor que ha sido lanzado el 23 de junio de 2003, y desarrollado por Linden Lab y es accesible gratuitamente en Internet. Sus usuarios, conocidos como residentes con un inventario de cosas y artefactos para utilizar, al mismo tiempo acceden a SL mediante el uso de uno de los múltiples programas de interfaz llamados viewers (visores), lo cual les permite interactuar entre ellos mediante un avatar [Campazzo, 2010]. Es una plataforma que puede ser utilizada para numerosas actividades entre las que se destaca una importante comunidad y la representación de Juegos de Roles. Si bien su interfaz es similar, SL tiene varias características que lo diferencian del resto, como por ejemplo la creatividad al ofrecer absoluta libertad a sus residentes para lo que pueden hacer ellos y además la cuestión de la Propiedad en cuanto permite ser dueño de tierras para construir, vivir y trabajar en ella pagando un canon fijo anual o mensual, dependiendo de la tierra adquirida. [Campazzo, 2010]

Para construir las herramientas de software de MV que interactuarán con los instructores y estudiantes en la proyectada infraestructura de Mundos Virtuales se estudiaron varias opciones. Una opción es la tecnología ofrecida en licencia Pública del originalmente llamada Project DarkStar que recientemente ha sido renombrada como RedDwarf Server o también la fuente OpenWonderland. Esto representa una solución de middleware de código abierto para el desarrollo del lado del servidor de juegos multijugador masivo en línea. En la actualidad se pueden descargar el código fuente e iniciar un desarrollo independiente debiendo seguir con la licencia Open Source. [Wikipedia,2012]

La definición en cuanto a utilizar la tecnología Open Source para desarrollar un Mundo Virtual propio se basa en mantener el control total de la totalidad de los contenidos y artefactos que, por estar almacenados en servidores de terceros, a merced de un eventual cierre o cambio de dueño por las frecuentes adquisiciones entre empresas del mercado tecnológico, podrían perderse sin posibilidad de reposición.

RedDwarf ofrece una infraestructura en desarrollo, con casos exitosos de mundos virtuales y algún nivel de soporte, expresado en foros de desarrolladores. [Wikipedia,2012]

En forma paralela se está revisando la alternativa de utilizar OpenWonderland, derivado del proyecto Darkstar como el anterior RedDwarf, como alternativa con mayor actualización y porque ha sido adoptado por Universidades de habla hispana . [openwonderland, 2012]

También se observa una redefinición de las plataformas, ya que MMOLE ha sido definido recientemente como “Los ambientes de realidad aumentada o virtual, o mixtas,

construidas en servidores de mundos virtuales que proveen aprendizaje interactivo a través de interfaces en dos dimensiones (2D), dos dimensiones y media (2.5) y tres dimensiones (3D) o completamente inmersivas, apropiadas para crear y administrar plataformas de aprendizaje colaborativo en línea, en los que los individuos participan por ellos mismos o figurativamente a través de un avatar". [milmogame, 2012]

Durante estos últimos años han surgido nuevas plataformas a incluir en los estudios y deben ser integradas al estudio para seleccionar la mejor atención de los requerimientos enunciados, como OpenCroquet, basado en SmallTalk del que se ha podido descargar el software durante mayo del 2012, aunque el sitio de la organización desarrolladora, presenta fallas en su funcionamiento (13).

#### D. Ingeniería de Software

Los entornos virtuales de multi usuarios con acceso masivo (MMVEs sigla en ingles de Massively Multi User Virtual Environments), tales como World of Warcraft, EVE o Second Life se caracterizan por atender a cantidades de usuarios interactuando en gran escala: miles de usuarios se interconectan entre sí, en tiempo real, de a pares o en grupos. La escala de MMVEs afecta en gran medida la capacidad de respuesta y la consistencia del mundo de juego percibida por los jugadores. Cuantos más usuarios y más acciones sean generadas por el usuario y al mismo tiempo compartidos, más difícil será para el servidor, apoyar las interacciones, y la dificultad radica en que esto no aumenta de forma lineal sino cuadráticamente dependiendo del número de usuarios. Por lo anterior los MMVEs son diseñados usando todo tipo de optimizaciones. Estas optimizaciones obstaculizan la escalabilidad y tolerancia a fallos, convirtiéndolos en equipos difíciles de mantener y evolucionar. En general, han habido pocas recomendaciones de características de configuración de arquitecturas MMVE que puedan generalizarse, lo que hace que cada aplicación termine siendo específicamente diseñada para sus necesidades únicas. Se ha analizado bibliografía que orienta a que una arquitectura basada en REST podría soportar los niveles de carga de trabajo que generarían unos pocos miles de usuarios que interactúen entre sí (entre mil y dos mil) [Lopes , 2012]

#### Formación de Recursos Humanos

En los subsiguientes pasos del proyecto se convocara a dos estudiantes avanzados para que cumplan con los

requerimientos de la materia Practica Profesional Supervisada de la carrera de Ingeniería de Software a fin de desarrollar los programas en Java para poner en marcha el Servidor y el cliente dentro de una misma Universidad.

Se prevé también ingresar un proyecto de Investigación y desarrollo en la Universidad Tecnológica Nacional, FRC para hacer participar a docentes y estudiantes de las materias Algoritmos y Estructuras de Datos en la construcción conjunta con los estudiantes de la UES21.

Entre los autores Pérez Cota y Neder son Doctores en Ingeniería de software, Groppo ha finalizado, presentado y aprobado su tesis de Doctorado del programa mencionado, durante el mes de diciembre del año 2011, Maldonado y Marciszack están en la etapa final de elaboración de tesis de Doctorado y Etcheverry es Especialista en Didáctica y TICs

#### Referencias

- [Souto, 1999] Souto, Marta (1999) Grupos y dispositivos de formación. Colección Formación de Formadores. Buenos Aires. Ediciones Novedades Educativas y Facultad de Filosofía y Letras, UBA.
- [linden lab, 2007] Second Life (2007) <http://secondlife.com/whatis/faq.php#02>.
- [Poveda, 2012] Poveda, L. A., & de la Educación, D. F. MOODLE como recurso didáctico.
- [Wikipedia, 2012] Wikipedia Project RedDwarf [http://en.wikipedia.org/wiki/RedDwarf\\_Server](http://en.wikipedia.org/wiki/RedDwarf_Server). Visitado Febrero 2012
- [Torres Vallejo, 2012] Torres Vallejo, Nicolás et al. Creación De Un Metaverso En Opensim Para La Universidad Distrital Dentro De La Red Rita-Ud – Revista Redes de Ingeniería . Vol 3, No 2 (2012)
- [Campazzo, 2010] Campazzo et al - DE LA PRESENCIALIDAD A LA INTERACCION VIRTUAL 3D (2010) Universidad Estatal a Distancia ISSN 1659-4703 C Rica
- [Lopes , 2012] Lopes Cristina V. et al RESTful Massively Multi-User Virtual Environments: A Feasibility Study
- [openwonderland, 2012] <http://openwonderland.org/> visitado en Abril del 2012
- [milmogame, 2012] <http://www.milmogame.com/> Sitio del servidor de Juegos basado en visitado en Abril 2012



## Aplicación de Requerimientos Didácticos y Desarrollo de Software de Mundos Virtuales con licencia Open Source

Calixto Maldonado / Groppo, Mario Alberto / Ligorria, Laura / Ligorria, Karina / Pollioto, Martin Dario / Romani, Germán Ariel / Odetti, María Alejandra  
Facultad Regional Córdoba

Maestro M. López esq. Cruz Roja Argentina – Ciudad Universitaria – Córdoba – Argentina C.P. (X5016ZAA)

calixtomaldonado@hotmail.com  
vigo@groppo.com  
ligorrialaura@hotmail.com  
karinaligorria@gmail.com

### Resumen

**Palabras clave:** Blended Learning, Darkstar, Didáctica, MMO, MMVE, Open Croquet, Open Source.

### Contexto

Es un proyecto de investigación en proceso de homologación por la Secretaría de Ciencia y Técnica de la Universidad Tecnológica Nacional – Facultad Regional Córdoba (UTN-FRC) y forma parte de los proyectos de investigación del área de Ingeniería de Software del Departamento de Ingeniería en Sistemas de Información de la FRC. El grupo está integrado por ingenieros, que en su rol de docentes y consultores dedicados al desarrollo de software, contribuirán al desarrollo del

### Introducción

“La primer generación de e-learning o los programas de aprendizaje basado en la Web, se focalizaron en presentar en Internet contenido instruccional basado en el aula. Además esta primera generación de programas de e-learning (aprendizaje provisto digitalmente) tendían a ser una repetición y compilación de versiones de cursos basados en el aula. La experiencia adquirida de esta primera generación de e-learning, a veces con extensas secuencias de páginas y preguntas con opciones, dieron paso al concepto de que un solo modo de entrega instruccional no puede proporcionar suficientes opciones, compromiso, contacto social, la pertinencia ni el contexto necesario para facilitar el éxito del aprendizaje con alto rendimiento.

En la segunda ola de e-learning, el aumento de diseñadores de aprendizaje que está experimentando con modelos de “blended Learning” o aprendizaje combinado o mixto que combinan varios modos de provisión. La evidencia de los relatos sobre las experiencias

indican que esta modalidad mixta o combinada ofrece más opciones y efectividad” (15)

“Blended Learning (BL), que es una mixtura de aprendizaje en línea y cara a cara, puede combinar los beneficios de ambos entornos, el tradicional en aula y e-learning” (17)

El proyecto se propone lograr, a través de software de Mundos Virtuales (MV), una presencia virtual, requerida en BL y así evitar que los estudiantes tengan que asistir físicamente a las clases para recibir la experiencia presencial.

En las consideraciones relacionadas a la Didáctica, distinguimos a las nuevas corrientes de pensamiento de la didáctica y sus autores más importantes.

Para el primer aspecto Didáctico del proyecto consideramos importante, con el software que se desarrollará, contribuir en lograr una mayor atención y, por ende, mayor recordación, de algunos contenidos formales de la educación, algunos de ellos que deben ser aprendidos para cumplir los objetivos primordiales proponiendo para esto, usar una estrategia integral que tenga la tecnología de Mundos Virtuales como soporte de los métodos pedagógicos estudiados.

Son consideradas entonces las ideas expuestas por Marc Prensky (8) que promueve el uso de los juegos como estrategia de apoyo para transmitir conocimiento a los nuevos estudiantes, a los que llama “nativos digitales” habituados a usar herramientas tecnológicas para jugar y entretenerse en largas horas de ocio.

Edgar Morin (9), afirmaba en su trabajo de 1999 “Los siete saberes necesarios a la educación del futuro” que “En esta evolución hacia los cambios fundamentales de nuestros estilos de vida y nuestro comportamiento, la educación, en su sentido más amplio, juega un papel preponderante. La educación es uno de los

instrumentos más poderosos para realizar el cambio. Uno de los desafíos más difíciles será el de modificar nuestro pensamiento de manera que enfrente la complejidad creciente, la rapidez de los cambios y lo imprevisible que caracterizan nuestro mundo”. Se aplica a este proyecto por que se pretende usar tecnología compleja, actual que cumple con las características de ser un entorno familiar a los nativos digitales, con la inmersividad que tienen los juegos que usan mundos virtuales.

Jean Piaget (10) en “A dónde va la educación” afirmaba que ésta “...debe orientarse hacia una reducción general de las barreras o hacia la apertura de múltiples puertas laterales que permitan a los alumnos (universitarios y secundarios) el libre paso de una sección a otra con la posibilidad de elegir múltiples combinaciones”, lo que viene a confirmar que se debe trabajar en múltiples dimensiones y que éstas estén conectadas. La brecha de comunicación está dada porque las generaciones crecen en compartimientos estancos. En esta misma obra. Piaget, enuncia que uno de los problemas “...es la preparación de los maestros, y que ésta es una cuestión previa a cualquier reforma pedagógica, ya que sin esta preparación los mejores programas y atractivas teorías no será posible llevar a cabo los cambios necesarios”. En este caso se pretende trabajar con materias de la carrera de sistemas de información y obtener los requerimientos pedagógicos que pueden ser atendidos con estos entornos virtuales manejados con el software a construir, para lograr con las teorías actuales y el software un entorno atractivo para los estudiantes.

Sigue Prensky marcando algunos problemas adicionales. "Los inmigrantes digitales (profesores y padres) no creen que los jóvenes puedan aprender exitosamente mientras miran TV o escuchan música, debido a que ellos mismos como inmigrantes, no pueden hacerlo. Los nativos digitales tienen poca paciencia para las lecturas y la lógica paso a paso. Los inmigrantes suponen que los aprendices son los mismos de siempre y que los mismos métodos que funcionaron para sus maestros cuando eran estudiantes, van a funcionar para sus estudiantes”. Sigue Prensky “Hay profesores que piensan que no merece la pena aplicar tecnología en la educación, y no puede ser comparado con todo lo que experimentan a diario con la tecnología. Por todo esto, a menos que queramos olvidar de educar a los nativos digitales hasta que ellos crezcan y lo hagan por sí solos, tenemos que enfrentar este problema y

reconsiderar nuestra metodología y nuestros contenidos.” (9)

La Didáctica es una disciplina en cuyo marco se construye nuevo conocimiento, está en constante desarrollo y en relación a diferentes contextos de aplicación, ubicando al profesor en rol de docente-investigador quien al mismo tiempo que interviene también aporta hacia la construcción de su corpus o marco conceptual. Sin olvidar la intencionalidad del dispositivo didáctico en función de provocar cambios en el sentido de aprendizaje en los jóvenes, que les permita incorporarse activamente en la cultura y la sociedad. Este dispositivo deberá promover, desde las situaciones de enseñanza, transformaciones en lo cognitivo, afectivo y social siempre orientados hacia la transmisión de saberes (15). Desde este punto de vista, el software será el dispositivo y buscará con una interfaz gráfica inmersiva abrir un canal válido de comunicación con el docente y el contenido a aprender.

Este cuadro de situación contiene los fundamentos didácticos del proyecto para buscar pautas metodológicas que aporten contenido formal al uso de mundos virtuales para lograr mejorar los resultados de la educación, aportando un canal de comunicación conocido y atractivo para la actual generación de estudiantes.

Un ingrediente más es que los estudiantes actuales o nativos digitales ya cuentan con equipos. El programa Conectar Igualdad, desde su inicio hasta 2012, ya ha distribuido en Córdoba 110,195 Netbooks a 102,481 Alumnos de 349 Escuelas (6).

Esto significa que en la UTN - FRC existen estudiantes, mayoritariamente provenientes de escuelas públicas de la provincia, que cuentan con equipo con capacidades de navegar en internet y correr software que le permita mejorar su aprendizaje.

Es evidente la necesidad de estudiar la aplicación de tecnología en el proceso educativo como contribución en revertir la crisis social en la que estamos inmersos (12)

En el aspecto de software de Mundos Virtuales (MV), como herramienta a aplicar para apoyar al profesor para generar experiencias vividas de aprendizaje, dirigido a los Nativos Digitales.

Antes de avanzar en el software, podemos presentar a los MV definiendo primero el término Virtual, asociado a los entornos o MV,

por analogía (13) "lo real es lo que ES, lo imaginario es lo que NO ES, y lo virtual NO ES, pero tiene la forma o efecto de lo que ES. Y los MV son lugares donde lo imaginario puede transformarse en algo similar a lo real."

Los MV son implementados por una computadora o red de computadoras que simulan un entorno o ambiente. Algunas, pero no todas, de las entidades en este entorno actúan bajo el control de otras personas. Debido a que varias personas pueden afectar el mismo entorno simultáneamente, al MV se lo denomina compartido o Multi-usuario. El entorno continúa existiendo y desarrollándose internamente, aún cuando no haya gente interactuando, es decir que es persistente.

Es importante destacar que los mundos virtuales no son lo mismo que la Realidad Virtual, ya que ésta es mucho más específica a mecanismos en los que las personas pueden interactuar con simulaciones de computadoras. En la actualidad en MV como Second Life se puede usar dispositivos que realizan tareas específicas, pero el MV sigue siendo el entorno y no sus medios internos para realizar acciones en él.

El MV contiene reglas internas y automáticas, que permiten al usuario efectuar cambios al entorno esto se denomina la Física del mundo. Así encontramos gravedad, viento, iluminación, sonido ambiental. Los usuarios son representados como individuos dentro del mundo. Así, pueden ejercer influencia total o parcial sobre un ejercicio, tripulación o partido, pero hay una sola entidad de juego que los representa en el mundo y con el que el usuario se identifica fuertemente. Este es su carácter, denominado avatar. Esta interacción es en tiempo real, y se espera obtener una retroalimentación inmediatamente. Una sala de chat, no es un mundo virtual, ya que no tiene esta física

Según Richard Bartle (13) la evolución de los mundos virtuales ha tenido etapas y que la distinguen en edades:

La primera edad es desde 1978 a 1985, en la que los juegos basados en MV se llamaban MUDs del inglés Multi User Dungeon.

La segunda edad se desarrolla entre 1985 y 1989 con los primeros juegos típicos MV como Shades de Neil Newel, Gods de Ben Lauries y MirrorWorld.

La tercera edad se desarrolla entre 1989 y 1995 con la difusión del programa AberMUD, por haber sido desarrollado en la Universidad de Gales en Aberystwyth. Fue escrito por Alan

Cox, en lenguaje B, antecesor del C, en un mainframe Honeywell L66 en 1987, pero fue su pase a C en 1989 lo que le permitió posicionarse muy rápidamente al poder ser ejecutado en equipos con Sistemas Operativos Unix. Esto contribuyó a una gran expansión, manifestado en que mediciones en el tráfico del backbone de NSFnet en 1993 mostraron que el 10% del tráfico de bits pertenecía a MUD activos. Puesto en perspectiva, quiere decir que era el 10% del tráfico de Internet, antes de la WWW. (13)

La cuarta edad entre 1995 y 1997 encuentra un estancamiento en Inglaterra y Francia por los costos de teléfono que generaba esta forma de conexión, y un gran desarrollo en EEUU, donde el costo del tráfico de llamadas locales era gratuito.

La quinta edad 1997 al Presente: Los MV gráficos se transformaron de ser multi-jugadores a ser multi-jugadores masivos y comienzan a emerger como recurso de uso "serio" y social más allá del lo meramente lúdico de las tecnologías de MV (14).

Se han analizado software de construcción de MV y su estado actual de desarrollo. Hoy en día, se pueden contar cerca de una docena de MV, consolidados y en aumento, pero por cuestiones de tiempo se ha elegido para realizar ensayos a Second Life ([www.secondlife.com](http://www.secondlife.com)). Second Life (SL) es un servidor lanzado el 23 de junio de 2003, desarrollado por Linden Lab, y que es accesible gratuitamente en Internet. Sus usuarios, conocidos como residentes, cuentan con un inventario de cosas y artefactos para utilizar y que acceden a SL mediante el uso de uno de los múltiples programas de interfaz llamados viewers (visores), lo cual les permite interactuar entre ellos mediante un avatar (2).

Investigadores de la Universidad de Barcelona (7), definen a un sitio definido como paradigmático de MV, con esta descripción "Second Life (SL) es un mundo virtual en 3D que permite interacciones entre usuarios, con los avatares, simulando comportamientos humanos: hablar, bailar, cambiarse de ropa o incluso intimar con otros usuarios, ver documentos y compartirlos, ver vídeos, escuchar audio, visitar lugares imaginarios o recreaciones de localizaciones reales, crear nuevos objetos, comprar y vender servicios, crear una casa y decorarla, recibir o impartir formación, asistir a obras de teatro, manifestarse, visitar museos. La idea de Linden Lab es demostrar la viabilidad de un modelo de economía virtual o sociedad virtual, algo que Philip Rosendale el creador de la empresa, definió: «no estoy construyendo un juego, estoy construyendo un nuevo continente». (11)

Si analizamos las acciones que son posibles en SL veremos que no se puede considerar un juego sino una realidad paralela:

- Crear un personaje y diseñar su apariencia (incluyendo rasgos físicos y ropa)
- Conversar con otras personas (utilizando el teclado o la voz), establecer redes sociales y grupos.
- Interactuar con otras personas (a través de gestos, de movimientos corporales, intercambiando objetos...)
- Visitar lugares o recreaciones virtuales de ciudades existentes, entrar en Edificios, crear un lugar propio.
- Interactuar con los objetos que ya existen (sentarse en un sillón, mover una silla, encender la radio, lanzar un objeto, etc.)
- Crear nuevos objetos (desde una esfera a un edificio).
- Ejecutar vídeos, audio o archivos de presentaciones llamados Machinimia.
- Asistir a conciertos, eventos de presentación de películas, sesiones de moda, obras de teatro.
- Impartir o recibir cursos de formación y/o asesoramiento
- Realizar simulaciones para aprendizaje (una simulación de accidentes, una Simulación de una oficina bancaria...)"

Para construir las herramientas de software que interactuarán con los instructores y estudiantes en la proyectada infraestructura de MVs, se estudiarán varias opciones que ofrecen código abierto en licencia Pública del originalmente llamado Project DarkStar y recientemente renombrado como RedDwarf Server o de otra fuente como es OpenWonderland. Esta es una solución de middleware de código abierto para el desarrollo del lado del servidor de juegos multi jugador masivo en línea. En la actualidad se pueden descargar el código fuente e iniciar un desarrollo independiente debiendo seguir con la licencia Open Source (5).

La decisión de utilizar la tecnología Open Source para desarrollar un Mundo Virtual propio se basa en mantener el control total de todos los contenidos y artefactos que, por estar almacenados en servidores de terceros, a merced de un eventual cierre o cambio de dueño por las frecuentes adquisiciones entre empresas del mercado tecnológico, podrían perderse y no se tendría alternativa alguna de reposición. RedDwarf ofrece una infraestructura en desarrollo, con casos exitosos de MV(3), y algún nivel de soporte, expresado en foros de desarrolladores.

En forma paralela se está revisando la alternativa de utilizar OpenWonderland, derivado del proyecto Darkstar como el anterior RedDwarf, como alternativa con mayor actualización y porque ha sido adoptado por Universidades de habla hispana (5).

También se observa una redefinición de las plataformas, ya que MMOLE ha sido definido recientemente como "Los ambientes de realidad aumentada o virtual, o mixtas, construidas en servidores de mundos virtuales que proveen aprendizaje interactivo a través de interfaces en dos dimensiones (2D), dos dimensiones y media (2.5) y tres dimensiones (3D) o completamente inmersivas, en los que los individuos participan por ellos mismos o figurativamente a través de un avatar" (1).

En estos últimos años han surgido nuevas plataformas a incluir en los estudios y deben ser integradas al análisis para seleccionar la mejor atención de los requerimientos enunciados, como OpenCroquet, que está desarrollado con SmallTalk, del que se ha podido descargar el software durante mayo del 2012, pero el sitio de la organización desarrolladora, presenta fallas en su funcionamiento (4). Sobre esta definición se estudiará OpenCroquet que está desarrollado con SmallTalk ofreciendo una oportunidad de conocer en profundidad este lenguaje, que puede servir en aplicaciones de ejemplo para la cátedra de Paradigmas de Programación de la carrera de Ingeniería en Sistemas de Información de la FRC de la UTN.

IBM ha lanzado múltiples iniciativas en 2007 y 2008, incluyendo los foros de Interoperabilidad de MV, con Linden Labs (desarrollador de SL) pruebas y demos con ActiveWorlds y Forterra, y más recientemente asistiendo a la "tele transportación" desde SL a un MV basado en OpenSim, llamado 'Ruthed5'. (14)

En conclusión desde este contexto actual, es urgente adoptar TICs para mejorar la comunicación entre docentes y estudiantes, los MV presentan perspectivas de interés en su aplicación en educación y es necesario incluir los requerimientos que la didáctica tiene para la implementación de estrategias integradas para mejorar el aprendizaje.

### **Líneas de investigación y desarrollo**

Las líneas de investigación son dos una, hacia la didáctica para rescatar la mayor cantidad de requerimientos que el software debe cumplir y la otra es estudiar y aplicar los mejores métodos de la ingeniería de software para lograr un



software de servidor de MV que soporte un número importante de usuarios simultáneos y un cliente de MV, ambos software serán, cada uno una línea de desarrollo del proyecto

## Resultados y objetivo

Es un grupo que iniciará su tarea a partir de la Homologación, a partir del 01/05/2013, por lo que no se han logrado aún resultados.

Los Objetivos Generales del Proyecto son:

- Investigar los fuentes originales y Desarrollar un software de servidor y cliente que permita aprovechar las características beneficiosas de un Mundo Virtual aplicables en conjunto con las estrategias de la Didáctica moderna para ser una innovación aplicable a contenidos de materias de nivel medio, universitario y laboral.

Los Objetivos Particulares del Proyecto son:

- Estudiar las estrategias Didácticas actuales para obtener requerimientos formales de estas prácticas de los profesores y asistentes a la enseñanza de contenidos de materias curriculares y laborales.
- Seleccionar la plataforma más adecuada de las disponibles.
- Realizar pruebas de concepto de la forma de cumplir los requerimientos con un MV ya existente.
- Diseñar un modelo de datos que permitan almacenar la información referida a las actividades a llevar a cabo dentro del MV
- Desarrollar un software de Servidor y otro de Cliente de mundo Virtual
- Desarrollar las interfaces entre el Software de MV y la plataforma Moodle, y su base de datos

## Formación de Recursos Humanos

Buscamos concretar con el proyecto la formación de profesionales capacitados en el desarrollo de software guiado por las pruebas.

Esta formación deberá incluir desde los aspectos teóricos hasta los de implementación práctica. Pero también buscamos que los integrantes, becarios y tesistas logren adquirir experiencia en la investigación y en el futuro puedan replicar y llevar adelante otros proyectos de investigación.

Creemos que la experiencia que adquieran en este proyecto la pueden aplicar dentro de la universidad en grupos de investigación, cátedras de grado y en el ámbito privado.

Con las habilidades que adquieran también se propiciará que los participantes adquieran categoría de investigador e inicien una carrera dentro de la docencia investigativa.

Otro aporte a la formación de recursos humanos es la elaboración de tesis de maestría de integrantes del proyecto, con lo cual se podrá alcanzar mayores niveles académicos que permitan el progreso en el ámbito docente y profesional.

Realizar tareas de difusión de los tópicos del trabajo y sus avances en las cátedras relacionadas en la UTN-FRC y público en general

## Referencias

- (1) Lorenzo, Carlos Miguel  
[http://es.wikipedia.org/wiki/Massively\\_Multiuser\\_Online\\_Learning\\_Wikipedia](http://es.wikipedia.org/wiki/Massively_Multiuser_Online_Learning_Wikipedia), visitado en 15 Mayo 2012
- (2) Second Life (2007)  
<http://secondlife.com/whatis/faq.php#02>.
- (3) Wikipedia Project RedDwarf Visitado Febrero 2012  
[http://en.wikipedia.org/wiki/RedDwarf\\_Server](http://en.wikipedia.org/wiki/RedDwarf_Server)
- (4) Wwww.opencroquet.org en funcionamiento restringido solo a bajar el SDK v 1.0, ampliar en:  
[http://en.wikipedia.org/wiki/Croquet\\_Project](http://en.wikipedia.org/wiki/Croquet_Project)
- (5) www.openwonderland.org/ visitado en Abril del 2012
- (6) www.conectarigualdad.com.ar visitado en Febrero del 2013
- (7) Mariona Grané, Joan Frigola, Miguel Angel Muras Second Life: Avatares para aprender - Universitat de Barcelona 2007
- (8) Marc Prensky Digital Natives, Digital Immigrants - On the Horizon (NCB University Press, Vol. 9 No. 5, October 2001)
- (9) Edgar Morin, "Los siete saberes necesario a la educación del futuro" UNESCO. París, 1999
- (10) Piaget, Jean "A dónde va la educación" Ed. Teide Barcelona 1978
- (11) Terdiman, Daniel (May 8, 2004). "Fun in Following the Money". Wired Magazine.  
<http://www.wired.com/gaming/news/2004/05/63363?currentPage=all> . Visitado en Febrero 2013
- (12) Maldonado, Calixto ESTUDIO DE MUNDOS VIRTUALES PARA MEJORAR EL APRENDIZAJE (artículo Id 2448) XI Workshop de Investigadores en Ciencias de la Computación, WICC 2009. ISBN 978-950-605-570-7
- (13) Bartle, Richard A. Designing Virtual Worlds. New Riders-California-Pearson Education ISBN 0-1310-1816-7
- (14) Christian Renaud and Sean F. Kane, Esq. Virtual World Industry Outlook 2008-2009.  
[http://digis.ewha.ac.kr/data/project/virtual\\_world\\_outlook\\_20082009.pdf](http://digis.ewha.ac.kr/data/project/virtual_world_outlook_20082009.pdf) Visitado en Febrero 2013
- (15) Souto, Marta (1999) Grupos y dispositivos de formación. Colección Formación de Formadores. Buenos Aires. Ediciones Novedades Educativas y Facultad de Filosofía y Letras, UBA
- (16) Singh, G., O'Donoghue, J., & Worton, H. (2005). A study into the effects of elearning on higher education. Journal of University Teaching & Learning Practice.
- (17) Al-huneidi, Ahmad. Constructivism Based Blended Learning in Higher Education.  
<http://doclib.uhasselt.be/dspace/handle/1942/12273>. Visitado en febrero 2012.

## Descifrando el Sustrato Atencional en el Entorno Telemático: ¿Atención Digitalizada?

Patricia Etcheverry  
Calixto Maldonado  
Universidad Empresarial Siglo 21  
Córdoba – Argentina

patriciaetcheverry@yahoo.com.ar  
calixtomaldonado@hotmail.com

### Resumen

El imaginario social formador de conductas, valores y preferencias se halla enmarcado en un nuevo Paradigma: el entorno telemático. En el área de Educación Media y frente a la implementación del Programa Nacional Digital “Conectar Igualdad” y el modelo 1:1, resulta inminente abordar una Investigación Motivacional de base Cualitativa acerca de los nuevos modos adquiridos por los estudiantes en el procesamiento de la información desde este novedoso entorno social digital. En la renovación de procesos se hallan implicados, la percepción, la atención selectiva y la memoria a corto plazo, condición sine qua non para la apropiación de los contenidos de la enseñanza y la generación de conocimiento, constituyendo todo ello una herramienta esencial para el apoyo didáctico del profesorado

**Palabras clave:** imaginario social, telemático, modelo 1:1, Cualitativa, atención selectiva, herramienta, profesorado

### Contexto

La Escuela Media de Bellas Artes Manuel Belgrano de la Ciudad de Buenos Aires junto con su Equipo Directivo y Cuerpo Docente brinda el apoyo necesario y operativo para llevar adelante una Investigación Cualitativa, tanto en su fase exploratoria como en sus etapas sucesivas previstas. Desde la Universidad Empresarial Siglo 21 se ha desarrollado un equipo de investigación para estudiar los requerimientos didácticos, desde el punto de vista del software, que contemple la problemática del docente a fin de poder desarrollar un software de mundos virtuales que apoye a un proceso de Blended Learning.

### Introducción

#### 1. Imaginario Social y Nuevo Paradigma

El imaginario social, asiento de valores, usos y costumbres, preceptos y preferencias actúa, a partir de una comunidad dada, como regulador de hábitos y conductas [1]. Señala tendencias e ideales hacia los cuales tender a modo de parámetros epocales. Las Tecnologías de la Información y la Comunicación (TICs) han ido influenciando las prácticas sociales configurando un nuevo espacio social con características propias, el de la conectividad digital, definido por Javier Echeverría como Entorno 3[2]

Los estudiantes de nivel medio, todos ellos nativos digitales, son efecto y al mismo tiempo retroalimentan su entorno, al que se lo ha denominado telemático. Hecho paradigmático [3] que también define nuevos modos de vinculación con los procesos de enseñanza y aprendizaje

En palabras de Cecilia Sagol...”la introducción de las TICs en las distintas actividades humanas no implica solamente la realización de las mismas actividades que antes pero por otros canales, sino que modificó rotundamente hábitos, procedimientos y la calidad y cantidad de información, lo que dio lugar a transformaciones profundas, calificadas por algunos como un cambio de paradigma e incluso como el paso de una época a otra: de la sociedad industrial a la sociedad del conocimiento”[4]

En el caso puntual de los procesos de enseñanza y aprendizaje, los estudiantes parecen preferir nuevas modalidades cognitivas ya sea para percibir como para procesar la información, desarrollando así originales habilidades y modos de captar la realidad. Sólo por mencionar

algunas de ellas, la puesta en práctica de la inteligencia lateral para la resolución de casos problemáticos y la nueva competencia social de las multitareas en paralelo colaboran en definir un nuevo sujeto para la educación [5]

En este sentido, los siguientes interrogantes resultan esenciales para el autor:

¿Conocemos realmente al estudiante de Media de hoy? ¿Es el mismo que el de hace unos pocos años?

¿Sabemos hasta qué punto han influenciado las TICs en su modo de ver el mundo, de percibirlo y de aprehenderlo? ¿Sabemos de las nuevas competencias adquiridas o necesariamente por adquirir?

¿Sabemos fehacientemente cuáles son los puentes cognitivos y procesos desde los cuales los jóvenes se vinculan a fin de construir su identidad, mundo propio y proyecto de vida?

## 2. El Programa Nacional de Alfabetización Digital “Conectar Igualdad” dependiente de Presidencia de la Nación

En el marco de la Ley de Educación Nacional N° 26.206 se creó el Programa Nacional de Alfabetización Digital “Conectar Igualdad” que viene implementando el modelo 1:1 – una Netbook por estudiante - a fin de achicar la brecha digital tecnológica y garantizar la inclusión digital, en sintonía con el objetivo nacional de promover la igualdad y justicia social esta vez desde la perspectiva de la educación [6]

El modelo 1:1 en su definición conceptual y funcional remite no sólo a la entrega uno a uno de una computadora personal sino a la impronta de un cambio didáctico y académico sustancial en donde la figura del docente cobra protagonismo como principal ejecutante del proceso [7]

Desde el sitio [www.conectarigualdad.com.ar](http://www.conectarigualdad.com.ar), puede observarse el contador de entrega de Netbooks a nivel Nacional. Al día 8 de marzo de 2013 se han entregado 2.212.000 Netbooks a estudiantes de Colegios Medios de la Republica Argentina.

Teniendo en cuenta la entrega en las 24 provincias argentinas y la tabla de registros del sitio, puede precisarse por ejemplo que sólo en la Provincia de Buenos Aires se han entregado 854.809 Netbooks mientras que en Ciudad Autónoma de Buenos Aires se han entregado 82.820

Durante el 2013, año en curso, el Programa prevé alcanzar una meta de entrega de 3.000.000 de Netbooks habiendo cumplimentado, hasta el presente, el 74% de la entrega prevista.

## 3. La Investigación Cualitativa. Motivación y Predictibilidad de la Conducta

La investigación cualitativa como metodología encargada del desciframiento de motivaciones, hábitos, costumbres, valores, preferencias [8] y en definitiva, los modos en que los individuos se vinculan con su medio cultural y social, es la indicada a fin de decodificar el imaginario social telemático de los nativos digitales, espacio donde se instalan procesos de enseñanza y aprendizaje.

La metodología cualitativa en relación a la didáctica –modos de enseñar- puede aportar básicamente en términos de investigación situada y en contexto, es decir, sus resultados no podrían ser representativos de un universo replicado en términos cuantitativos, pero sí tendría la ventaja de aportar datos significativos y estratégicos en términos de especialización conductual dentro de una comunidad determinada, por ejemplo, los usos, costumbres y preferencias en el procesamiento de la información en un determinado contexto cultural.

Siguiendo con esta pauta metodológica, los datos pueden sistematizarse [9] y resultar en líneas de condensación de sentido pregnantes permitiendo ser reutilizadas con valor de hipótesis en contextos cada vez más amplios de validación aportando al conocimiento de las relaciones humanas en particular y en general.

En términos macro y en acuerdo con la diversidad socio cultural existente en nuestra Nación, esta metodología acuerda respecto de la contemplación de las diferencias culturales y la iniciativa en la planificación Nacional por la inclusión social a nivel República.

En nuestro ámbito, esta metodología se presenta como la oportunidad de esclarecer significados y aportar líneas de acción situadas al quehacer del profesorado frente a la impronta, todavía oculta, de la didáctica por emerger efecto del nuevo escenario telemático.

En terreno y recolectando información motivacional situada y en contexto, esta metodología permitiría elaborar didácticas eficaces para ese universo en particular, y al

mismo tiempo generar la plataforma como para que en el estudiante despierte el deseo genuino por la instrucción y el conocimiento de los saberes socialmente válidos en la adquisición de competencias tanto sociales como laborales. Saberes que en el caso de ser impuestos per se, no sólo no se estaría en línea con las construcciones teóricas más valoradas acerca del aprendizaje, sino que además se tendería a la profundización de la distancia, ausencia de código vincular y abulia del estudiante de media que en general se observa en los Colegios de la Ciudad de Buenos Aires.

### **Líneas de investigación y desarrollo**

En una primera etapa se realizó una investigación cualitativa de carácter piloto exploratoria con construcción de muestra sustentable respecto de las distintas especializaciones del Nivel Medio Superior de los Colegios de la Ciudad de Buenos Aires.

En ese estudio [10] y mediante el instrumento de Focus Group se hizo hincapié en la relación existente y posible entre las Nuevas Tecnologías de la Información y la Comunicación y el Aprendizaje, obteniéndose líneas de sentido generales derivadas de la índole exploratoria de la investigación y posibilitando la creación de hipótesis de trabajo a futuro.

Los ejes de la investigación actual implican la colaboración en la participación protagónica del profesorado en la implementación Nacional del modelo 1:1, y en ese sentido se definen los siguientes propósitos:

Crear espacios de enseñanza y aprendizaje con TICs acorde con las expectativas afectivas y cognitivas de los estudiantes

Esclarecer y brindar líneas de acción al profesorado a fin de favorecer las situaciones de enseñanza y aprendizaje mediante TICs

### **Resultados y objetivos**

Para este contexto en particular, estudiantes de Colegios Medios del Ciclo Superior de la Ciudad de Buenos Aires, surge del estudio exploratorio cualitativo previo y como línea de significación con poder modelizador desde el imaginario social, que la relación entre TICs y aprendizaje no es espontáneo. Más aún, surgen líneas de sentido asociadas a resistencias y el uso de la computadora más vinculada al ocio y a redes sociales que al aprendizaje escolar, al que a veces ni siquiera se lo ve como posible.

Efectivamente, no alcanza con brindar una computadora por estudiante, es decir, el artefacto por sí mismo no genera el conocimiento, sino más bien se presenta la urgencia de investigar los modos en que tomaría nueva forma la impronta de los procesos cognitivos para el aprendizaje enmarcados desde el nuevo entorno y Paradigma Digital.

En cuanto a los objetivos y a partir del diagnóstico obtenido del primer relevamiento exploratorio se propone una segunda etapa de investigación cualitativa. La misma se realizará durante el año 2013 también en el ámbito de Colegios Medios de la Ciudad de Buenos Aires pero sumando valor en cantidad y calidad de los insumos de trabajo. Concretamente mayor cantidad de grupos y mayor ajuste conceptual en los objetivos a investigar.

A modo de hipótesis se supone existirían nuevos y renovados modos de percibir y focalizar la atención -¿atención digital?- debido a la influencia del Paradigma de la Conectividad. Modos que todavía permanecen velados al profesorado y a su práctica didáctica cotidiana

Se debe tener presente que tanto la forma de percibir como la capacidad de seleccionar estímulos y prestar atención selectiva participan del inicio de la secuencia lógico perceptiva, la misma que posibilita la entrada de la información significativa a la memoria a corto plazo y su posible tratamiento hacia la memoria permanente, en definitiva que sea posible apropiarse del conocimiento. En otras palabras, sin percibir y/o atender resultaría estéril plantear la condición de posibilidad del aprendizaje.

En este sentido y en relación a los propósitos antes mencionados, se proponen los siguientes objetivos:

Profundizar en la comprensión de los procesos perceptivo atencionales espontáneos implicados y asociados a las TICs en términos de ocio – juegos en red entre otros- y redes sociales.

Explorar y comprender el comportamiento de la percepción o memoria sensorial en el entorno TICs en relación con situaciones de aprendizaje reales e hipotéticas.

Explorar y comprender el comportamiento de la atención y memoria a corto plazo en el entorno TICs en relación con situaciones de aprendizaje reales e hipotéticas.

Clasificar variables surgidas del campo de la experiencia que puedan ser utilizadas en



estudios cuantitativos a fin de obtener datos de proyección y validez estadística.

Como se ha mencionado, la construcción de la nueva muestra estará en sintonía con el primer estudio cualitativo exploratorio. El instrumento de investigación será también en este caso el de Focus Group aunque se pretende que parte de la muestra se efectivice esta vez mediante grupos de discusión creativos.

Frente al encuadre grupal, la técnica de indagación será la de asociación libre y se utilizarán técnicas indirectas proyectivas que permitan la expresión del imaginario social telemático en toda su magnitud, con sus motivaciones tanto conscientes como inconscientes.

### **Formación de Recursos Humanos**

En el marco de la mejora de la enseñanza (Plan Mejora Institucional (PMI) implementado en los Colegios Medios por el Gobierno Nacional y en este caso puntual, en la Escuela de Bellas Artes Manuel Belgrano, la Prof. Lic. Patricia Etcheverry forma parte de un equipo de trabajo junto al equipo directivo y docentes de la institución. Puntualmente, durante el año 2013 llevará adelante junto con el Coordinador del Área de Arte, el Vice Director y la Asesora Pedagógica la implementación de un Proyecto de Aplicación Didáctica en Herramientas TICs a ser dictado al resto de la comunidad docente del colegio frente a la llegada inminente de las Netbook

La Prof. Lic. Patricia Etcheverry expuso una serie de dos clases teóricas expositivas bajo el título "Escuela Media y TICs" a los docentes formadores de profesores de Psicología. Se trata de la Cátedra Didáctica Especial y Prácticas de la Enseñanza a cargo de la Prof. Adjunta (Int.) Livia García Labandal, correspondiente al último año del Profesorado de Enseñanza Media

y Superior en Psicología, de la Facultad de Psicología de la Universidad de Buenos Aires (UBA) [11]

### **Referencias**

- [1] Castoriadis, Cornelius; Nasart, Pierre y otros. *El Imaginario Social*. Colombo, Eduardo; Ed. Tupac, 1989
- [2] Echeverría, Javier. *Cosmopolitas domésticos*. Barcelona: Anagrama, 1995
- [3] Kuhn, Thomas. *La estructura de las revoluciones científicas*. ESTUDIOS. Filosofía-historia-letras Primavera 1985
- [4] [6] [7] Sagol, Cecilia. *El Modelo 1 a 1: notas para comenzar*. 1º ed. Buenos Aires: Ministerio de Educación de la Nación, 2011
- [5] M. Prensky, Marc, *Mom don't bother me, I'm Learning*. 2006
- [8] Flick, Uwe. *Introducción a la Investigación Cualitativa*. Ed. Morata S.L. Madrid. Abril 2004
- [9] Anselm Strauss. *Juliet Corbin. Bases de la investigación cualitativa: técnicas y procedimientos para desarrollar la teoría fundamentada*. Ed. Universidad de Antioquia. 2002
- [10] Etcheverry, Patricia. *Imaginario Social y Construcción de Subjetividad desde el Entorno Telemático: Hacia una Investigación Posible*. Publicado en Revista TE&ET (Argentina). 2013
- [11] Blog de aplicación de TICs en educación [www.didacticaencontextotics.blogspot.com.ar](http://www.didacticaencontextotics.blogspot.com.ar) visitado en Marzo 2013

## El desafío de las nuevas tecnologías de la información y las comunicaciones en los contextos educativos.

Hugo Ramón, Claudia Russo, Mónica Sarobe, Mariana Saenz, Tamara Ahmad, Leonardo Esnaola, Nicolás Alonso, Paula Lencina, Juan Pablo Tessore

Instituto de Investigación y Transferencia en Tecnología (ITT), Escuela de Tecnología, Universidad Nacional del Noroeste de la Provincia de Buenos Aires (UNNOBA)

{hugoramon, crusso, monicasarobe, marianasaenz, tamaraahmad}@unnoba.edu.ar,  
{leonardo.esnaola, nicolas.alonso, paula.lencina, juanpablo.tessore}@nexo.unnoba.edu.ar

### Resumen

Se presenta un proyecto de I+D+i (Investigación, Desarrollo e innovación) centrado en las tecnologías educativas y su relación con el contexto. En el marco del presente proyecto y considerando que las tecnologías actuales plantean nuevos paradigmas que implican un impacto directo en los modelos educativos, se investigará sobre temas relacionados con las tecnologías informáticas aplicadas a la educación, haciendo énfasis en las diversas variantes y evoluciones del e-Learning como {t, c, u, m}-Learning entre otras.

Se propone además un análisis de técnicas y herramientas que permiten la interconexión de los ambientes virtuales de aprendizaje con los Entornos Virtuales 3D; como así también el desarrollo de aplicaciones informáticas que faciliten el uso de Laboratorios Virtuales centrados en la Web, a ser utilizados en los procesos de enseñanza y aprendizaje del ámbito educativo en forma presencial, no presencial (e-Learning) o mixta (b-Learning).

Se indagará también sobre distintas metodologías y técnicas que permitan dotar de interoperabilidad a los sistemas de información presentes en la UNNOBA.

Además se relevarán e investigarán las capacidades y posibilidades del desarrollo de soluciones computacionales ubicuas aplicadas al binomio enseñanza y aprendizaje fomentando la innovación en los ambientes virtuales de aprendizaje mediante el uso de sistemas ubicuos.

Se estudiarán modelos de gestión de calidad y heurísticas de accesibilidad aplicadas a sistemas de e-Learning y e-Government.

**Palabras clave:** eLearning, eGovernment, Calidad, Entorno Virtual, Ubicua.

### Contexto

Este proyecto fue acreditado mediante evaluación externa y financiado por la Secretaría de Investigación, Desarrollo y Transferencia de la UNNOBA en el marco de la convocatoria a Subsidios de Investigación Bianuales (SIB2013) y es una continuación del proyecto “UNNOBA Virtual. Una plataforma para la integración de sistemas, metodologías y herramientas de enseñanza y aprendizaje”, acreditado en las convocatorias Subsidios de Investigación Bianuales (SIB2010).

El proyecto se desarrolla en el Instituto de Investigación en Tecnologías y Transferencia (ITT) dependiente de la mencionada Secretaría, y se trabaja en conjunto con la Escuela de Tecnología de la UNNOBA.

Este proyecto se coordina con otros dos proyectos presentados por el Instituto de Investigación y Transferencia en Tecnología (ITT), uno de ellos relacionado con el desarrollo y aplicaciones de los sistemas ubicuos, y el otro con la gestión de contenidos digitales y manual de buenas prácticas para ambientes educativos, demostrando una clara interdisciplinaridad y transversalidad del proyecto presentado.

### Introducción

La utilización de las nuevas tecnologías aplicadas a la educación han contribuido a la expansión de nuevas metodologías de enseñanza y aprendizaje entre las que podemos mencionar el {e,c,u,m}-Learning. La enseñanza universitaria cada vez más internacionaliza estas metodologías que generalmente complementan, en unos casos, y sustituye, en otros, a la educación presencial tradicional.

Las redes virtuales de educación que se han ido desarrollando por medio de internet aparecen como resultado de un largo proceso y dan cuenta de las relaciones que han ido aconteciendo entre los medios de comunicación y las transformaciones sociales, y los modelos educativos.

La sociedad actual está inmersa en un mundo de productos tecnológicos, este hecho está directamente relacionado con la revolución científico-técnica iniciada en el siglo XX y que se profundiza en el siglo XXI [1].

Los procesos de enseñanza y aprendizaje se han hechos más complejos gracias al auge de

las nuevas tecnologías, especialmente de internet. Las posibilidades de estos entornos han despertado el interés de docentes e investigadores por conocer y profundizar las condiciones y características de nuevas formas de enseñar y aprender. Se observa en el contexto actual de educación superior que los docentes están optando por la modalidad de aprendizaje combinado como estrategia para incorporar la virtualidad en la propuesta de enseñanza [2].

Esta nueva forma de aprender permite además la desaparición de las condiciones espacio-temporales y las posibilidades de conectarse en espacios dispersos geográficamente.

Durante las últimas décadas ha crecido notablemente el interés por la EaD y esto se ve reflejado en las numerosas publicaciones, congresos y experimentos que tienen lugar tanto en el ámbito nacional como internacional. De hecho, muchas universidades tradicionales han empezado a experimentar, después de mucho tiempo de haber ignorado este paradigma, con EaD [3]. Tal es el caso de las universidades de Oxford, Stanford, Yale y Princeton, con su proyecto de interconexión para crear una universidad en línea para sus egresados; o la interconexión entre la Universidad de Cambridge y el Massachusetts Institute of Technology (MIT) y su proyecto “*el puente de las mentes*”. Otto Peters (Profesor emérito de la FernUniversitaet in Hagen), sostiene que la razón principal de este creciente interés por la EaD radica en el sorprendente avance en las telecomunicaciones y expresa que, para los educadores a distancia, hay una serie de innovaciones asombrosas que cobran importancia: las mejoras en la tecnología de las computadoras personales, la tecnología multimedia, la tecnología de compresión digital de video e Internet. Sumadas a otras

tecnologías hacen posible una serie de ventajas logísticas y pedagógicas inesperadas, como la entrega rápida de formación a cualquier hora y a todas partes, verdaderas posibilidades para el aprendizaje autónomo, calidad de los programas y eficacia en el aprendizaje. Estas sorprendentes ventajas conllevan, al mismo tiempo, numerosos problemas de los cuales hay que ocuparse. En este sentido, el autor expresa: “... *entre más experimentos hagan los educadores a distancia con la informática y los medios de comunicación, más conscientes se vuelven de que la educación a distancia está metida en un proceso de cambio estructural profundo*”. Un cambio que no solo tiene que ver con los medios a utilizar, sino también con cómo utilizar esos medios, los métodos y los contenidos.

La Web 3.0 ha modificado la forma en la que nos comunicamos en Internet, nadie puede negar que, al menos en parte, la irrupción de las denominadas Redes Sociales, como Facebook y Twitter, que han afectado nuestra forma de informarnos y relacionarnos. Estamos viviendo en una sociedad muy cambiante, y el sistema educativo no puede permanecer ajeno a este hecho. Algunos autores [4], sostienen que el sistema educativo actual está más influenciado por la Era Industrial que por la Era Digital. Una cita del educador norteamericano William T. Harris describe con absoluta claridad esta afirmación: “... *el primer requisito de la escuela es el Orden: cada alumno debe ser enseñado primero y principal para adecuar su comportamiento a un estándar general.*” Y continúa, “[...] *El alumno debe cumplir sus deberes en el tiempo fijado, debe levantarse al sonido de la campanilla, moverse en línea, retornar; en suma, realizar todos los movimientos con igual precisión*”. Sin embargo, los procesos industriales se desarrollan hoy muy diferentemente a aquellos del siglo XVIII, y se basan

fuertemente en el conocimiento mucho más que en la fuerza de las máquinas. En este sentido, los mismos autores se preguntan: “*¿Por qué seguimos utilizando un sistema educativo que mantiene las reglas creadas para una cultura y una sociedad que ya no existe?*”, parafraseando a los autores, hoy los estudiantes deben adquirir otras destrezas, como habilidades para la toma de decisiones, trabajo en grupo, relacionarse e interactuar con personas de diferentes culturas, manipular grandes volúmenes de información disponible, pudiendo discernir entre aquella que resulta útil para un determinado propósito de aquella que no lo es, transformar dicha información en conocimiento y poner ese conocimiento al servicio de su propio desarrollo personal y profesional.

Manuel Castells (profesor de Sociología y director del Internet Interdisciplinary Institute de la Universitat Oberta de Catalunya) planteaba: “*Si hay una institución que puede ser transformada y provocar transformaciones es la educación. Lamentablemente es la menos transformada y puede apreciarse un desfase creciente entre los niños y su entorno pedagógico. Muchos jóvenes [y los propios docentes] viven en un entorno digital que no encuentran en la escuela. Aquí hay un problema institucional que no se reduce al equipamiento tecnológico, sino que implica la necesidad de un cambio cultural y organizativo. No hay que culpabilizar a los docentes sino ayudarlos, puesto que sin educadores no hay enseñanza*”. Uriel Cukierman (Especialista en Sistemas de Información y Master en Tecnologías por la Información de la Universidad Politécnica de Madrid) sostiene que los estudiantes forman parte hoy de nuevas generaciones acostumbradas a manejarse con las TIC de una manera cuasi innata y a aprovecharlas para modificar la forma en la cual se relacionan con sus pares,



acceden a la información y al conocimiento, trabajan y, en síntesis, participan de la vida en sociedad.

Hoy en día los nuevos dispositivos móviles tienen la capacidad de integrarse a conexiones de Internet (GPRS, 3G, 4G, WAP, entre otras), lo que ha permitido que estos aparatos se conviertan en verdaderos PCs, posibilitando comunicaciones inalámbricas de alta calidad. Con el avance de los dispositivos móviles, los dispositivos que integran la informática entorno a las personas y la masificación de la televisión digital se desarrollan nuevas metodologías que están revolucionando el ámbito de la educación a distancia, dando origen a nuevos modelos tales como {t, m, u}learning.

El concepto de e-government o administración electrónica ha experimentado en los últimos años un gran desarrollo, esto fue acompañado por el avance de las TIC. Es importante destacar la importancia que tiene las mencionadas tecnologías para la gestión pública moderna. En tal sentido a partir del año 2000 se ha realizado en diferentes países un considerable esfuerzo para lograr la incorporación de los avances de la informática y la conectividad de Internet en forma integral a la gestión de los distintos servicios públicos, en la perspectiva de avanzar hacia un Estado Moderno, ágil y eficiente. Las universidades e instituciones educativas de todos los niveles han ido incorporando este concepto en su desarrollo tanto para ser aplicado en educación como en la gestión de la tecnología.

### **Líneas de investigación y desarrollo**

El actual proyecto se centrará en tres líneas de investigación y desarrollo principales, planteándose para cada una de ellas una serie de objetivos específicos, metas y resultados esperados:

*El e-Learning 3.0, variantes y evoluciones: T-learning, C-learning, U-learning y M-learning:*

- Investigar las singularidades y aplicaciones de los sistemas T-learning, C-learning, U-learning y M-learning.
- Relevar herramientas de la Web 3.0 aplicables al campo de la educación.
- Investigar aplicaciones, desarrollos y experiencias sobre Entornos Virtuales 3D y herramientas informáticas orientadas al desarrollo de laboratorios virtuales y remotos, realizando, al mismo tiempo, un análisis de ventajas y desventajas de su utilización en los sistemas de e-Learning.
- Analizar y recolectar experiencias sobre estudios que investiguen el uso de agentes pedagógicos en distintos niveles de la educación para posteriormente proponer y desarrollar un modelo de agente pedagógico con capacidad de ser integrado en un sistema de e-Learning presente en la UNNOBA.
- Relevar, seleccionar y analizar protocolos y tecnologías que permitan, por medio de la implementación de un módulo, dotar de interoperabilidad a los sistemas de información presentes en la UNNOBA, tales como su gestor integrado de bibliotecas y su sistema de gestión de aprendizaje.
- Seleccionar un portfolio de aplicaciones de e-Learning 3.0 que ofrezcan funcionalidades específicas y cubran metodologías pedagógicas concretas.

- Implementar actividades colaborativas, utilizando un Entorno Virtual 3D, con capacidad de ser aplicadas en los diferentes niveles de educación presentes en la UNNOBA.
- Implementar materiales educativos digitales y aplicarlos en los diferentes niveles educativos de la UNNOBA, incorporando diferentes metodologías de diseño estudiadas en el marco del proyecto.

*Gestión de calidad y heurísticas de accesibilidad aplicadas a sistemas de e-Learning y e-Government:*

- Investigar las singularidades de los principales sistemas de e-Government.
- Investigar los modelos de gestión y los principales estándares de calidad aplicados los sistemas e-Learning y e-Government.
- Releva las principales heurísticas de accesibilidad aplicadas a los sistemas de e-Learning y e-Government.
- Adquirir capacitación en el dominio de conceptos y herramientas que permitan planificar y gestionar la calidad aplicada a los sistemas de e-Learning y e-Government.
- Proposición de un modelo de gestión de calidad aplicable a los sistemas de e-Learning y de e-Government presentes en la UNNOBA.
- Proponer recomendaciones de accesibilidad en base a determinadas heurísticas que sean aplicables a los sistemas de e-Learning y e-Government presentes en la UNNOBA

*Innovación en los ambientes virtuales de aprendizaje: los sistemas ubicuos y los dispositivos móviles:*

- Recopilar y analizar bibliografía, revistas y publicaciones referentes a tecnologías de computación ubicua aplicables al desarrollo de ambientes de aprendizaje.
- Releva y documentar la infraestructura tecnológica presente en UNNOBA con capacidades de aplicación en ambientes de aprendizaje ubicuos.
- Analizar, releva y aplicar distintos estándares al desarrollo de soluciones ubicuas que permitan la interoperabilidad con las demás tecnologías y sistemas presentes en el ámbito de la UNNOBA.
- Contactar y conformar un grupo reducido de estudiantes de la UNNOBA que habrán de probar las soluciones creadas; verificando su correcto funcionamiento a través del uso de distintos dispositivos móviles.
- Implementar un instrumento de medición, y distribuirlo entre los integrantes del grupo de pruebas, para obtener una retroalimentación respecto de las soluciones implementadas.

### **Resultados y Objetivos**

Se espera que los aspectos investigados contribuyan al fortalecimiento de las experiencias de enseñanza y aprendizaje llevadas a cabo por nuestra universidad.

Al mismo tiempo, se pretende propiciar un marco formal para el establecimiento de líneas de investigación en estas temáticas, atendiendo a la formación de recursos humanos, en su rol de investigadores o partícipes activos en equipos de investigación; fomentando la culminación de

sus estudios superiores, promoviendo la redacción, exposición y defensa de Trabajos Finales de Grado y Postgrado, y la realización de Prácticas Profesionales Supervisadas.

Así mismo, al finalizar el proyecto se buscan obtener los siguientes resultados:

*Generar:*

- Informes técnicos consignando las diferentes tecnologías aplicables al ámbito del e-Learning, y sus derivados, tales como T-Learning, C-Learning, U-Learning y M-Learning, y e-Government, destacando aquellas que se consideren más aplicables para la UNNOBA.
- Un informe técnico con recomendaciones de accesibilidad, en base a determinadas heurísticas, aplicadas a los sistemas de e-Learning y e-Government.
- Un modelo de un posible Entorno Virtual 3D aplicable a los procesos de EaD de la UNNOBA, así como también un listado de posibles actividades colaborativas que hagan uso de este entorno, aplicadas a los diferentes niveles educativos de la UNNOBA (pre-grado, grado y posgrado).
- Una serie de recomendaciones para el diseño de materiales digitales adaptados al sistema de gestión del aprendizaje de la UNNOBA.
- Un informe técnico conteniendo recomendaciones para la implementación de distintos sistemas computacionales ubicuos susceptibles de ser utilizados con las distintas tecnologías presentes en la UNNOBA.

*Desarrollar:*

- Un modelo de agente pedagógico con capacidad de ser integrado en un sistema de e-Learning presente en la UNNOBA.
- Un módulo que permita dotar de interoperabilidad al gestor integrado de bibliotecas y su sistema de gestión de aprendizaje.
- Una aplicación computacional ubicua que pueda ser utilizada en los procesos de enseñanza de la UNNOBA y verificar su implementación mediante trabajo de campo. Al mismo tiempo, se pretende generar un documento que resuma la experiencia y permita extraer conclusiones.

*Proponer:*

- Un modelo utilizando herramientas de la Web 3.0 para publicar contenidos educativos de los diferentes niveles educativos de la UNNOBA, entre los que se destacan pre-grado, grado, posgrado y extensión.

### **Formación de Recursos Humanos**

En esta línea de I/D se ha concluido tres tesinas de grado y hay dos trabajos de grado que se están desarrollando y se vinculan con el proyecto.

Se espera desarrollar cuatro tesinas de grado y tres tesis de magister, en el área de Tecnología y Educación, dirigidos por miembros de este proyecto.

En esta misma línea se han obtenido dos Becas de Entrenamiento y dos Becas de Estudio, ambas otorgadas por la Comisión de Investigaciones Científicas (CIC), y una beca de estímulo a las vocaciones científicas otorgada por el Consejo Interuniversitario Nacional (CIN).

## Referencias

[1] Patricia Halaban. "La comunicación virtual en educación a distancia, un estudio en interacciones comunicacionales y procesos pedagógicos en internet". CICCUS Febrero 2010.

[2] Stella Briones. "Delegados y Horizontes para el siglo XXI. Problemas y desafíos siempre vigentes de la enseñanza y aprendizaje a distancia en nuestros contextos institucionales". RUEDA 2011.

[3] Peters O. La educación a distancia en transición. Nuevas tendencias y retos. Primera Edición. Universidad de Guadalajara. Guadalajara, México. 2002.

[4] Cukierman U., Virgili J. La tecnología educativa al servicio de la educación tecnológica. Experiencias e Investigaciones en la UTN. Universidad Tecnológica Nacional. edUTecNe. Argentina. 2010.

## Bibliografía

- Casamayor, G.; Alós, M.; y otros. La Formación On-line, Una mirada integral sobre el e-learning, b-learning. Primera edición. Editorial Graó. Barcelona, España. 2008.
- Krumm J., Creating a New Book of Tutorials: Ubiquitous Computing Fundamentals, Microsoft Research Redmond, UbiComp 2009.
- Lion, C. Imaginar con tecnologías, Relaciones entre tecnologías y conocimiento. Primera edición. Editorial Stella. Buenos Aires, Argentina. 2006.
- Barberà, E. La incógnita de la Educación a Distancia. Primera edición. Editorial Horsori. Barcelona, España. 2001.
- Carmona M.; González S.; Castro Ruiz, Innovación Tecnológica en Comunicaciones Móviles Desarrollada Con Software Libre: Campus Ubicuo Novática: Revista de la Asociación de Técnicos de Informática, ISSN 0211-2124, N°. 190, 2007.
- Litwin, E.; Maggio, M.; Lipsman, M. Tecnologías en las aulas, Las nuevas tecnologías en las prácticas de la enseñanza. Casos para el análisis. Primera edición. Editorial Amorrortu. Buenos Aires, Argentina. 2005.
- Bravo C.; Redondo M.; Ortega M.; Bravo J. Evolución de un Entorno Colaborativo de Enseñanza Basado en Escritorio hacia la Computación Ubicua, (2002) COLINE'02 Investigación en Entornos de Interacción Colectiva, Workshop de Investigación sobre nuevos paradigmas de interacción en entornos colaborativos aplicados a la gestión y difusión del Patrimonio cultural. Granada, 11 y 12 de Noviembre del 2002.
- Litwin, E. Tecnologías educativas en tiempos de Internet. Primera edición. Editorial Amorrortu. Buenos Aires, Argentina. 2009.
- Levis, D.; Gutierrez Ferrer, M. ¿Hacia la herramienta educativa universal?, Enseñar y aprender en tiempos de Internet. Editorial Ciccus. Buenos Aires, Argentina. 2000.
- Manzanedo, J. El e-Learning en España, Modelos actuales y tendencias de actuación. Primera edición. Colección EOI Tecnología e Innovación. España. 2003.
- Weiser M. The Computer for the Twenty-First Century. Scientific American, Vol. 265, No. 3., pp. 94-104. 1991.



## **Innovación en los ambientes virtuales de aprendizaje: Las TIC y su relación con los sistemas ubicuos.**

Hugo Ramón, Claudia Russo, Mónica Sarobe, Daniela Pérez, Leonardo Esnaola, Nicolás Alonso

Instituto de Investigación y Transferencia en Tecnología (ITT), Escuela de Tecnología, Universidad Nacional del Noroeste de la Provincia de Buenos Aires (UNNOBA)

{hugoramon, crusso, monicasarobe, danielaperez}@unnoba.edu.ar, {leonardo.esnaola, nicolas.alonso}@nexo.unnoba.edu.ar

### **Resumen**

Para las universidades de hoy en día es imprescindible la interacción con las aplicaciones que ofrecen las Tecnologías de la Información y la Comunicación (TIC), en particular con aquellas relacionadas a la educación.

Las TIC pueden ayudarnos a integrar y potenciar diferentes instancias del proceso enseñanza-aprendizaje. En tiempos de la ubicuidad, de los dispositivos móviles y del aprendizaje en red, podemos complementar diferentes recursos y herramientas digitales para ampliar la comunicación y el intercambio de saberes.

La línea de investigación presentada pretende implementar un ambiente de aprendizaje ubicuo, a través de la identificación, utilización y adaptación de las distintas tecnologías disponibles en la UNNOBA, con el propósito de potenciar la experiencia de aprendizaje del estudiante a través de los distintos entornos educativos medidos por tecnologías que la universidad ofrece.

**Palabras clave:** Sistemas ubicuos, TIC, EVEA.

### **Contexto**

Esta línea de investigación fue aprobada por la Comisión de Investigaciones Científicas (CIC) y forma parte del proyecto “*El desafío de las nuevas tecnologías de la información y las comunicaciones en los contextos educativos*” que fue acreditado mediante evaluación externa y financiado por la Secretaría de Investigación, Desarrollo y Transferencia de la UNNOBA en el marco de la convocatoria a Subsidios de Investigación BIANUALES (SIB2013) y es una continuación del proyecto “*UNNOBA Virtual. Una plataforma para la integración de sistemas, metodologías y herramientas de enseñanza y aprendizaje*”, acreditado en las convocatorias Subsidios de Investigación BIANUALES (SIB2010).

El proyecto se desarrolla en el Instituto de Investigación en Tecnologías y Transferencia (ITT) dependiente de la mencionada Secretaría, y se trabaja en conjunto con la Escuela de Tecnología de la UNNOBA.

El equipo está constituido por docentes e investigadores pertenecientes a los departamentos de Informática y Tecnología, Humanidades y Afines y Complementarias y estudiantes de las carreras de Licenciatura en

Sistemas, Ingeniería en Informática y Diseño Gráfico de la UNNOBA.

### Introducción

En los últimos años han surgido una enorme variedad de desarrollos tecnológicos, muchos de los cuales han sido incorporados exitosamente en las experiencias educativas y, particularmente, en los entornos virtuales de enseñanza y aprendizaje. Aunque, cabe aclarar, es un hecho innegable que esa irrupción continua de desarrollos tecnológicos plantea un gran desafío a la hora de adoptarlos (y adaptarlos), de manera tal de enriquecer las experiencias de aprendizaje mediado por tecnologías.

Particularmente, en la última década, hemos ido incorporando nuevos dispositivos tecnológicos a nuestra vida diaria, en especial, dispositivos de comunicación: teléfonos celulares inteligentes, notebooks y netbooks, tabletas, etc. Gracias a estos dispositivos, estamos “conectados” más tiempo que nunca antes y no es difícil reconocer que, en mayor o menor medida, nuestra forma de relacionarnos ha cambiado, nuestras costumbres y por qué no también, nuestra forma de aprender.

Hay que reconocer también que no siempre se puede tener lo último en tecnología, de hecho hay personas que no pueden acceder a ella, porque es costosa y no es un artículo de primera necesidad, o, tal vez, porque ni siquiera se tiene acceso a la infraestructura para utilizarla provechosamente. Sin embargo, es una realidad que los gobiernos y las instituciones están haciendo un esfuerzo por hacer llegar la tecnología a más y más personas. Este es el caso de nuestro país, podemos citar programas como “Conectar Igualdad”, que tiene como meta reducir las brechas digitales, educativas y sociales en toda la extensión de nuestro país. El

programa plantea distribuir tres millones de netbooks entre cada uno de los estudiantes y docentes de educación secundaria de escuela pública, educación especial e institutos de formación docente [1]. También hay iniciativas en las instituciones de formación universitaria, como es el caso de UNNOBA y su programa de distribución de netbooks entre sus estudiantes y docentes.

De esta manera, el acceso a la tecnología se está resolviendo, el desafío ahora es implementarlos de manera provechosa para explotar su máximo potencial en los procesos de aprendizaje mediados por tecnología. En este sentido, un concepto que cobra mayor relevancia a medida que aparecen más y mejores TIC, es el de “computación ubicua”.

*El concepto de ubicuidad en las TIC, fue introducido por Mark Weiser en 1988 y adquirió reconocimiento mundial en 1991 con el trabajo “The Computer for the Twenty-First Century”. “...Se refiere, en general, a la presencia de una entidad en todas partes; pero en la computación adquiere la característica de ser, además, invisible. Este paradigma pretende brindar sistemas de cómputo inteligentes que se adapten al usuario, y cuyas interfaces permitan que éste realice un uso intuitivo del sistema. De allí que la meta, de la computación ubicua, de integrar varias computadoras (dispositivos) al entorno físico busca habilitar los beneficios de éstas y de la información digitalizada en todo momento y en todas partes.*

*La influencia de la computación ubicua es en gran parte de carácter teórico, y dadas las condiciones de integración entre educación y sistemas de cómputo explicada antes, comparte créditos junto a otros conceptos como usabilidad, interactividad e hipertextualidad. Una posible aplicación en*

*educación a distancia, sería que el docente presentara el conocimiento en varias formas que se diferencien en sus características técnicas y didácticas. Así el alumno ejercerá su poder de decisión para optar por una vía conveniente de aprehensión del conocimiento, con el consecuente mejoramiento de la calidad en el aprendizaje.” [2].*

Weiser [3] menciona que: “*vamos camino a ambientes ubicuos, ambientes poblados de numerosos sensores que gracias a la miniaturización de los dispositivos son invisibles al usuario y están en permanente rastreo de la actividad humana*”. En estas palabras de Weiser radica el objetivo del presente trabajo, pues se pretende llevar a cabo un relevamiento de las diferentes herramientas tecnológicas disponibles en el ámbito de UNNOBA, integrarlas para conseguir que sean interoperables y lograr explotar al máximo su potencialidad, en provecho de las necesidades de aprendizaje de los estudiantes, sin intervención humana. No se está hablando de remplazar el rol crítico que cumple un docente ni al docente mismo, sino simplemente integrar y potenciar las tecnologías disponibles para crear un ambiente virtual de aprendizaje más adaptado a las preferencias de cada usuario, pero sin la necesidad de contar con un docente distinto para cada estudiante y al servicio de ese único estudiante.

### **Líneas de investigación y desarrollo**

Se mencionan aquí las principales líneas de investigación y desarrollo abordadas en el marco del proyecto:

- Recopilar y analizar bibliografía, revistas y publicaciones referentes a tecnologías de computación ubicua aplicables al desarrollo de ambientes de aprendizaje ubicuos.

- Relevar y documentar la infraestructura tecnológica presente en UNNOBA con capacidades de aplicación en ambientes de aprendizaje ubicuos.
- Analizar y relevar los distintos estándares que conciernan a las tecnologías identificadas y que permitan la interoperabilidad con otras tecnologías y sistemas presentes en el ámbito de UNNOBA.
- Aplicar los estándares relevados al diseño y desarrollo de por lo menos dos soluciones ubicuas.
- Contactar y conformar un grupo reducido de estudiantes que habrán de probar las soluciones creadas.
- Distribuir las soluciones ubicuas desarrolladas entre el grupo de pruebas, con el objetivo de ponerlas en práctica y verificar su funcionamiento.
- Implementar un instrumento de medición, y distribuirlo entre los integrantes del grupo de pruebas, para obtener una retroalimentación respecto de las soluciones implementadas.
- Analizar los resultados obtenidos en base al instrumento de medición seleccionado y obtener conclusiones al respecto.

### **Resultados y Objetivos**

El objetivo consiste en implementar un ambiente de aprendizaje ubicuo, a través de la identificación, utilización y adaptación de las distintas tecnologías disponibles en UNNOBA, con el propósito de potenciar la experiencia de aprendizaje del estudiante a través de los distintos entornos educativos medidos por tecnologías que la universidad ofrece.

Asimismo pueden nombre los siguientes objetivos particulares:

- Obtener el máximo provecho de las TIC presentes en el ámbito de UNNOBA.
- Identificar e implementar nuevas y mejores formas de diseñar y desarrollar materiales (o recursos educativos), en beneficio de la calidad de aprendizaje de los estudiantes.
- Contribuir al desarrollo y mejoramiento del entorno virtual de aprendizaje adoptado por UNNOBA.
- Reforzar los conocimientos y buenas prácticas adquiridas en el marco de los trabajos de investigación anteriores, contribuyendo a la formación de nuevos investigadores en el ámbito de UNNOBA.
- Difundir los logros alcanzados a la comunidad entera, para su adopción definitiva en toda la comunidad educativa.

### Formación de Recursos Humanos

En esta línea de I/D se ha concluido dos tesinas de grado.

Se espera desarrollar una tesina de grado y una tesis de magister, en el área de Tecnología y Educación, dirigidas por miembros de este proyecto.

En esta misma línea se han obtenido una beca de entrenamiento para alumnos universitarios y una beca de estudio, ambas otorgadas por la Comisión de Investigaciones Científicas (CIC), y una beca de estímulo a las vocaciones científicas otorgada por el Consejo Interuniversitario Nacional (CIN).

### Referencias

- [1] Programa “Conectar Igualdad”: <http://www.conectarigualdad.gob.ar/sobre-el-programa/que-es-conectar/>
- [2] Filippi J.; Lafuente G.; Perez D.; Aguirre S. Ambientes Ubicuos como facilitador del Proceso Educativo. Facultad de Ingeniería. Universidad Nacional de La Pampa. [http://sedici.unlp.edu.ar/bitstream/handle/10915/19310/Documento\\_completo.pdf?sequence=1](http://sedici.unlp.edu.ar/bitstream/handle/10915/19310/Documento_completo.pdf?sequence=1)
- [3] Weiser M. Ubiquitous Computing, IEEE Computer "Hot Topics", October 1993, <http://www.ubiq.com/hypertext/weiser/UbiCompHotTopics.html>

### Bibliografía

- Casamayor, G.; Alós, M.; y otros. La Formación On-line, Una mirada integral sobre el e-learning, b-learning. Primera edición. Editorial Graó. Barcelona, España. 2008.
- Krumm J., Creating a New Book of Tutorials: Ubiquitous Computing Fundamentals, Microsoft Research Redmond, UbiComp 2009.
- Lion, C. Imaginar con tecnologías, Relaciones entre tecnologías y conocimiento. Primera edición. Editorial Stella. Buenos Aires, Argentina. 2006.
- Barberà, E. La incógnita de la Educación a Distancia. Primera edición. Editorial Horsori. Barcelona, España. 2001.



- Carmona M.; González S.; Castro Ruiz, Innovación Tecnológica en Comunicaciones Móviles Desarrollada Con Software Libre: Campus Ubicuo Novática: Revista de la Asociación de Técnicos de Informática, ISSN 0211-2124, N°. 190, 2007.
- Litwin, E.; Maggio, M.; Lipsman, M. Tecnologías en las aulas, Las nuevas tecnologías en las prácticas de la enseñanza. Casos para el análisis. Primera edición. Editorial Amorrortu. Buenos Aires, Argentina. 2005.
- Bravo C.; Redondo M.; Ortega M.; Bravo J. Evolución de un Entorno Colaborativo de Enseñanza Basado en Escritorio hacia la Computación Ubicua, (2002) COLINE'02 Investigación en Entornos de Interacción Colectiva, Workshop de Investigación sobre nuevos paradigmas de interacción en entornos colaborativos aplicados a la gestión y difusión del Patrimonio cultural. Granada, 11 y 12 de Noviembre del 2002.
- Litwin, E. Tecnologías educativas en tiempos de Internet. Primera edición. Editorial Amorrortu. Buenos Aires, Argentina. 2009.
- Levis, D.; Gutierrez Ferrer, M. ¿Hacia la herramienta educativa universal?, Enseñar y aprender en tiempos de Internet. Editorial Ciccus. Buenos Aires, Argentina. 2000.
- Manzanedo, J. El e-Learning en España, Modelos actuales y tendencias de actuación. Primera edición. Colección EOI Tecnología e Innovación. España. 2003.
- Almenara, J.; Graván, P. E-actividades, Un referente básico para la formación en Internet. Primera edición. Editorial MAD. Sevilla, España. 2008.
- Weiser M. The Computer for the Twenty-First Century. Scientific American, Vol. 265, No. 3., pp. 94-104. 1991.

**UNNOBA blogs: Blogs de asignaturas y contenidos abiertos.**

Hugo Ramón, Claudia Russo, Mónica Sarobe, Mariana Saenz, Tamara Ahmad, Paula Lencina, Nicolás Alonso, Leonardo Esnaola, Claudio Gómez Luengo

Instituto de Investigación y Transferencia en Tecnología (ITT), Escuela de Tecnología, Universidad Nacional del Noroeste de la Provincia de Buenos Aires (UNNOBA)

{hugoramon, crusso, monicasarobe, marianasaenz, tamaraahmad}@unnoba.edu.ar,  
{paula.lencina, nicolas.alonso, leonardo.esnaola,  
claudio.gomezluengo}@nexo.unnoba.edu.ar

**Resumen**

Se presenta a continuación una línea de investigación relacionada con la búsqueda de la mejora de propuestas educativas en educación superior a través de la incorporación del uso de la herramienta blog.

La investigación estará basada en el hecho de ofrecer a los docentes de la UNNOBA una herramienta que les permita contar con un espacio institucional desde la cual puedan aumentar sus aulas, y con un acompañamiento constante, es decir, recursos humanos capacitados en el uso de los blogs, en técnicas de comunicación, en el uso de herramientas dentro de una propuesta educativa; así como otras cuestiones que se relacionen con la comunicación como eje central de un proceso educativo.

A partir de esta investigación se sugieren ciertas pautas a seguir a fin de formalizar la solicitud y el uso de la herramienta blog en la UNNOBA, que van desde los pasos a seguir para solicitar la apertura de un blog, las capacitaciones en el uso de los blogs, el intercambio de experiencias docentes, el seguimiento del uso de los blogs, entre otros. Esta propuesta permite ir más allá del dictado de la clase magistral o convencional; y lo hace reformulando el uso de una antigua TIC como lo es el blog.

**Palabras clave:** propuesta educativa, herramienta, blog, TIC, comunicación.

**Contexto**

Esta línea de investigación forma parte del proyecto “*El desafío de las nuevas tecnologías de la información y las comunicaciones en los contextos educativos*” que fue acreditado mediante evaluación externa y financiado por la Secretaría de Investigación, Desarrollo y Transferencia de la UNNOBA en el marco de la convocatoria a Subsidios de Investigación Bianuales (SIB2013) y es una continuación del proyecto “*UNNOBA Virtual. Una plataforma para la integración de sistemas, metodologías y herramientas de enseñanza y aprendizaje*”, acreditado en las convocatorias Subsidios de Investigación Bianuales (SIB2010).

Docentes e investigadores pertenecientes a los departamentos de Informática y Tecnología, Humanidades y Afines y Complementarias y estudiantes de las carreras Licenciatura en Sistemas, Ingeniería en Informática y Diseño Gráfico de la UNNOBA, conforman el equipo que está abocado a este proyecto, el cual se desarrolla en el Instituto de Investigación en Tecnologías y Transferencia (ITT) dependiente de la mencionada Secretaría, y se

trabaja en conjunto con la Escuela de Tecnología de la UNNOBA.

### Introducción

Si se toma al término “educar” en un sentido como el de “guiar o conducir”, colaborando en la “formación e instrucción” de un sujeto; se entiende a la comunicación como uno de los principales ejes en el ejercicio de educar. Entonces la comunicación, en el proceso educativo, debe atenderse si lo que se desea es brindar una buena propuesta educativa.

En el libro *La comunicación: un arte que se aprende* de Lisa Langevin Hogue, se nos plantea: “...tanto si se trata de una relación entre dos individuos como de una interacción dentro de un grupo, antes de descubrir que en el origen de la relación deteriorada exista un problema de comunicación, por desgracia ya se han imputado muchos errores a uno u otro de los interlocutores...”.

Muchos son los canales de comunicación que pueden tomarse para llevar a cabo en este contexto. Si se habla de proceso educativo se entiende que acompaña a una propuesta educativa, y para llevarla a cabo habrá herramientas que asistan al proceso desde la propuesta.

Cabe preguntarse entonces: ¿es el uso de los blogs un canal de comunicación viable para llevar adelante una propuesta educativa?, ¿es un blog una herramienta que complementa la comunicación en pos del éxito de un proceso educativo?

A partir de estos interrogantes y como educadores se debe “pensar la herramienta” como un recurso que ayuda al estudiante a desarrollar el proceso de aprendizaje de forma autónoma. Ya que permite desarrollar en ellos una serie de competencias tales como: análisis de contenidos, facilidad en la

comprensión de discusiones, habilidad para crear y desarrollar su pensamiento y escribir correctamente, capacidad de formar opinión criticando constructivamente el trabajo de compañeros de curso y del docente, habilidad para compartir producciones y trabajo propio, capacidad para comentar y acordar puntos comunes entre diferentes trabajos, habilidad para formar nuevas opiniones usando el trabajo de los compañeros como base y el poder evaluar críticamente el trabajo propio y el de sus pares. Sin embargo, y pensando nuevamente en la herramienta pero esta vez del lado del docente, es este último quien debe proponer el uso de la misma de un modo estratégico destacando las potencialidades del blog, en este caso.

Por otra parte, hace tiempo que la evolución tecnológica se tornó un tanto vertiginosa, por lo que no siempre se llega a conocer la potencialidad de cada una de las herramientas que pueden asistir a un proceso educativo. Entonces será necesario ahondar en el estudio de las potencialidades de las diferentes herramientas.

### Líneas de investigación y desarrollo

En busca de mejorar una propuesta educativa universitaria a través del uso de la herramienta blog en una asignatura dada será necesario:

- Indagar acerca de la importancia de la comunicación en los procesos educativos, siendo la comunicación uno de los ejes centrales en los procesos educativos en cualquier nivel.
- Investigar los diferentes canales de comunicación en una propuesta educativa y comparar la eficiencia de cada uno.

- Investigar sobre el uso de los blogs en otras instituciones educativas, realizando un relevamiento sobre el uso de blogs en diferentes universidades.
- Estudiar de qué manera se puede mejorar la calidad de los procesos educativos, por ejemplo proponiendo una forma más de extender o aumentar el aula través del uso de blogs.
- Pensar de un modo estratégico cómo redefinir una herramienta con historia.
- Investigar si mediante la propuesta del uso de blogs se incentiva la producción de materiales propios de los docentes.
- Impulsar la producción de materiales propios de los docentes.

En la siguiente figura, a modo de ejemplo, se puede graficar el espacio del blog institucional que contendrá a los blogs de las asignaturas:



## Resultados y Objetivos

Emprender todas las tareas antes detalladas permitirá plasmar un espacio de blog institucional, un blog que sirva como eje o base para que luego cada asignatura o proyecto (de investigación, extensión o voluntariado) de la UNNOBA, pueda a partir de allí tener su propio espacio. Además de poder:

- Establecer la relevancia de la comunicación en los procesos educativos.
- Establecer un nuevo y complementario canal de comunicación en una propuesta educativa.
- Aprender de la experiencia de otras instituciones educativas.
- Mejorar la calidad de los procesos educativos.
- Redefinir la herramienta blog en el ámbito de una asignatura o proyecto dentro de la universidad.

Fundamentalmente se persigue el objetivo de formar recursos humanos (RRHH) con las habilidades necesarias para capacitar a docentes que requieran el uso de blogs en sus asignaturas o proyectos. Profesionales con habilidades pedagógicas, preparación didáctica, y la suficiente formación en la utilización de diferentes herramientas. Estos RRHH deben poder planear y brindar capacitaciones en el uso de la herramienta blog. Y, desde luego, estar al tanto de las nuevas corrientes pedagógicas en pos del dictado de nuevas capacitaciones al cuerpo docente.

Además de este acompañamiento pedagógico se persigue el objetivo de generar RRHH que cumplan con ciertas funciones administrativas tales como: recibir y gestionar las solicitudes de blogs, gestionar los expedientes de cada blog, instruir al solicitante en el proceso de solicitud de blog y comunicar al solicitante las novedades, entre otras.



### Formación de Recursos Humanos

En esta línea de I/D se ha concluido una tesina de grado y hay dos becarios alumnos del ITT trabajando en el proyecto.

En el contexto de esta investigación se pretende concluir el trabajo de grado de dos integrantes del equipo del ITT de la UNNOBA dirigido por miembros de este proyecto.

### Bibliografía

- García Aretio, L; Ruiz Corbellá, M; Domínguez Figaredo, D. De la educación a distancia a la educación virtual. Primera edición. Editorial Ariel. 2007.
- García Aretio, L; Ruiz Corbellá, M; García Blanco, M. Claves para la Educación: Actores, agentes y escenarios en la sociedad actual. Narcea SA de ediciones y Universidad Nacional de Educación a Distancia. 2009.
- Barberá, E; Romiszowski, A; Sangrá, A; Simonson, M. Educación abierta y a distancia. Primera edición. Editorial UOC. 2006.
- Barberà, E. La incógnita de la Educación a Distancia. Primera edición. Editorial Horsori. Barcelona, España. 2001.
- Levis, D.; Gutierrez Ferrer, M. ¿Hacia la herramienta educativa universal?, Enseñar y aprender en tiempos de Internet. Editorial Ciccus. Buenos Aires, Argentina. 2000.
- Litwin, E. Tecnologías educativas en tiempos de Internet. Primera edición. Editorial Amorrortu. Buenos Aires, Argentina. 2009.
- Litwin, E.; Maggio, M.; Lipsman, M. Tecnologías en las aulas, Las nuevas tecnologías en las prácticas de la enseñanza. Casos para el análisis. Primera edición. Editorial Amorrortu. Buenos Aires, Argentina. 2005.
- Almenara, J.; Graván P. E-actividades. Un referente básico para la formación en Internet. Primera edición. Editorial Eduforma. España, 2006.
- De la Torre Solarte, G; Narvéez Paredes, E; y otros. Pensamiento Universitario. Propuesta Educativa. Segunda edición. Educ. Editorial Universidad Cooperativa de Colombia. 2006.
- Pardo Kuklinski, H; Cobo Romani, C. Planeta web 2.0: inteligencia colectiva o medios fast food.
- Silva Salinas, S. Medios didácticos multimedia para el aula: Las nuevas tecnologías como Herramientas Didácticas en los Centros Educativos. Primera edición. Ideaspropias editorial. 2005.
- Prieto Castillo, D. La comunicación en la educación. Buenos Aires: La Crujía, capítulos 6, 8 y 9. 1991.
- Langevin Hogue, L. La comunicación: un arte que se aprende. Sal Terrae. España. 2000.

## Entornos Virtuales 3D como herramienta para promover el trabajo colaborativo en educación.

Hugo Ramón, Claudia Russo, Mónica Sarobe, Nicolás Alonso, Leonardo Esnaola, Maximiliano Fochi, Rodrigo Rincón, Eliana Serrano

Instituto de Investigación y Transferencia en Tecnología (ITT), Escuela de Tecnología, Universidad Nacional del Noroeste de la Provincia de Buenos Aires (UNNOBA)

{hugoramon, crusso, monicasarobe}@unnoba.edu.ar, {nicolas.alonso, leonardo.esnaola, maximiliano.fochi, rodrigo.rincon, eliana.serrano}@nexo.unnoba.edu.ar

### Resumen

La incorporación cada vez más acelerada de las Tecnologías de la Información y la Comunicación (TIC) a la educación, está generando una serie de cambios y transformaciones en las formas en que nos representamos y llevamos a cabo los procesos de enseñanza y aprendizaje. Estos cambios pueden observarse en los entornos tradicionales de educación formal, pero también en la aparición de nuevos entornos educativos basados total o parcialmente en las TIC.

La línea de investigación presentada se centra en el diseño, creación y coordinación de un entorno virtual 3D de enseñanza y aprendizaje. Para llevar a cabo dicho propósito se plantea la definición e implantación de un modelo de entorno virtual 3D haciendo enfoque, no solo en el plano técnico y pedagógico, sino que también en una metodología de transición que permita mejorar la calidad de los métodos y técnicas de enseñanza.

**Palabras clave:** Entorno Virtual 3D, PACIE, SBS, TIC.

### Contexto

Esta línea de investigación forma parte del proyecto “*El desafío de las nuevas tecnologías de la información y las comunicaciones en los contextos educativos*” que fue acreditado mediante evaluación externa y financiado por la Secretaría de Investigación, Desarrollo y Transferencia de la UNNOBA en el marco de la convocatoria a Subsidios de Investigación Bianuales (SIB2013) y es una continuación del proyecto “*UNNOBA Virtual. Una plataforma para la integración de sistemas, metodologías y herramientas de enseñanza y aprendizaje*”, acreditado en las convocatorias Subsidios de Investigación Bianuales (SIB2010). A su vez se conforma de proyectos específicos aprobados por la Comisión de Investigaciones Científicas (CIC) y el Consejo Interuniversitario Nacional (CIN) en el marco de las convocatorias a las becas de entrenamiento para alumnos universitarios y becas de estímulo a las vocaciones científicas del año 2012.

El proyecto se desarrolla en el Instituto de Investigación en Tecnologías y Transferencia (ITT) dependiente de la mencionada Secretaría, y se trabaja en conjunto con la Escuela de Tecnología de la UNNOBA.

El equipo está constituido por docentes e investigadores pertenecientes a los departamentos de Informática y Tecnología, Humanidades y Afines y Complementarias y estudiantes de las carreras de Licenciatura en Sistemas, Ingeniería en Informática y Diseño Gráfico de la UNNOBA.

### Introducción

Desde hace algunos años se está observando que la educación ha sufrido transformaciones debido al avance e introducción de nuevas tecnologías en esta área, especialmente desde la introducción de los “Sistemas de Gestión del Aprendizaje”, comúnmente conocidos como LMS (- por sus siglas en inglés de *Learning Management System* -). Éstos favorecieron en gran parte al crecimiento de la educación a distancia y colaboraron con la mejora en los procesos de aprendizaje. Pero el problema que se viene observando, es que los proveedores de los LMS, están compitiendo cada vez más por intentar posicionar su herramienta como elemento central en el e-learning [1], dejando de lado lo que debería ser el objetivo principal, permitir que el estudiante asuma más responsabilidad en el proceso de su aprendizaje.

El enfoque tradicional de los LMS hace que los docentes diseñen los cursos y actividades con una visión centrada en el estudiante, pero esto no permite a los estudiantes establecer sus propias metas y plazos, organizar su trabajo, evaluar el uso del tiempo y trabajar en colaboración con sus compañeros [2].

Mucho tiene que ver en el impacto de estas nuevas tecnologías o modalidades de enseñanza, el advenimiento de una nueva generación; una generación que nació con la tecnología a su alcance, lo que se conoce como los nativos digitales [3], quienes tienen más predilección por el aprendizaje basado en

la experiencia y el uso de las herramientas provistas por las TIC. Pero esto genera también un reto a los docentes tradicionales, quienes deben generar o adaptarse a una nueva metodología de enseñanza que permita, mediante el uso de tecnologías, explotar las capacidades de los estudiantes [4].

De lo antes expuesto consideramos que para poder plantear el diseño, creación y coordinación de un entorno virtual 3D de enseñanza y aprendizaje (EV3D) a utilizarse en el ámbito de la UNNOBA, no debemos enfocarnos solamente en un plano técnico y pedagógico, sino que además debemos pensar en una metodología de transición que permita mejorar la calidad de los métodos y técnicas de enseñanza. Es así que basaremos parte de la propuesta del EV3D en la metodología PACIE (Presencia, Alcance, Capacitación, Interacción, E-learning), desarrollada por el ingeniero Pedro Camacho [5]. PACIE es una metodología que permite el uso de las TIC como un soporte a los procesos de aprendizaje y autoaprendizaje, dando realce al esquema pedagógico de la educación real.

Centraremos la propuesta del EV3D en tres aspectos fundamentales, dejando abiertas ciertas incógnitas que podrán ser subsanadas luego de la realización de distintas experiencias realizadas con el EV3D una vez puesto en marcha [6]:

*Primer aspecto:* Imagen del EV3D. Se pretende lograr un entorno amigable y atractivo y para ello se debe tener en cuenta las siguientes consideraciones:

- Utilizar una misma tipografía y tamaño de texto para títulos.
- Utilizar una misma fuente tipografía para la información.

- Diferenciar la información relevante del resto, utilizando una tipografía y color diferente.
- Las imágenes deben mantener un mismo tamaño.
- Utilizar enlaces con imágenes que faciliten el acceso a los distintos elementos de la universidad.
- Incorporar recursos atractivos de la web 2.0 como animaciones, video y otros.
- Crear la necesidad de descubrir formas nuevas, llamativas, en el EV3D que motive al estudiante a continuar investigando y utilizando el mismo.
- Tener y mantener una identidad. Es decir, hacer sentir al estudiante que cada aula virtual es parte de un mismo centro de educación virtual [7].

*Segundo aspecto:* Manejo y organización de la información. En base al primer aspecto ya se tiene una visión de cómo presentar el EV3D. En este aspecto se da paso a la importancia del manejo y la organización de la información dentro de nuestra aula virtual.

Para conseguir un buen aprendizaje por parte del estudiante se debe tener en claro que se busca. Dentro del modelo que se propondrá, el cual basará parte de sus aspectos en la metodología PACIE, se tomó la decisión de utilizar los SBS. Mediante los SBS, del inglés *Standards, Benchmarks and Skills* (Estándares, Patrones y Destrezas); determinamos qué destrezas (*Skills*) deben generar, desarrollar o mejorar nuestros estudiantes, para cuando salgan de los procesos educativos hacia un punto productivo al que se propenda, hayan cumplido con los requerimientos deseados y puedan afrontar

los nuevos esquemas. Todas estas destrezas se agrupan por patrones o marcas académicas (*Benchmarks*) que guarden características afines, todas estas personalizadas para nuestra cátedra de forma exclusiva. Luego reunimos los patrones académicos y apuntamos hacia estándares (*Standards*) académicos que deben ser fijados por el área académica a la que nuestra cátedra pertenezca [8] [9].

*Tercer aspecto:* Organización, gestión y administración de la educación virtual. Según lo propuesto en la metodología PACIE, se considera que para poder llevar a cabo nuestra propuesta del EV3D, es de gran importancia poder contar con un área especializada en educación a distancia que deba orientarse a gestionar todos lo relacionado con la educación virtual, y ser la encargada de organizar, gestionar y administrar la misma, obteniendo poder de decisión sobre el EV3D.

Según lo indicado en la metodología PACIE [10], la composición mínima recomendada de recursos humanos que deben estar presentes en esta área se compone de un experto en pedagogía, un experto en informática y un experto en comunicaciones sociales.

De esta manera podemos hacer una mejor división de tareas, donde:

- El comunicador debe manejar la comunicación de la información tanto de los procesos de educación virtual, como del EV3D en general y de las aulas virtuales en particular. Debe ser el nexo entre el pedagogo y el informático.
- El informático debe apoyar en cuestiones que respectan a la tecnología y a la plataforma brindando soluciones para que el EV3D funcione óptimamente.



- El pedagogo debe ser el encargado de implementar metodologías de aprendizaje acordes al EV3D, y de esta manera obtener mejores técnicas para aprender dentro del mismo, y una forma de enseñar pensada para el EV3D con el que se está trabajando.

### Líneas de investigación y desarrollo

Se identificaron una serie de aspectos que en el modelo a proponer se busca abordar con soluciones concretas y que sientan las bases de lo que consideramos debería estar presente en el EV3D:

- *Aspecto administrativo:* un EV3D es un sistema que integra diferentes soluciones, pero su principal objetivo es proveer un adecuado sustento a las actividades de enseñanza aprendizaje de quienes serán sus principales actores, estudiantes y docentes por igual. Estos necesitan de un fuerte acompañamiento que debe ser brindado por recursos entrenados con el apoyo de procesos bien definidos.
- *Aspecto organizacional de la información:* debemos hacer hincapié en el manejo y la organización de la información, analizar que se hace con la misma y como utilizarla para generar el aprendizaje del estudiante. Podemos lograr un buen manejo y organización de la información utilizando estándares y o metodologías vigentes pudiendo ser una opción el uso de los SBS.
- *Aspecto pedagógico:* es necesario que el docente rompa con el esquema tradicional de enseñanza, debe acercarse al estudiante hacia el autoaprendizaje, que el entorno sea una puerta de contenidos para que se interese, continúe investigando y aprenda por sí mismo. Esto se lleva a cabo desarrollando contenido apropiado para el entorno y de esta

manera guiar a sus estudiantes a través de él [11].

- *Aspecto técnico:* todo buen EV3D debe ser acompañado de personas capacitadas para su uso, que actúen de guía de aquellas personas que no posean los conocimientos necesarios específicos para su utilización.
- *Aspecto evolutivo:* todo sistema virtual debe considerar una evolución previsible si pretende estar vigente con una realidad tan cambiante como la actual. Así, es necesario realizar un proceso de control y de permanente actualización que busque adaptar el sistema a nuevas herramientas y entrar en un círculo virtuoso que lo optimice permanentemente.
- *Aspecto gráfico:* Se debe comprender que, en los tiempos actuales, el diseño y aspecto visual del EV3D debe ser atractivo para los estudiantes, dado a que en caso contrario se asume el riesgo de que no ingresen frecuentemente al EV3D y por ende no podamos aprovechar sus ventajas.
- *Aspecto funcional:* tareas indispensables como la gestión de las aulas virtuales: su creación, el proceso de matriculación de alumnos, la asignación de docentes, el manejo de los roles, etc., deberían realizarse de una manera estandarizada y por personal específico. Por un lado, porque dichas tareas no deberían ser responsabilidad de un docente, el docente es un usuario y no un gestor en el entorno virtual; y por otro lado, es necesario que haya una homogeneidad en la realización de esas actividades, ya que un entorno heterogéneo da la sensación de desorden.

## Resultados y Objetivos

En función a lo expresado, el principal objetivo es lograr la correcta implantación de un EV3D, en el ámbito de la UNNOBA, basándonos en la metodología investigada. Para ello se realizó un exhaustivo análisis de la metodología PACIE. Se investigó sobre el uso de los EV3D aplicados a educación y se relevaron y analizaron diversas soluciones de software, tales como *SecondLife*, *OpenSim*, *OpenWonderland*, para finalmente decidir cuál será la que mejor se adecúe a las necesidades actuales presentes en el ámbito de la UNNOBA.

Actualmente se está trabajando en la integración del EV3D propuesto con el entorno virtual de enseñanza y aprendizaje presente en la UNNOBA, denominado UNNOBA Virtual [12], para poder llevar adelante una experiencia del uso del EV3D en el marco de una actividad colaborativa determinada a desarrollarse en el curso de ingreso con modalidad semi-presencial que dicta la UNNOBA.

## Formación de Recursos Humanos

En esta línea de I/D se ha concluido una tesina de grado y hay dos trabajos de grado que se están desarrollando y se vinculan con el proyecto.

Se espera desarrollar una tesina de grado y una tesis de magister, en el área de Tecnología y Educación, dirigidos por miembros de este proyecto.

En esta misma línea se han obtenido una beca de entrenamiento para alumnos universitarios de la Comisión de Investigaciones Científicas (CIC) y una beca de estímulo a las vocaciones científicas otorgada por el Consejo Interuniversitario Nacional (CIN).

## Referencias

[1] Alan Palme. Enhancing learning and teaching through the use of technology: a

revised approach to HEFCE's strategy for e-learning. HEFCE. United Kingdom, 2009.

[2] Maestría Diseño Instruccional por José Luis Córca  
[http://cvonline.uaeh.edu.mx/Cursos/Maestria/MGIEMV/DisenoProgramasEV12/materiales/Unidad%204/Cap4\\_Disenoinstruccional\\_U4\\_MGIEV001.pdf](http://cvonline.uaeh.edu.mx/Cursos/Maestria/MGIEMV/DisenoProgramasEV12/materiales/Unidad%204/Cap4_Disenoinstruccional_U4_MGIEV001.pdf)

[3] Prensky, M. (2001). Digital natives, digital immigrants. *On the Horizon*, 9 (5), 1-6.

[4] Revista La educación y la Virtualidad Editorial: Grupo Dseta Editora: Francia Tovar Romero  
<http://www.youblisher.com/p/173589-Please-Add-a-Title-La-Educacion-y-la-Virtualidad/>

[5] Metodología PACIE, ensayo publicado por el Ing. Luis Oñate para la Fundación para la Actualización Tecnológica de Latinoamérica (FATLA), 2009,  
[www.iuetaebvirtual.wikispaces.com/file/view/22234756-La-Metodologia-Pacie.pdf](http://www.iuetaebvirtual.wikispaces.com/file/view/22234756-La-Metodologia-Pacie.pdf)

[6] Ensayo "Plataformas de educación a distancia", Ing. Rambo Alice,  
[http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/Educacion\\_Distancia\\_Alice\\_2009.pdf](http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/Educacion_Distancia_Alice_2009.pdf)

[7] Metodología PACIE aplicada a entornos virtuales, <http://pacie-en-muves.wikispaces.com/home>

[8] Exposición profesional "Aplicación PACIE en los Estándares Académicos de la Educación Virtual" por Deizi Carolina De Jesús Lobo  
<http://pacieeducavirtual.jimdo.com/aplicaci%C3%B3n-de-pacie-en-los-est%C3%A1ndares-acad%C3%A9micos-de->

[la-educaci%C3%B3n-virtual-aula-virtual-moodle/](#)

[9] Experiencia educativa con entornos 3D, UPEL (Universidad Pedagógica Experimental Libertador),  
[www.ugr.es/~sevimeco/revistaeticanet/numero10/Articulos/Formato/articulo5.pdf](http://www.ugr.es/~sevimeco/revistaeticanet/numero10/Articulos/Formato/articulo5.pdf)

[10] Implementación en módulos según metodología PACIE, 2011,  
[www.moodle.org.uy/moodlemoot\\_2011/moodlemoot/moodlemootuy2011\\_submission\\_25.pdf](http://www.moodle.org.uy/moodlemoot_2011/moodlemoot/moodlemootuy2011_submission_25.pdf)

[11] Ensayo “Entornos Virtuales 3D, Alternativa Pedagógica para el Fomento del Aprendizaje Colaborativo y Gestión del Conocimiento”,  
[http://www.scielo.cl/scielo.php?pid=s0718-50062011000200006&script=sci\\_arttext](http://www.scielo.cl/scielo.php?pid=s0718-50062011000200006&script=sci_arttext)

[12] Entorno Virtual de Enseñanza y Aprendizaje (EVEA) de la UNNOBA,  
<http://virtual.unnoba.edu.ar>

## Aplicaciones de las TIC en Educación y Ciencias

**Adair Martins, Carina Fracchia, Claudia Allan, Susana Parra,  
Renzo García, Estefania Zurbrigk, Natalia Baeza, Lorena Robles, Daniel Benilla**

Departamento Computación Aplicada/ Facultad de Informática  
Universidad Nacional del Comahue

**Dirección:** Buenos Aires 1400, 8300 – Neuquén

**Teléfono:** 0299 - 4490300 int. 429

**e-mails:** mgavbm@yahoo.com.ar,

{allanclau, sbparra, rgarcia.inf, baeza.natalia}@gmail.com,

{cfracchi@hotmail.com, zureste, dopanga, dymbe2}@hotmail.com

### Resumen

En la actualidad las Tecnologías de la Información y Comunicación (TIC) están presentes en el día a día en todas las áreas del conocimiento y principalmente con significativas aplicaciones en las áreas de educación y de las ciencias en general. Ya no es posible estar ajeno a su crecimiento y a los desafíos proporcionados por todos sus recursos. En este trabajo se presentan los avances y resultados de la investigación que se está realizando en el estudio y desarrollo de nuevas herramientas tecnológicas.

Se ha avanzado en el diseño de herramientas computacionales basadas en software libre para el diseño de algoritmos en la enseñanza de la programación en materias de primer año. También se ha avanzado en la implementación de una herramienta para el tratamiento de datos del viento para la simulación de generación de energía eólica, y una herramienta para la traducción del Lengua Española a la Lengua de Señas Argentina (LSA) destinada a personas con discapacidad auditiva.

El objetivo general de estas herramientas es su integración en los procesos de enseñanza y aprendizaje colaborativos.

**Palabras clave:** TIC, Herramientas Computacionales, Software libre, Entorno Colaborativo, Algoritmos, Accesibilidad, Energía Eólica

### Contexto

Las líneas de investigación en desarrollo se encuentran enmarcadas en el Proyecto: “Simulación y Métodos Numéricos en Ciencias de la Computación” de la Facultad de Informática, Universidad Nacional del Comahue (UNCo). El proyecto está integrado por un grupo interdisciplinario, teniendo como participantes docentes, estudiantes de la carrera de Licenciatura en Ciencias de la Computación de la Facultad de: Informática (FaIF) e investigadores de las facultades de: Ingeniería (FI), Economía y Administración (FaEA), Ciencias Agrarias (FaCA) y de las Universidades Brasileñas de Itajubá (UNIFEI), del Estado de São Paulo (UNESP) y Universidade Católica de Brasilia (UCB).

### Introducción

En la temática del proyecto de investigación se está trabajando fundamentalmente con el Uso de TIC, Accesibilidad Web, Aplicación de Técnicas de Simulación. Las TIC y todos sus recursos permiten complementar la educación presencial y en este sentido se ha avanzado en el diseño e implementación de distintas herramientas basadas en software libre para la evaluación de los aprendizajes de diferentes modalidades de educación.

La concientización sobre la utilización de recursos energéticos basados en fuentes renovables con el propósito de satisfacer la creciente demanda de energía viene siendo



incrementada cada vez más. En función de esto, una muy buena alternativa técnica y económica es la utilización de la energía eólica, energía obtenida de la fuerza del viento, para generar energía eléctrica sin producir contaminantes ambientales, promoviendo de esta forma un futuro energético limpio. El estudio de las características de los vientos es una etapa imprescindible para la evaluación de la eficiencia energética y la evaluación económica para la realización del proyecto de un parque eólico. Es posible utilizar distintos modelos estocásticos para estimar y estudiar las características del viento de una determinada localidad. En general la distribución Weibull es la más utilizada para caracterizar de forma eficiente las clases de velocidad del viento. Teniendo en cuenta la relevancia del tema, se desarrolló una herramienta computacional basada en software libre para ser utilizada en el tratamiento de datos de viento. Con la herramienta se pueden analizar los datos generados por equipos de medición de viento, particularmente datos de velocidad, dirección y altura, mediante modelos numéricos y estadísticos [1-2].

En los últimos años muchos países han avanzado en forma diferente en el tratamiento de la Accesibilidad Web. Es cada vez más importante brindar igualdad de posibilidades en relación al acceso a la información a personas con diferentes discapacidades. En Argentina el concepto de accesibilidad ha cobrado importancia a partir de la legislación y reglamentación relacionada. La ley N° 26653 de “Accesibilidad a la Información en Páginas Web”, sancionada el 3 de Noviembre de 2010, fue publicada en el boletín oficial el 30 de Noviembre del mismo año [3]. Especifica las normas y requisitos que deben cumplir los sitios Web, y las técnicas de diseño y programación para su elaboración con el objeto de facilitar el acceso a sus contenidos a todas las personas independientemente de las discapacidades que puedan tener. Esta ley trata de garantizar la igualdad de oportunidades para el acceso a los contenidos en la Web, evitando cualquier tipo

de discriminación. El tema de la accesibilidad ha tomado mayor reconocimiento a partir de la sanción de la ley y está incentivando que se comiencen a tomar medidas y a realizar acciones para mejorar la accesibilidad de las tecnologías usadas. En base a lo mencionado anteriormente, se propusieron modificaciones y la implementación de nuevas herramientas basadas en software libre para mejorar la accesibilidad Web en la Universidad Nacional del Comahue (UNCo) y reducir la brecha hacia la inclusión de personas con capacidades diferentes [4-5].

### **Líneas de Investigación y Desarrollo**

Se presentan en forma sintética las líneas de investigación y desarrollo:

- Uso de TIC
- Diseño e implementación de una herramienta computacional para el trabajo colaborativo de edición de algoritmos
- Accesibilidad Web.
- Diseño e implementación de una herramienta computacional para el tratamiento de datos de viento en el estudio y simulación de la generación de energía eólica.
- Desarrollo e implementación de herramientas automatizadas para la evaluación de contenidos y del aprendizaje en las modalidades semipresencial y a distancia.

### **Resultados y Objetivos**

Se pueden mencionar los siguientes resultados alcanzados en las líneas de investigación que dieron origen a publicaciones en congresos internacionales, nacionales y dictados de distintos cursos de capacitación.

Existen actualmente distintas herramientas tecnológicas que permiten el diseño de algoritmos utilizando metodologías tales como pseudocódigo y diagramas de flujo. En

las materias introductorias de programación es fundamental lograr que los alumnos puedan adquirir habilidades para resolver problemas mediante el diseño de algoritmos y su posterior implementación en un lenguaje de programación determinado. La programación visual mediante la utilización de herramientas automatizadas para el diseño de algoritmos ha permitido observar importantes ventajas en la enseñanza y aprendizaje en materias iniciales de programación. Permite que el alumno realice sus diseños eligiendo alguna metodología y a continuación pueda automáticamente compilarlo y ejecutarlo, obviando la etapa de la codificación. La ventaja fundamental es que en esta etapa inicial no existe la necesidad de recordar la sintaxis de los lenguajes de programación utilizados en la programación convencional. El tiempo ganado se invertirá en el aprendizaje de la lógica y la verificación del correcto funcionamiento del mismo.

Para la programación se usan entornos como Eclipse y NetBeans, pero los mismos carecen de funcionalidades que permitan trabajar las instancias previas correspondientes al diseño de los algoritmos. Se está realizando el diseño e implementación de una herramienta basada en software libre que contempla la edición individual o colaborativa de algoritmos y al mismo tiempo su codificación en un lenguaje seleccionado como C++ o Java. El objetivo de esta herramienta es permitir la ejecución paso a paso y la confección de la traza resultante. Será utilizada e integrada a los procesos de enseñanza y aprendizaje en los cursos iniciales de programación. Se pretende con el uso de esta herramienta mejorar el rendimiento del desempeño académico de los alumnos de las primeras materias de las carreras de Computación de la Facultad de Informática de la Universidad Nacional del Comahue disminuyendo la deserción en el primer año de la universidad [6-10].

Los resultados de esta investigación fueron presentados en el XVIII Congreso Argentino de Ciencias de la Computación (Cacic 2012) y la exposición de la integrante del proyecto resultó distinguida y premiada como mejor

exposición del día del Workshop: Tecnología Informática Aplicada en Educación (WTIAE) [10].

El uso masivo de las TIC y el diseño no accesible de la mayoría de los sitios Web, hizo que estos se convirtieran en un factor de exclusión en el ámbito educativo de los usuarios con algún tipo de discapacidad. Como respuesta a esta problemática, se han propuesto modificaciones e implementaciones de nuevas herramientas computacionales como alternativa para permitir la adquisición de competencias por parte de los usuarios con necesidades diferentes. Del relevamiento de referencias bibliográficas sobre el estudio de los lenguajes de señas en distintos idiomas y software relacionados [11-15] se observó la importancia y la necesidad de nuevos desarrollos de herramientas tecnológicas para potenciar la inclusión de las personas con discapacidad auditiva en el ámbito educativo.

Se está desarrollando para personas con discapacidad auditiva una herramienta de software libre que acopla las modernas tecnologías de la computación gráfica para realizar la traducción de la Lengua Española a la Lengua de Señas Argentina con el propósito de beneficiar la no exclusión en el proceso de enseñanza y aprendizaje. Se está utilizando el entorno NetBeans y el lenguaje de programación Java, lo que permite la portabilidad de la aplicación, haciendo posible utilizar la herramienta sobre diferentes sistemas operativos sin inconvenientes. Se ha desarrollado un módulo, que traduce texto a lengua de señas.

Los resultados de la investigación fueron presentados en los siguientes congresos: XVII Congreso Internacional de Informática Educativa (TISE) [15] y en las 3ras Jornadas de Educación Mediada por Tecnología [16]. Esta investigación es el tema de la tesis de grado de la alumna de la Licenciatura en Ciencias de la Computación y ya tiene su plan de tesis aprobado por el Comité Académico de Tesis

En relación a la aplicación de las TIC a la simulación de la generación de energía eólica en la Patagonia, se ha desarrollado una

herramienta computacional para el tratamiento de datos viento. Con esta herramienta se permite analizar los datos generados por equipos de medición de viento, particularmente datos de velocidad, dirección y altura, mediante modelos numéricos y estadísticos. Permitirá clasificarlos y mostrarlos en forma gráfica a través de histogramas de velocidad, rosa de viento, etc. En esencia, el uso de esta herramienta ayudará predecir escenarios de producción de energía eólica a partir de los análisis estadísticos de los datos de velocidad del viento.

Esta herramienta será integrada a los procesos de enseñanza y aprendizaje colaborativos. Esta investigación es el tema de la tesis de grado de un estudiante de la Licenciatura en Ciencias de la Computación que ya tiene su plan aprobado por el Comité Académico de Tesis.

Se han realizado varias experiencias utilizando la plataforma Plataforma de Educación a Distancia de la Universidad Nacional del Comahue (PEDCO) y distintos recursos TIC. Integrantes del proyecto han participado como tutores en cursos destinados a docentes del nivel medio y nivel superior dictados en la modalidad presencial y a distancia. El objetivo fue promover la articulación entre los niveles medio y universitario para:

- El acceso equitativo a herramientas de aprendizaje, recursos y tecnologías de calidad.
- Apoyar comunidades de aprendizaje que faciliten a los educadores colaborar, compartir mejores prácticas e integrar las habilidades dentro de las aulas.
- Ofrecer instancias para el aprendizaje individual y colaborativo.

Los resultados fueron presentados en los congresos: Cuarto Congreso Virtual Iberoamericano de Calidad en Educación a Distancia (EduQ@2012) [17], I Congreso Internacional de Educación a Distancia (La Educación a Distancia en América Latina, Desafíos, Alcances y Proyecciones) [18],

También se presentó una propuesta para una experiencia basada en el aprendizaje en forma colaborativa, en el abordaje por proyecto y en la construcción colectiva en la materia Programación del Profesorado de Matemática. Dio origen al trabajo [19].

Siguiendo con la línea Uso de TIC se ha realizado el relevamiento de nuevos recursos tecnológicos y diseño de nuevas herramientas computacionales para la evaluación de aprendizajes en modalidad presencial y a distancia. Se puso énfasis a las principales estructuras de control para la enseñanza de la programación estructurada en las materias de computación de primer año. Esta investigación contribuirá a potenciar las modalidades de educación en la Facultad de Informática. Esta investigación forma parte de la de tesis de la Maestría en Enseñanza de las Ciencias Exactas y Naturales de una docente integrante del proyecto. Se ha finalizado y se está en espera de la programación de la fecha de defensa.

## Formación de Recursos Humanos

El objetivo más importante del Proyecto de Investigación es la formación de recursos humanos. El Proyecto cuenta con 13 integrantes, entre los cuales se encuentran docentes, asesores del país y del exterior y estudiantes de la carrera de Licenciatura en Ciencias de la Computación que se encuentran desarrollando sus tesis de grado.

Se puede destacar que el proyecto contó con dos becarios, ambos estudiantes de la carrera de Licenciatura en Ciencias de la Computación. Uno obtuvo la beca de Estímulo a las Vocaciones Científicas del Consejo Interuniversitario Nacional (CIN), fue renovada en 2012 con el informe anual aprobado. El otro becario obtuvo la beca TICs de la Agencia Nacional de Promoción Científica y Tecnológica (FONCyT), en el marco del “Programa para promover la innovación productiva a través del fortalecimiento y consolidación de capital humano aplicado a la industria de la Tecnologías de la Información y las Comunicaciones (TICs), para la finalización de estudios de grado” en año 2013. Una

integrante del grupo ha presentado su tesis de la maestría en Enseñanza de las Ciencias Exactas y Naturales, se encuentra en la espera de la programación de la fecha de la defensa. Otra integrante finalizó en 2012 los cursos de la misma maestría. Se concluyó una tesis de grado y tres se encuentran en desarrollo, previendo sus defensas para fines del 2013.

## Referencias

- [1] Gary, L. J., Wind Energy Systems, Manhattan, KS, (2001).
- [2] Ting D., Bin, S., Sheng-Wen, S., Study on the wind energy resources assessment in Wind Power Generation, Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC), 10.1109/AIMSEC.2011.6009905, IEEE, (2011).
- [3] Acceso a la Información Pública, Ley 26.653, <http://infoleg.gov.ar/infolegInternet/anexos/175000-179999/175694/norma.htm>.
- [4] Zurbrigk E., Alonso de Armiño A., Martins A., Accesibilidad Web y Educación para Todos, III Jornadas de educación mediada por tecnología. Centro Universitario Zona Atlántica UNCo, (2012).
- [5] Martins, A., Fracchia, C., Allan, C., Parra, S., García, R., Benilla, D., Robles, L., Zurbrigk, E., Baeza, N. Avances en Simulación y TIC, XIV WICC 2012. Posadas. Misiones. ISBN 978-950-766-082-5, (2012).
- [6] Watts, T., The SFC editor a graphical tool for algorithm development, JSSC, USA, (2004)
- [7] Garner, S., Learning Resources and Tools to Aid Novices Learn Programming, Informing Science InSITE, pp. 214-222, (2003)
- [8] Garner, S., A Program Design Tool Help Novices Learn Programming, Proceeding Ascilite, Singapore, (2007)
- [9] Manso, A., Marques, C. G., Dias, P., Portugol IDE v3.x: A New Environment to Teach and Learn Computer Programming, EDUCON, IEEE, pp. 1007-1010, (2010)
- [10] Fracchia, C., Baeza, n., Martins, A. ECDIA: Entorno Colaborativo para el Diseño e Implementación de Algoritmos. X Workshop Tecnología Informática aplicada en Educación. Cacic 2012. UNS, Bahía Blanca, (2012).
- [11] Guimaraes, C.; Antunes, D.R.; Trindade, D.F.G.; Silva, R.A.L.; Garcia, L.S.; Fernandes, S., Evaluation of a computational description model of Libras (Brazilian Sign Language): Bridging the gap towards information access Research Challenges in Information Science (RCIS), Fifth International Conference on Computer Science, pp. 1-10, (2011)
- [12] V.C M., Dolores, Rojano Caceres, J. R., García Gaona, R., Design of Educational Software to Integrate Kids with Hypoacusia in Preschool, Computer Science (ENC), 2009 Mexican International Conference, pp. 294 – 297, (2009).
- [13] Lengua de Señas, <http://www.lenguajedelsilencio.es.tl/LENGUA-DE-SE%D1AS.htm>
- [14] NetBeans, <http://netbeans.org>
- [15] Zurbrigk E., Alonso de Armiño A., Martins A., Finger: Herramienta Educativa para Personas con Discapacidad Auditiva, XVII Congreso Internacional de Informática Educativa – pp. 49-52, TISE 2012, Santiago- Chile, (2012).
- [16] Zurbrigk E., Alonso de Armiño A., Martins A., Accesibilidad Web y Educación para Todos, III Jornadas de educación mediada por tecnología. Centro Universitario Zona Atlántica, UNCo, (2012).
- [17] Pérez, T., Plaza, J., Fracchia, C. Acuña, A. Pérez, F., Las TIC: Una ventana hacia nuevas relaciones sociales y de aprendizaje colectivo. IV Congreso Virtual Iberoamericano de Calidad en Educación a Distancia, EduQ@2012, Mendoza, (2012).
- [18] Fracchia, C., Perez, F., Plaza J., Acuña, A., Perez, T. (2012). Consolidación de Redes de Conocimiento: Experiencia Comahue. I Congreso Internacional de Educación a Distancia: Desafíos, Alcances y Proyecciones. Universidad del Salvador. Buenos Aires.
- [19] Parra S., Allan C., Martins, A., Una Experiencia Educativa basada en Proyectos de Trabajo Colaborativo y Construcción Colectiva, III Jornadas de educación mediada por tecnología. Centro Universitario Zona Atlántica, UNCo, (2012).



# BIGDATA EN LA EDUCACIÓN

**Autores:** Doc. Ing. Rodolfo Bollatti (Director Investigación)

Universidad Abierta Interamericana – Buenos Aires  
Tecnología Informática  
Secretaría de Investigación – CAETI

**E-mail:** Rodolfo.Bollatti@UAI.edu.ar

**Telefono:** 011-1536584700

## Resumen

BigData se está convirtiendo en todos los entornos informatizados en el gran tema del momento, también lo va a ser en la educación. Con “BigData” nos referimos al hecho de que la producción y almacenamiento de información on-line está creciendo de forma tan desmesurada que las soluciones que en general han gestionado esta información a nivel corporativo, ya no son suficientes. Así como los profesionales y los modelos que tienen que analizar estos datos y convertirlos en elementos que resulten útiles para la toma de decisiones

El objetivo general de este proyecto es revelar la tendencia actual y futura sobre modelos basados en bigdata que ayuden a las instituciones educativas a identificar estudiantes en riesgo, de forma de intervenir con el fin de reducir la deserción y aumentar las tasas de graduación de los alumnos.

**Palabras clave:** BigData, Academic Analytics.

## Contexto

El proyecto se basa en el desarrollo de conocimiento sobre la tecnología innovadora del BigData y el análisis de los modelos necesarios sobre dicha tecnología que aportan valor agregado a las instituciones educativas. Desarrollado desde la facultad de Tecnología de Información de la Universidad Abierta Interamericana, y la Secretaría de Investigación de UAI (CAETI)

## Introducción

Por Big data nos referimos exactamente al tratamiento y análisis de enormes repositorios de datos, tan desproporcionadamente grandes y desestructurados que resulta imposible tratarlos con las herramientas de bases de datos y analíticas convencionales. La tendencia se encuadra en un entorno que no nos suena para nada extraño: la proliferación de páginas web, aplicaciones de imagen y vídeo, redes sociales, dispositivos móviles, aplicaciones, sensores, *internet de las cosas*, etc. capaces de generar, según IBM, más de 2.5 quintillones de bytes al día, hasta el punto de que el 90% de los datos del mundo han sido creados durante los últimos dos años. Hablamos de un entorno absolutamente relevante para muchos aspectos, desde el análisis de fenómenos naturales como el clima o de datos sismográficos, hasta entornos como salud, seguridad o, por supuesto, el ámbito educativo.

Las características del bigdata se centran en las tres v:

*Volumen:* Grandes volúmenes de datos, a partir de TeraBytes o incluso PetaBytes.

*Variación:* El concepto de Big Data también suele venir acompañado de diversos tipos de fuentes de datos, ya sean estructurados o no estructurados.

*Velocidad:* La frecuencia de las actualizaciones de estas grandes bases de datos también es un punto muy a tener en cuenta. Es por ello por lo que su procesamiento y posterior análisis también ha de realizarse prácticamente en tiempo real para poder mejorar la toma de decisiones en base a la información generada

Bigdata permite procesar gigantescos volumen de datos, no por medio de una supercomputadora, sino a través de computadoras pequeñas unidas en redes de procesamiento. Utilizaremos tecnología Hadoop (código abierto y gratuito) la cual permite a las aplicaciones trabajar con miles de nodos y petabytes de datos. Hadoop se inspiró en documentos Google.

Las soluciones de bigdata para en ámbito académico se pueden centrar en el aspecto institucional (Academic Analytics) como en el aprendizaje (Learning Analytics). En este proyecto nos centraremos en el aspecto institucional, relevando y desarrollando modelos

que permitirán por medio del BigData optimizar la deserción en el ámbito universitario como así también realizar un seguimiento más exhaustivo de los alumnos.

Nuestro alcance propone utilizar la información disponible (redes sociales, sistemas educativos, webs institucionales, etc) con el objetivo de generar e implementar modelos que permitan cumplir con nuestros objetivos planteados.

También abordaremos las problemáticas éticas que acompañan el desarrollo de dichos avances tecnológicos para el ámbito académico, la privacidad y la manipulación de datos sensibles ponen en evidencia los riesgos de tales prácticas.

## Resultados y Objetivos

En la actualidad el proyecto se encuentra en su primera fase de maduración y análisis detallado.

A continuación se describen los correspondientes objetivos específicos a desarrollar:

- a) Generación de un entorno de desarrollo sobre tecnología BigData (Hadoop y plataforma de análisis).
- b) Estudio de modelos actuales y desarrollo de modelos específicos, que permitan identificar estudiantes en riesgo, de forma de tomar decisiones tendientes a intervenir con el fin de reducir la deserción y aumentar las tasas de graduación.
- c) Estudio de riesgos y cuestiones éticas.
- d) Desarrollo de un caso práctico sobre una institución académica, análisis de resultados y objeciones.

## Formación de Recursos Humanos

A continuación se describen los productos asociados a dicho proyecto:

<b>Producto</b>	<b>Nº</b>
Artículos en Revistas Internacionales con referato	1
Artículos en Revistas nacionales con referato	2
Presentaciones en Congresos	3
Tesis aprobadas	1
Recursos Humanos formados	5

## Referencias

- Campbell, J.P. DeBlois, P.B., & Oblinger, D.G. (2007). Academic analytics: A new tool for a new era.

- Carmean, C., & Mizzi, P. (2010). The case for nudge analytics.

- Norris, D., Baer, L., Leonard, J., Pugliese, L., & Lefrere, P. (2008). Action analytics: Measuring and improving performance that matters in higher education.

- Sarah Lacy, "Peter Thiel: We're in a Bubble and It's Not the Internet, It's Higher Education,"

- Bollier, D. (2010) 'The Promise and Peril of Big Data'

- Berry, D. (2011) 'The Computational Turn: Thinking About the Digital Humanities', Culture Machine.

- Boyd, d. and Ellison, N. (2007) 'Social Network Sites: Definition, History, and Scholarship',

- Acquisti, A. & Gross, R. (2009) 'Predicting Social Security Numbers from Public Data'

## Objetos de Aprendizaje en ambientes centrados en el alumno.

Zulema Beatriz Rosanigo<sup>(1)</sup>, Pedro Bramati<sup>(2)</sup>, Claudia López de Munain<sup>(1)</sup>, Silvina Bramati<sup>(2)</sup>, Leda Cotti<sup>(2)</sup>  
[brosanigo@yahoo.com.ar](mailto:brosanigo@yahoo.com.ar), [pedrobramati@speedy.com.ar](mailto:pedrobramati@speedy.com.ar)

<sup>(1)</sup> Departamento de Informática – <sup>(2)</sup> Departamento de Ingeniería

Facultad de Ingeniería – Sede Trelew – U.N.P.S.J.B. – Te-Fax (02965) 428402

### RESUMEN

Se presentan objetivos y principales resultados de la línea de investigación sobre Objetos de Aprendizaje (OA).

El grupo de investigación viene trabajando desde 1997 en Informática Educativa y desde 2007 en este tema.

Los principales ejes abordados son: metodologías, procesos y herramientas para el diseño de OA en ambientes centrados en el alumno, repositorios de OA, estándares de e-learning, protocolos de interoperabilidad, búsqueda federada y aspectos relacionados con el acceso abierto, derechos de autor y licencias tipo creative commons.

**Palabras clave:** Objeto de aprendizaje, Repositorio de Objetos de Aprendizaje, metadatos, reusabilidad.

### CONTEXTO

La investigación se desarrolla en Facultad de Ingeniería (Sede Trelew) de la Universidad Nacional de la Patagonia San Juan Bosco.

Intervienen docentes e investigadores provenientes de distintas disciplinas y alumnos de las carreras de Ingeniería Civil y Licenciatura en Informática/ Sistemas.

### 1. INTRODUCCIÓN

En la educación universitaria, cada vez más se procura que el alumno sea el centro del proceso enseñanza-aprendizaje, enfocándose en el alumno que aprende más que en el profesor que enseña. Según Mohanan [1], la enseñanza puede definirse como una actividad que facilita el aprendizaje. El que aprende es el

alumno y la tarea del docente es facilitar ese aprendizaje.

El cambio educativo hacia este modelo centrado en el alumno y la aparición de nuevos entornos virtuales de aprendizaje han propiciado el desarrollo de los objetos de aprendizaje.

Un OA es un conjunto de recursos, autocontenible, diseñado y creado en pequeñas unidades *digitales*, con un *propósito educativo* para maximizar el número de situaciones en las que se puede utilizar [2,3,4,5,6].

A partir de Internet ha cambiado la manera en que se diseña, desarrolla y distribuye el material educativo y aumenta la necesidad de intercambiar y reutilizar recursos educativos en procesos de aprendizaje apoyados por tecnología [2].

El paradigma de Objetos de Aprendizaje plantea como objetivos [6]:

- Favorecer el acceso a los contenidos educativos.
- Optimizar los recursos destinados a su producción mediante su reutilización.

Para alcanzarlos se plantean tres estrategias:

- Descripción de los contenidos con metadatos normalizados, tal como LOM (Learning Object Metadata), ampliamente utilizado, desarrollado por IEEE (Institute of Electrical and Electronics Engineers) [7], cuyo propósito es simplificar las operaciones de búsqueda, gestión e intercambio de OA.
- Interoperabilidad de los contenidos en entornos tecnológicos diferentes, esto significa que materiales creados para un entorno de enseñanza aprendizaje como



puede ser Moodle, podría utilizarse sin inconvenientes en otro entorno como aTutor o WebCT.

- Modularidad y agregación de los contenidos. Generando OA modulares se favorece la reutilización. Por un lado, se pueden utilizar en forma independiente, y por otro, al agrupar varios OA en un diseño curricular mayor, con criterios concretos y estableciendo entre ellos una relación determinada, permiten logros que cada uno por sí solo nunca alcanzaría [8].

Con los OA se pretende compartir y reutilizar recursos educativos en procesos de aprendizaje apoyados por tecnología.

Un aspecto a tener en cuenta en relación a la reutilización, es la adopción de estándares que garanticen que los OA puedan ser utilizados en diferentes entornos tecnológicos y almacenados en repositorios, propiciando su reutilización a partir de la descarga de los mismos. Existen varias organizaciones dedicadas al desarrollo de estándares, especificaciones y modelos de referencia, que incluyen, entre otros, la estructuración de los datos, su descripción a través de los metadatos, el empaquetamiento de los contenidos y su secuenciación, a fin de facilitar la interoperabilidad, reusabilidad, adaptabilidad, accesibilidad y durabilidad de los OA. Si bien los estándares no garantizan la reutilización, indudablemente la propician.

Más allá del potencial de reusabilidad que tienen, los OA constituyen un medio para la construcción del conocimiento y permiten lograr un aprendizaje flexible y adaptable al itinerario pedagógico que elija cada alumno.

## 2. LÍNEAS DE INVESTIGACIÓN y DESARROLLO

Se viene trabajando en la línea de Informática Educativa desde 1997, comenzando a investigar el paradigma de los Objetos de Aprendizaje en 2007, a través de los proyectos PI 628 “Hacia un repositorio de Objetos de Aprendizaje” que dio origen al repositorio de

OA Graduate! y PI 912 “Procesos y herramientas para el desarrollo de Objetos de Aprendizaje en ambientes de aprendizaje centrados en el alumno” que continúa la línea de investigación comenzada con el PI 628 y promueve el desarrollo de OA en la comunidad educativa.

Los principales ejes abordados en relación a los OA son:

- Metodología y patrones de diseño de OA, criterios de evaluación de la calidad del OA, consideraciones para la evaluación de los contenidos.
- Criterios de clasificación de OA, estudio de taxonomías y ontologías.
- Estándares existentes para metadatos, para paquetes de contenidos y para repositorios.
- Diseño educativo basado en objetos de aprendizaje. Ambientes centrados en el alumno.
- Análisis de repositorios existentes, con énfasis en la funcionalidad, prestaciones y distinción de aspectos positivos y negativos. Puesta en marcha del repositorio de OA Graduate, desarrollado en el marco de una tesina de grado.
- Técnicas de búsqueda. Protocolos de interoperabilidad. Búsqueda federada. Acceso Abierto.
- Modelos y escenarios para la integración de objetos de aprendizaje y herramientas de gestión de contenido.
- Análisis de herramientas de diseño de material educativo. Experimentación con herramientas de software libre que facilitan la generación de materiales educativos, especialmente aquellas que permitan crear actividades interactivas y puedan ser usadas por docentes con pocos conocimientos informáticos.
- Derechos de autor y licencias tipo Creative Commons.

### 3. RESULTADOS OBTENIDOS/ESPERADOS

Se han investigado y analizado varios modelos y metodologías para el diseño y desarrollo de OA, propuestas por diferentes autores [9,10,11,12,13,14,15]. Las mismas han servido de base para establecer una guía metodológica para el diseño de OA y generación de cursos basados en OA [8] que ha sido ampliamente probada.

Siguiendo la metodología propuesta, se están implementando OA en diversas áreas, tales como Informática, Matemática, Sistemas de Representación, Instalaciones en edificios, Física.

Para el desarrollo de contenidos educativos se analizaron diversas herramientas y se programaron cursos y talleres para su aprendizaje.

Para concientizar a los docentes en la importancia de generar materiales educativos compartibles para sus cátedras, se dicta un curso de posgrado, de cuatro meses de duración, para dar a conocer técnicas, procedimientos, especificaciones, estándares y herramientas que permitan la creación de objetos de aprendizaje interoperables, reutilizables y compartibles para desarrollar contenidos educativos en línea.

El material del curso ha sido desarrollado como OA siguiendo la metodología propuesta [8] y utilizando la misma herramienta con que los participantes crearán los OA (eXelearning).

También se han dictado varias conferencias sobre esta temática, algunas de ellas de carácter internacional.

Se han analizado distintas alternativas de software abierto para crear repositorios observando principalmente la documentación provista, las funcionalidades ofrecidas, el diseño arquitectónico y las posibilidades de evolución y personalización, tales como: DOOR (Digital Open Objects Repository), ARIADNE (Alliance for Remote Instructional Authoring and Distribution Networks for Europe), PlanetDR y DSpace.

Se diseñó y construyó el repositorio Graduate! (<http://www.roa.unp.edu.ar:8080/graduate/>) utilizando DSpace, y se lo está poniendo en marcha.

### 4. FORMACIÓN DE RECURSOS HUMANOS

Desde el punto de vista de formación de recursos humanos se ha logrado un gran crecimiento y afianzamiento de los integrantes en el trabajo interdisciplinario, motivando a los docentes y alumnos a participar de actividades de investigación y en jornadas y eventos científicos.

Se capacitaron varios docentes en la generación de OA y así poder desarrollar y socializar material didáctico para ser utilizado en diferentes entornos educativos virtuales.

Se ha logrado una fuerte interacción y sinergia con otros grupos de investigación en áreas afines, tanto de nuestra propia universidad como de otras del país y de España, Chile y Ecuador, con los que se han realizado diferentes actividades.

A través del curso de posgrado anteriormente descrito, se capacitaron a integrantes de la Unidad Ejecutora de los Proyectos “TIC y formación Web 2.0 para la inclusión social y el desarrollo sostenible” y “Capacitación y Gestión del Conocimiento con Herramientas Web 2.0 para Docencia Universitaria, Gestión Administrativa y Educativa y Desarrollo Profesional Continuo en Argentina, Chile y Ecuador”, lo que se vio reflejado en la generación de OA que sirvieron como material didáctico en los cursos e-learning impartidos desde esos proyectos.

### 5. BIBLIOGRAFÍA

1. Mohanan, K.P. (2005) *Assessing Quality of Teaching in Higher Education* <http://www.cdtl.nus.edu.sg/publications/assess/default.htm>, Centre for Development of Teaching and Learning, National University of Singapore
2. Wiley, David A. (2001) “Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy”.

- <http://www.elearning-reviews.org/topics/technology/learning-objects/2001-wiley-learning-objects-instructional-design-theory.pdf>
3. Chiappe, A. (2009). Objetos de aprendizaje 2.0: una vía alternativa para la re-producción colaborativa de contenido educativo abierto. Colección: Univirtual Objetos de Aprendizaje Prácticas y perspectivas educativas ISBN: 958-8162-65-3 Pontificia Universidad Javeriana – Cali.
  4. Paur, A.B. (2008): Objetos de Aprendizaje. Factores que potencian su reusabilidad. Tesina Master en Ingeniería de Medios para la Educación.
  5. Polsani (2003). Use and abuse of reusable Learning Objects Pithamber R. Journal of Digital Information, Volume 3 Issue 4 Article No. 164, 2003-02-19
  6. Rosanigo, Z. B. (2012) Objetos de Aprendizaje en “Capacitación y Gestión del Conocimiento a través de la Web 2.0” DYKINSON S.L. Madrid. En prensa.
  7. IEEE (2002). Draft Standard for Learning Object Metadata. [http://ltsc.ieee.org/wg12/files/LOM\\_1484\\_12\\_1\\_v1\\_Final\\_Draft.pdf](http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf).
  8. Rosanigo, Z. B., Paur, A.B, Saenz Lopez M., (2010) “Nuevas tendencias para el diseño de materiales en entornos virtuales: Objetos de aprendizaje”. Ed. Universidad Nacional de la Patagonia, ISBN 978-950-763-100-9.
  9. Alvarado Silva, Alan. (2001). Metodología de Objetos de Aprendizaje en el e-learning como Herramienta para la Construcción de Competencias. <http://repositorial.cuaed.unam.mx:8080/jspui/bitstream/123456789/2989/1/2-cl-Alan%20Alvarado%20Silva.pdf> Accedido: Mayo 2011
  10. Contreras Paredes, Jorge Polo. (2006). Ambiente Colaborativo Multidisciplinar, para la Construcción de Objetos de Aprendizaje. 2º Encuentro en línea de Educación y software Libre. EDUSOL
  11. Contreras, Jorge, Sánchez Arias, Victor Germán, Hernández, Norma (2007). COBÁ: Un ambiente colaborativo para el diseño y desarrollo de material educativo basado en la tecnología de objetos de aprendizaje y Software Libre. “Tercer Encuentro en línea de Educación y software Libre”
  12. Instituto Tecnológico de Sonora. (2009). Introducción al proceso de diseño y desarrollo de Objetos de Aprendizaje. [http://biblioteca.itson.mx/oa/formacion\\_profesores/induccioa/oa/index.html](http://biblioteca.itson.mx/oa/formacion_profesores/induccioa/oa/index.html)
  13. Morales, Erla, García, Francisco J., Barrón, Ángela, Berlanga, Adriana J., López, Clara. (2005). Propuesta de Evaluación de Objetos de Aprendizaje. <http://www.uoc.edu/symposia/spdece05/pdf/ID06.pdf>.
  14. Osorio Urrutia, Beatriz, Muñoz Arteaga, Jaime, Álvarez Rodríguez, Francisco, Arévalo Mercado, Carlos. (2008). Metodología para elaborar Objetos de Aprendizaje e integrarlos a un Sistema de Gestión de Aprendizaje. [http://www.colombiaaprende.edu.co/html/mediateca/1607/articles-172721\\_archivo.pdf](http://www.colombiaaprende.edu.co/html/mediateca/1607/articles-172721_archivo.pdf)

Taller Revisión de Temas de Matemática, modalidad e-learning y presencial.

**Amigone Jorge<sup>1</sup>, Parodi Carlos<sup>2</sup>, Mansilla Alejandra<sup>2</sup>.**

Facultad de Ingeniería / Universidad Nacional de La Pampa.

Calle 110 N° 390, General Pico, La Pampa

Teléfono: 02302 – 422780. INT 6411

[jamigone@hotmail.com](mailto:jamigone@hotmail.com)<sup>1</sup>; {parodic, mansilla}@ing.unlpam.edu.ar<sup>2</sup>.

## Resumen

Presentamos una propuesta desarrollada por profesores del área Matemática y del área de Informática, de la Facultad de Ingeniería de la UNLPam, pertenecientes a asignaturas de primer año de las distintas carreras que se dictan en la institución. La propuesta está basada en la selección de contenidos y metodologías de enseñanza adecuadas que tiendan a favorecer la articulación entre el Nivel Medio y la Universidad, en un contexto de reflexión y trabajo.

Las acciones se implementan en el último año de Nivel Medio, para tratar de ampliar los conocimientos de los estudiantes, intentado mejorar sus habilidades con el aprendizaje de la matemática y que éstos puedan ser aprovechados por aquellos que pretendan incursionar en carreras pertenecientes o no a nuestra Facultad. Estas acciones intentan ser una extensión universitaria de la institución al conjunto de la sociedad que la sustenta tanto en lo local, como en lo regional

Se trabaja con dos tipos de metodología de enseñanza, una en forma Presencial y la otra en un Entorno Virtual de Aprendizaje (EVA), con el propósito de llegar a una comparación cualitativa de las mismas. Cuyo resultado permitirá tomar decisiones institucionales al momento de direccionar los esfuerzos.

**Palabras Claves:** Matemática de Nivel Medio, Modelos de enseñanza: Presencial y Entorno Virtual de Aprendizaje (EVA), Articulación.

## Contexto

El proyecto “Aportes para la enseñanza de la Matemática a ingresantes a la Universidad, con recursos y herramientas bajo la modalidad presencial y virtual”, está inserto en la línea de Investigación y Desarrollo, comprendida en el área de Tecnología Informática Aplicada en Educación, desarrollándose en la Facultad de Ingeniería de la Universidad Nacional de La Pampa.

Unas de las acciones propuestas como metodología de trabajo por los docentes investigadores en el transcurso de los años que durará el proyecto, son los Talleres de Revisión de Temas de Matemática, uno en la modalidad Presencial, y otro en un Entorno Virtual de Aprendizaje.

## Introducción

Desde ya unos años esta Facultad ha detectado problemáticas y dificultades relacionadas con el desempeño y el rendimiento académico de los alumnos ingresantes en todas las carreras<sup>1</sup> que constituyen su oferta formativa.

A lo largo de estos años se visualizaron fundamentalmente problemáticas de deserción inicial y un significativo desgranamiento a los largo del primer año, Las causas principales detectadas entre otras, pueden resumirse en un abrupto cambio que significa el ingreso al mundo universitario, desconocimiento del funcionamiento de la Facultad y sus normativas, ausencia de hábitos y costumbres para organizar el tiempo disponible, dificultades para interactuar con los docentes, carencia de estrategias de aprendizaje propias de este nivel educativo y ausencia de hábitos y técnicas de estudio.

Frente a las dificultades mencionadas, las propuestas y estrategias para el desarrollo de política de ingreso, se enmarcan en diferentes líneas de acción. Una de ellas es “diseño y organización de una serie de cursos de tipo introductorios y optativos en el área de matemática, física, informática y electromecánica, entre los que se destaca: Taller de Revisión de Temas de Matemática bajo la modalidad e-learning, Taller de Revisión de Temas de Matemática con modalidad presencial.

a) Taller de Revisión de Temas de Matemática bajo la modalidad e-learning.

En la actualidad surgen nuevas Tecnologías de la Información y la Comunicación (TIC) que nos hacen la vida más cómoda y fácil, también nos impactan en el sector educativo. La educación con medios electrónicos (e-learning) es considerada una de las alternativas más prometedoras para elevar el nivel educativo y la capacitación de la población a nivel mundial.

Dentro de este nuevo paradigma de la educación está surgiendo una amplia variedad de nuevos conceptos, como son las plataformas LMS (Learning Management System, sistemas de gestión de aprendizaje) diseño instruccional y objetos de aprendizajes.

Un sistema de gestión de aprendizaje es una aplicación residente en un servidor de páginas web en la que se desarrollan acciones formativas. Es el lugar donde estudiantes, tutores, profesores o coordinadores se conectan a través de Internet para descargar contenidos, ver el programa de la asignatura, enviar correo al profesor, charlar con los compañeros, debatir en un foro, participar en una tutoría, etc. Las plataformas que se utilizaron en el Taller fueron Claroline y Moodle, en las distintas etapas del proyecto.

*Migración del Aula Taller de Revisión de Matemática de la Plataforma Claroline a Moodle.*

<sup>1</sup> Carreras: Ingeniería en Sistemas, Analista Programador, Ingeniería Electromecánica, Ingeniería Industrial, Ingeniería Electromecánica con orientación en Automatización.



**Claroline** (<http://www.claroline.net>)

Aula Virtual: [MATEMATICA - Revisión de temas de Matemática](#)

En esta plataforma se debe de realizar 5 pasos para poder acceder al contenido “Unidad 1 - Clase 1 – Introducción a teoría de conjuntos”, como lo demuestran las siguientes Imágenes. Sólo era repositorio de Información, en el aula se debía de crear también archivos con la ejercitación y evaluación, subiéndolas a la sección Secuencia de Aprendizaje.



Imagen del Contenido

**Moodle - Entorno Modular de Aprendizaje Dinámico Orientado a Objetos**

Aula Virtual: [Taller de Revisión de Matemática](#).

En esta LMS, directamente se accede al contenido, se presetan las consignas, se puede ejecutar Videos, Presentaciones Mutimediales. El aula es construida teniendo en cuenta la mediación pedagógica en la virtualidad.

Se denomina *e-learning* (electronic learning) al desarrollo de programas de enseñanza y aprendizaje a través de entornos virtuales. Es una modalidad de estudio totalmente online, donde los estudiantes se desvinculan del docente; dándole paso a la interacción con las TIC's en plataformas o entornos virtuales de aprendizajes.

Creemos que algunas ventajas de utilizar esta modalidad de enseñanza son:

- La separación física entre el estudiante y el docente, ya que la misma favorece la flexibilidad de espacio y tiempo en el proceso de aprendizaje.
- La utilización de multimedios, que genera la posibilidad de atender a los estilos de aprendizajes de cada estudiante y establecer trayectos opcionales.
- Se puede realizar un seguimiento del aprendizaje, motivando la autonomía del estudiante.
- Se puede planificar el curso y el material para que no sean simples textos electrónicos, sino que haya equilibrio entre los recursos de aprendizaje, ayudas al estudio, actividades y sistemas de autoevaluación eficaces.

Un proyecto e-learning es un conjunto de ideas y planes que se desarrollan para armar una vía de acceso al conocimiento a través de soportes tecnológicos, programando los mecanismos pedagógicos, tecnológicos, y administrativos con el objeto de desarrollar nuevas e innovadoras alternativas de capacitación. Este plan debe de tener objetivo/s, etapas, recursos y diferentes áreas de realización.

Las diferentes áreas que componen un proyecto y sobre las cuales se sustenta son las siguientes.

- a) De planeamiento
- b) De tutorías
- c) De diseño de materiales
- d) Evaluación.
- e) De tecnología

Son las mismas que contiene un proyecto educativo, con el agregado de las áreas específicas: diseño de materiales y tecnología multimedia.

#### *Implementación*

En nuestro caso la etapa de *planeamiento* se concretó teniendo en cuenta que el curso tendría una duración de 40 horas reloj comenzando en el mes de octubre no siendo conveniente prolongarlo más allá del mes de noviembre, ya que en ese momento los estudiantes del nivel polimodal se encontrarán abocados a la finalización de esa etapa, lo que supone que estarán rindiendo las últimas evaluaciones en todas las materias y algunos de ellos recuperando objetivos no alcanzados.

Se dispuso que al inicio del curso haya un único encuentro presencial para que los participantes y tutores/profesores se conozcan personalmente, compartan

actividades y experiencias que favorezcan el acercamiento, el trabajo en grupo y el establecimiento de vínculos para el enriquecimiento de todos.

Una vez que los participantes ingresen en el entorno virtual, tendrán a su disposición el cronograma de trabajo general -el cual les permitirá organizar sus tiempos de estudio- y un instructivo que los acompañará en un primer recorrido por los espacios del entorno y los orientará en el manejo de las herramientas propias del mismo: foros, correos, materiales, instructivos.

Como metodología de trabajo, cada lunes se dará inicio a una nueva unidad. Para ello se pondrá a disposición de los alumnos el material de lectura correspondiente, más los instructivos adicionales que sean necesarios para el uso de las nuevas herramientas sugeridas para el desarrollo propio de la unidad. Como espacio de debate y reflexión sobre los temas incluidos en la misma, se habilitarán foros, que serán coordinados por un tutor asignado para tal fin.

El correo electrónico estará disponible como medio sumamente útil para desarrollar la comunicación entre docentes - alumnos y alumnos - alumnos.

Los participantes tendrán acceso a la plataforma las 24 horas, sugiriéndose conectarse a la plataforma una vez al día, leer la información publicada, reflexionar sobre los materiales, y participar en forma activa.

En cuanto a las *tutorías*, se estableció que habrá tutores académicos y técnicos.

Los tutores académicos serán los encargados de atender todas sus dudas respecto al desarrollo de los contenidos y de las evaluaciones, y sus funciones serán:

- Orientar pedagógicamente al alumno en la construcción de su conocimiento en forma autónoma.
- Motivar y contener al alumno desde la interacción personalizada, frente a posibles dificultades, obstáculos o inconvenientes que puedan identificarse anticipadamente o que se le presenten al tutor dentro de su proceso de aprendizaje.
- Asesorar en lo referido a aspectos organizativos y/o administrativos.

Los tutores técnicos serán los encargados de atender los aspectos relativos a la inscripción y al uso de la plataforma, encargándose además de asesorar en lo referido a aspectos organizativos y/o administrativos del proyecto.

Considerando que los *materiales* son los recursos didácticos mediados por tecnologías como el texto, audio, material navegable, etc. Se diseñaron Objetos de Aprendizajes de cada unidad temática. Si bien L'Allier (1997) los define como "la mínima estructura independiente que contiene un objetivo, una actividad de aprendizaje y un mecanismo de evaluación"; Wiley (2002) los explica como "Cualquier recurso digital que se puede utilizar como apoyo para el aprendizaje", entonces tenemos que un OA, va en función de una gama enorme de recursos digitales, con diferentes niveles de uso.

Los contenidos desarrollados en cada OA fueron los siguientes:

CONJUNTOS

NÚMEROS REALES

EXPRESIONES ALGEBRAICAS - POLINOMIOS

ECUACIONES E INECUACIONES

FUNCIONES

SISTEMAS DE ECUACIONES E INECUACIONES

b) Taller de Revisión de Temas de Matemática con modalidad presencial.

Los talleres presenciales, orientados a trabajar contenidos matemáticos que serán requeridos en la etapa a la que los alumnos aspiran alcanzar, presentaron cambios sustanciales que mostraron las adaptaciones necesarias propuesta y proyectadas cada año por el cuerpo de docentes involucrados. Modificaciones que surgieron del análisis realizado del trabajo con los grupos de alumnos en los años previos.

En los últimos años, y en especial en el 2012, se implementó el taller en el que los alumnos mostraron especial interés en trabajar los contenidos propuestos formando grupos principalmente a su elección. Esto último se pensó por la afinidad que ellos tenían al provenir de medios compartidos.

En ese año se experimentó una evaluación diagnóstica similar a la que se esperaba tomar en las diferentes Facultades de Ingeniería del país. Del análisis de los resultados surgieron las dificultades del trabajo matemático. Dificultades que se retomaron al momento del trabajo en el taller de los contenidos correspondientes.

Se analizaron las dificultades en el trabajo matemático, tanto desde el punto de vista de lo que se espera de los estudiantes, al considerar la transición de la escuela media a la educación superior y para algunas de las nociones en particular. Se observa (incluso en investigaciones) que en el paso de una etapa educativa a otra, los estudiantes (y profesores) siempre tratan de justificar sus dificultades debido al enfoque del concepto que se supone debe o no estar disponible (conocimientos: técnicos, movilizables y disponibles, según definiciones de A. Robert, 1997) en el alumno. La cuestión de las diferentes maneras de acercarse a esas nociones no está necesariamente vinculada únicamente a la cuestión de los conocimientos disponibles de los alumnos, sino que también depende de otros factores tales como los sistemas culturales, los diferentes obstáculos educacionales y las expectativas de trabajo de los estudiantes que son diferentes en la educación media y en la universidad. Artigue (2001) intenta demostrar la importancia de abordar las matemáticas en todas sus facetas a nivel universitario, mientras se dirigía a los futuros matemáticos o aquellos cuyas relaciones futuras con las matemáticas será menos central. Esto demuestra la importancia de estudiar y cuestionar los procesos de aprendizaje en esta etapa de escolarización. Parece que hay pocos estudios en este sentido y que, tanto en la escuela media y en la educación superior, los estudiantes tienen grandes dificultades, a menudo carecen de las destrezas y habilidades necesarias para su desarrollo educativo y profesional y los profesores tienen pocos recursos para ayudarlos. Muchas de estas dificultades no

están asociadas con la movilización (Robert, 1998, C. Parodi 2004) de los conocimientos matemáticos adquiridos previamente. Podemos ver que la falta de conocimientos movilizados tiende a aumentar durante las diferentes etapas de la escolarización. Esta situación se refleja en la calidad de la enseñanza de las matemáticas en los dos niveles educativos aquí nombrados. Por lo tanto, estudiar la transición entre la escuela secundaria y la universidad puede permitir identificar los elementos que se consideran herramientas disponibles en la mayor parte de la educación y que no alcanza ese nivel de conocimiento.

Por último, podemos decir que, en general, la asistencia de los talleres presenciales es muy elevada, lo que muestra una necesidad creciente del apoyo matemático a los futuros ingresantes. Inquietud intenta responder a las dificultades que venimos de nombrar y que no escapan al futuro estudiante universitario. Y, esperando acercarnos más al trabajo matemático en la universidad, en estos talleres, se procuró trabajar en un ámbito al cual los alumnos no están acostumbrados a participar, ya que los talleres se constituyen con un grupo de docentes y docentes alumnos de la Facultad de Ingeniería y en las instalaciones de dicha Facultad de la UNLPam, procurando poner en práctica la manera en que se trabaja en el nivel universitario.

### **Líneas de Investigación y desarrollo**

Consideramos que si al estudiante se le brinda la posibilidad de una apropiación significativa de los conocimientos básicos de matemática a través de una selección de estrategias y actividades con características que se aproximen a las habituales del Nivel Superior; con la utilización de material didáctico adecuado y de nuevas tecnología de la información y comunicación, se logrará estimularlos para que, a través de un trabajo sistemático, incrementen las posibilidades de lograr un buen desempeño al inicio de sus carreras.

Por eso nuestro Objetivo General es, acompañar al estudiante en instancias previas al ingreso al Nivel Superior a través de la implementación del sistema de enseñanza y de aprendizaje más apropiado, atendiendo las diversidades y propendiendo a la igualdad de oportunidades.

Teniendo en cuenta que la investigación es un proceso por el cual se construyen conocimientos acerca de una problemática de un modo sistemático y riguroso, esta investigación es de corte cualitativo porque nos permite entender el proceso utilizado al interactuar con ideas o problemas matemáticos. Manuel Santos Trigo (1996) nos afirma que mientras un análisis cuantitativo le da mucha importancia al producto final obtenido por el estudiante, es decir, a la solución del problema, el análisis cualitativo analiza las diferencias sutiles que se encuentran en el proceso utilizado al trabajar la solución del problema.

En referencia a las ventajas de los estudios cualitativos este mismo autor nos apunta que con datos cualitativos, uno puede preservar el flujo cronológico de la información, acceder a la causalidad local y derivar explicaciones fructíferas. Además, resultados inesperados y

nuevas integraciones teóricas pueden aparecer en el proceso.

La estrategia seleccionada es la triangulación metodológica, porque se combinan encuestas (método cuantitativo) con entrevistas y observaciones. Con ella se intenta abordar de manera más completa la compleja trama del fenómeno educativo. Su ventaja reside en que sirve para relativizar la distorsión que el fenómeno imprime en el objeto de estudio, y en este sentido, se vincula directamente con la confianza y la validez de los datos en su conjunto. Estrictamente hablando, se esperaría que la triangulación produjera los mismos datos a través del uso de técnicas de medición diferentes.

Al optar por este tipo de estrategia de investigación se es consciente de que los resultados estarán altamente influenciados por las interacciones sociales que se den en ese grupo en particular durante su actividad cotidiana, pero dado que se piensa realizar estas observaciones en grupos presenciales y virtuales, se prevé obtener variada información como para extraer conclusiones.

El análisis de objetos de aprendizajes que en la actualidad se viene desarrollando atendiendo a la creciente preocupación por mejorar los resultados de los ingresantes a las Universidades, permitirá establecer un marco que sustente la elaboración de nuevos materiales que pensamos proponer a nuestra Universidad para el trabajo futuro con estudiantes ingresantes, atendiendo a posibles adaptaciones.

### **Resultados y Objetivos**

**Resultados:** A partir de la implementación de este proyecto incluyen:

Mejorar el proceso de aprendizaje de los estudiantes que aspiran a ingresar al Nivel Superior, con la utilización de técnicas adecuadas que permitan la incorporación de conocimientos nuevos a partir de sus conocimientos previos.

Disponer de material didáctico organizado para la implementación de los cursos, como material impreso, digitalizado y creación del Objeto de Aprendizaje.

Evaluar el proceso de enseñanza y de aprendizaje llevado a cabo en cada instancia.

Contar con un grupo de docentes actualizados que integren una comunidad autocrítica, reflexiva, que tenga en cuenta el desarrollo de habilidades, destrezas y valores.

Mediante las actividades previstas se espera que los estudiantes construyan y/o reconstruyan conceptos matemáticos básicos; planteen y resuelvan situaciones problemáticas; desarrollen habilidades que les permitan, desde el pensamiento matemático, enfrentar nuevas situaciones buscando, además, caminos alternativos para su resolución; que interpreten los resultados obtenidos y que analicen la factibilidad de los mismos dentro del contexto de la situación planteada.

### **Objetivos:**

- Recordar y/o reforzar algunos conceptos fundamentales de matemática, desarrollados en el transcurso del Nivel Medio/Polimodal.

- Complementar la preparación básica del estudiante tendiendo a favorecer sus posibilidades de acceso a la educación superior.

### Formación de Recursos Humanos

La estructura del equipo de trabajo de la línea de I/D presentada es la siguiente:

**Director:** Mg. Jorge Amigone

**Co Director:** Mg. Carlos Parodi

**Investigadora:** Esp. Alejandra Mansilla

### Referencias

- Area Moreira, M. (2006). Veinte años de políticas institucionales para incorporar las tecnologías de la información y comunicación al sistema escolar. En J. M. Sancho (Coord.) *Tecnologías para transformar la educación* (pp. 199 - 232). Madrid: AKAL-UIA.
- Aparici, Roberto. (2005). Teorías de aprendizaje para el diseño de material pedagógico. El impacto del aprendizaje significativo, colaborativo y de la autogestión o aprendizaje independiente. Unidad I. Tema II. Curso de actualización profesional Medios y tecnología para la enseñanza a distancia. México: Organización Universitaria Interamericana (OUI), el Colegio de las Américas y la Universidad Veracruzana.
- Artigue, Michele. (1995). Ingeniería didáctica en educación matemática. La enseñanza de los principios del cálculo: problemas epistemológicos, cognitivos y didácticos (pp.97-135). Grupo Editorial Iberoamérica.
- Brousseau, G. (1987). Fondaments et Méthodes de la Didactique. Recherches en didactique des mathématiques. 7.2 (pp. 33-115). Córdoba: Versión en español publicada por la Facultad de Matemática, Astronomía y Física de la Universidad de Córdoba
- Burbules N. y Callister, T. (2001). Educación: riesgos y promesas de las nuevas tecnologías. Barcelona: Granica.
- Chan Núñez, M. (2005). Córdoba: Competencias mediacionales para la educación en línea. México: Revista Electrónica de Investigación Educativa. Redie. Vol 7\_2.
- D.A. Wiley, II (2002) Connecting learning objects to instructional design theory: A definition, a metaphor, and taxonomy, Utah State University.
- L'Allier, James (1997), Frame os Reference: NETg's Map to ist Products, Their Structures an Core Beliefs.



## Identificación de las estrategias de uso de recursos y actividades en las aulas virtuales en la Facultad de Ciencias Económicas de la UNPSJB

Marta Isabel Dans<sup>(1)</sup>, Maria Elena Sendín<sup>(2)</sup>, Adriana Almeida<sup>(1)</sup>, Pablo De Battisti<sup>(3)</sup>  
[mdans@cpsarg.com](mailto:mdans@cpsarg.com) [gonsen@infovia.com.ar](mailto:gonsen@infovia.com.ar) [adriana.almeira@gmail.com](mailto:adriana.almeira@gmail.com) [pablodebattisti@yahoo.com.ar](mailto:pablodebattisti@yahoo.com.ar)

<sup>(1)</sup> Departamento de Administración - Facultad de Ciencias Económicas - U.N.P.S.J.B.- Trelew

<sup>(2)</sup> Departamento de Matemática - Facultad de Ciencias Económicas - U.N.P.S.J.B.- Trelew

<sup>(3)</sup> ISFD N° 801 "Juana Manso" Avenida Rawson y Mitre - Trelew

### CONTEXTO

La unidad ejecutora del proyecto "Identificación de las estrategias de uso de recursos y actividades en las aulas virtuales en la Facultad de Ciencias Económicas de la UNPSJB", está integrada por docentes universitarios y del nivel medio, que en su mayoría tienen más de diez años desarrollando investigación en la línea de la incorporación de TIC en educación superior. El presente proyecto se desarrolla en la Facultad de Ciencias Económicas de la UNPSJB, cuenta con la participación de docentes de otras facultades de la UNPSJB. Este es el noveno proyecto que continúa en la línea de proyecto de investigación en el campo de informática educativa. Además de continuar con el crecimiento en las áreas ya experimentadas, aquí se propone extender el proyecto anterior incluyendo en el análisis de las estrategias de uso de los recursos y actividades que posee la plataforma Moodle de manera tal de sumar elementos fundamentales en el sistema en estudio la modalidad b-learning que ha comenzado a tomar fuerza en la UNPSJB y que indudablemente constituye la antesala de la modalidad a distancia. Por ser la unidad ejecutora en su mayoría de la Facultad de Ciencias Económicas de la UNPSJB, se han realizado aplicaciones en este ámbito. Aunque desde el año 2009 se han incorporado en la capacitación a docentes de otros niveles educativos. Además por haber incorporado en los cursos de capacitación la etapa de implementación de aulas virtuales y junto con ella corresponde un acompañamiento por parte de los tutores, miembros de la unidad ejecutora del proyecto, a los alumnos hemos alojado aulas de otras facultades de esta universidad y de instituciones educativas de nivel primario y medio.

### RESUMEN

El proyecto que antecede a la presente proposición, "Determinación de los principales actores y sus roles para los cursos de EaD en la Facultad de Ciencias Económicas de la UNPSJB", nos permitió analizar a los actores en sus roles relacionados a la plataforma MOODLE y determinar las habilidades que estos deben poseer para gestionar sus actuaciones en las aulas virtuales, a partir del análisis de la bibliografía y la experiencia en la capacitación impartida.

Esta propuesta permite completar el proyecto anterior. Dado que percibimos la necesidad de precisar las estrategias de uso de los distintos recursos y actividades que brinda MOODLE. Puesto que dicha plataforma propone una variada gama de posibilidades que hasta el momento no habíamos experimentado, en cuanto a recursos a disposición de los cursos y actividades a implementar durante el desarrollo de los mismos. Consideramos que se ha formado un grupo numeroso de docentes capacitados a lo largo de los cuatro últimos años que todos utilizan la plataforma deseosos de compartir, mejorar e incorporar nuevas propuestas a sus alumnos a través de las aulas virtuales.

La investigación de las distintas estrategias de uso será especialmente enriquecedora para las prácticas docentes en el nuevo entorno, teniendo en cuenta las posibilidades que brinda en cuanto al desarrollo de las capacidades de colaboración, cooperación, autoevaluación, socialización, aprendizaje autónomo, manejo de las TIC relacionadas al ámbito de la capacitación.

**Palabras clave:** *estrategias didácticas, e-learning, plataforma virtual, capacitación docente continua.*

### 1. INTRODUCCION

En el presente proyecto nos proponemos continuar con el estudio de la plataforma para ofrecer a los docentes un aporte de mayor profundidad como lo es analizar las posibilidades de uso de las distintas herramientas que posee la plataforma adoptada desde el año 2005 y comenzada a utilizar en el año 2006.

La trayectoria en investigación en el área de tecnologías informáticas y de comunicación nos permitió corroborar lo expresado en los párrafos anteriores y nos exige continuar mejorando las distintas estrategias tecnológicas ya incorporadas en nuestro ambiente.

En los anteriores proyectos se han desarrollado: un modelo de diseño para desarrollo de software multimedial, se implementó el sitio Web de la Facultad de Ciencias Económicas de la UNP, se incorporaron los servicios de Internet de WWW y foros de discusión para materias con profesores viajeros y profesores locales. Oportunamente se

comprobó que tales innovaciones mejoraron la calidad en cuanto a la comunicación docente - alumno durante el proceso de enseñanza aprendizaje.

Es el quinto proyecto que continuamos en la línea de investigación en informática educativa. Diseñamos e implementamos la plataforma virtual para los cursos de educación a distancia, diseñamos el método MeCoVED (DANS 2003) para la construcción de videos didácticos digitales y seleccionamos la plataforma virtual para cursos de educación a distancia.

Además de mantener el crecimiento en las áreas ya experimentadas, aquí se propone investigar a fondo los roles necesarios para potenciar el entorno virtual de los cursos de educación a distancia, al servicio del aprendizaje de los alumnos.

Algunas de las actividades están planeadas para realizarlas individualmente, otras son grupales. En todos los casos la plataforma Moodle brinda al profesor herramientas confiables de total seguimiento y control.

Las actividades son muy variadas. Algunas de las que se implementan en el curso son: foro, chat, consulta, glosario, tarea, consulta, lecciones, cuestionario, scorm (sharable content object reference model), encuestas, wiki, taller, base de datos y diario.

En el presente proyecto se estudiarán las posibles estrategias didácticas que se desprenden de la aplicación de las herramientas que ofrece la plataforma, relacionada estas con los objetivos de enseñanza de los cursos de la FCE.

#### *Mejoras introducidas en el plan de capacitación*

Se agregaron dos etapas en los cursos de capacitación. La primera es la construcción de un proyecto de curso sobre la plataforma Moodle y la segunda es el acompañamiento del curso. Es decir los cursantes deben implementar un aula y llevar a cabo un curso de capacitación. La unidad ejecutora realiza el acompañamiento del curso planificado para asistir a los docentes durante la ejecución del mismo. De esta forma se espera mejorar el impacto de los cursos brindados ya que se ha observado que el mismo es muy pobre en cuanto al uso de los recursos dentro de las aulas abiertas. También se observó una falta de estructura didáctica en los cursos virtuales por ello se vio la necesidad de investigar a cerca de la didáctica que subyace bajo los cursos virtuales.

#### *Estructura didáctica de los cursos*

En los primeros meses de la investigación se ha buscado y analizado distinta bibliografía sobre la didáctica de aulas virtuales. Concluimos que todos los elementos presentes en un curso presencial deben estar en un aula virtual, solo se establece la forma de

virtualizarlos teniendo en cuenta que no todo lo tecnológicamente viable es educativamente pertinente. Cristóbal Suarez Guerrero define una Estructura Pedagógica de Aprendizaje EPA para la formación virtual que consta de cinco elementos, los cuales son dinámicos y ajustables al contexto en que se apliquen, no deben verse como una linealidad, ya que cada participante accederá a cada parte en el orden en que así lo determine su interés y necesidad en pos de alcanzar los objetivos de tareas propuestas. Estos elementos son: Para Empezar, Ampliación, Definición de tareas, Intercambio, Consolidación. Elegimos este modelo ya que encontramos reflejado en él casi todos los elementos de nuestros cursos de capacitación.

Para empezar, contiene el impulso, la motivación que realiza el docente referente al tema que se desea que los alumnos aprendan.

Definición de tareas, es el producto final, lo que los alumnos deben presentar y en donde están implícitos los aprendizajes.

Intercambio, en este espacio se realiza todo tipo de conversación, en forma textual, oral, visual o combinación de ellas, es fundamental en la modalidad virtual.

Consolidación, es la etapa en la que los alumnos reciben la retroalimentación de su tarea a través de la evaluación de la tarea. Es fundamental para el aprendizaje la forma en que el docente devuelve los resultados de las evaluaciones, ya que el alumno así podrá conocer sus debilidades para corregir su aprendizaje.

Ampliación, en esta etapa, Suarez Guerrero propone, ampliar las ideas trabajadas en la fase de la consolidación, buscando ampliar las fuentes de información y buscando de resolver nuevas situaciones problemáticas en las cuales se pueda aplicar lo aprendido. Hace hincapié en la necesidad de estimular el conocimiento metacognitivo del proceso de aprendizaje virtual. Es decir hacer reflexionar a los alumnos como construyen el conocimiento a través del aula virtual.

De todos los estadios anteriormente expuestos, no hemos realizado la reflexión metacognitiva del proceso de aprendizaje en la virtualidad, lo cual consideramos que es importante para que el alumno comprenda todas las acciones que debe realizar en la plataforma para aprender con los otros, de los otros y de su propia reflexión. El estudiante aprende a aprender en la virtualidad.

#### *Reconociendo algunos modelos en el uso de MOODLE*

La mayoría de los docentes de la Facultad de Ciencias Económicas no poseen formación pedagógica-didáctica por lo que resulta imprescindible la incorporación de la dimensión

pedagógica en la capacitación sobre el uso de la plataforma virtual. Se observa en muchas situaciones que se concibe a la plataforma virtual como un repositorio de material de estudio sin crear un entorno didáctico adecuado. Los resultados preliminares permiten reconocer la presencia de los siguientes modelos: academicista, tecnológico, humanista y psicognitivo, dentro de este último se destaca un modelo formativo b-learning, centrado en la teoría/metodología del conocimiento constructivista.

A) Modelos academicistas: Las estrategias pedagógicas de esta tendencia son racionalistas, privilegian la exposición de conocimientos por parte del docente, destacando que su rol fundamental es transmitir contenidos para que el estudiante pueda asimilarlos. El predominio de las teorías académicas en Educación a Distancia se complementa en tres tipos de razones: 1) La posibilidad de combinar y complementar el enfoque tecnológico con aspectos de nuestra cultura y condiciones contextuales; 2) Se establece que los estudiantes alcance los objetivos propuestos; y 3) Estandarización de los aprendizajes, presentando los conocimientos de forma estándar, con pretensión de resultados comunes.

Bajo este modelo de enseñanza en Educación a Distancia, el docente el único poseedor de los conocimientos, utilizándose la lectura de textos, el visionamiento de documentos audiovisuales, la manipulación de programas de las computadoras, como distintas modalidades para que el estudiante "asimile" los contenidos definidos por el profesor. El modelo descansa en la presentación eficaz, convincente y magistral de los contenidos que realiza el profesor, estando considerada la intervención del estudiante para aclarar dudas relativas a los contenidos presentados por el docente.

B) Modelos tecnológicos: Son compatibles con la necesidad de masificar el acceso a la educación y la democratización de la misma. Los mismos, permiten la planificación rigurosa de la enseñanza, el desarrollo de contenidos validados y la mediatización profesional de los recursos didácticos que se utilizan. Dentro de este modelo se utilizan permanentemente los medios de comunicación y las tecnologías para representar de manera eficaz los contenidos.

El modelo tecnológico se compatibiliza bien con un enfoque industrial eficiente, típico de muchos sistemas de Educación a Distancia. Aquí predomina la planificación sistemática, la validación de los contenidos, el control del proceso de aprendizaje, la retroalimentación que este tipo de modelos promueve, asegura que mayor cantidad de personas lograrán resultados similares.

Desde el punto de vista de la enseñanza, implica planificar el proceso y mediatizar los

contenidos de manera muy rigurosa. En el que se modela el proceso de interacción como un diálogo, entre el estudiante y los medios (por ejemplo, correo, teléfono, computador, etc.). Esta interacción es a su vez fuente de retroalimentación en la medida que permite obtener información del estado de avance en relación con los objetivos esperados) y sobre la eficacia del sistema mismo.

C) Modelos humanistas: Se centran en el desarrollo del ser y en el crecimiento personal, se centra en un método personalista con una forma de enseñanza poco interactiva cuestionando el lograr objetivos precisos y predefinidos. Se busca desarrollar la expresión de los estudiantes de acuerdo a sus intereses, buscando un desarrollo más integral y objetivos emergentes de los estudiantes.

D) Modelos psicocognitivos: están centrados en el reconocimiento de los procesos cognitivos y metacognitivos implícitos en el aprendizaje humano. Bajo este modelo, es típico que el docente invite a sus estudiantes a hacer "una exploración de su universo cognitivo", cuyo fin es explicitar sus conocimientos previos. A partir de estos modelos se han ido desarrollando ambientes de aprendizaje informatizados. Ello porque como recurso de aprendizaje posibilita que el estudiante "navegue" explore y experimente un aprendizaje por descubrimiento.

E) Modelo formativo b-learning, deriva del anterior modelo centrado en la teoría/metodología del conocimiento constructivista, implica la puesta en común de los contenidos (la posible descarga de los mismos a través del portal de la asignatura), la interacción y comunicación entre los alumnos y el docente, una mayor implicación del estudiante en su propio desarrollo académico quien tiene posibilidades de construir los conocimiento por sí mismo. Permite desarrollar además en los estudiantes la idoneidad para su futuro profesional, una mejor capacidad de análisis, organización, gestión y responsabilidad, aumento de la perspectiva y actitud de trabajo en equipo, cooperativo y colaborativo.

Consideramos que el modelo ideal es el formativo b-learning. Podemos observar en los cursos activos que los modelos formativos no se dan puros sino combinación de ellos.

## 2. LINEAS DE INVESTIGACION y DESARROLLO

La línea de investigación es la aplicación de las TIC en educación superior.

Dentro de ella se ha incursionado en distintas aristas: construcción de multimedios, creación de videos didácticos digitales, la Web educativa y actualmente plataformas virtuales para cursos.

En el proyecto actual "Identificación de las estrategias de uso de recursos y actividades en las aulas virtuales en la Facultad de Ciencias Económicas de la UNPSJB", se propone identificar las estrategias de uso de las herramientas que ofrece MOODLE relacionadas a los objetivos de aprendizaje que se planteen en los cursos de la Facultad de Ciencias Económicas (FCE) de la UNPSJB.

### 3. RESULTADOS OBTENIDOS/ESPERADOS

Actualmente la plataforma de la FCE posee más de 80 aulas de distintas materias pertenecientes a todas las delegaciones y a la sede central.

Desde el proyecto pasado se dictaron 11 instancias del curso introductorio y 3 instancias del curso avanzado.

La unidad ejecutora está indagando sobre las estrategias didácticas para analizar en base a ellas el uso de los recursos y actividades que ofrece Moodle. Introducirá en los actuales cursos de capacitación los resultados de los análisis a los que se arribe en el presente proyecto, de manera tal que llegue al cuerpo docente que utiliza dicha plataforma. Con esto se espera mejorar sustancialmente las actividades que se realizan sobre el aula, siendo estas propuestas un aporte significativo para el aprendizaje de los alumnos de los conceptos que se imparten en las distintas disciplinas.

La mayoría de la unidad ejecutora compuesta por doce docentes utiliza la plataforma para sus cursos. Lo cual produce un enriquecimiento a la hora de proponer actividades en los cursos de formación de la herramienta.

### 4. FORMACION DE RECURSOS HUMANOS

#### Capacitación de los docentes

Con el fin de avanzar en las investigaciones y determinar las necesidades de capacitación como así también el éxito de la misma se analizan las aulas virtuales creadas por los docentes que realizaron los cursos, analizando el uso que se hace de las mismas.

Como se expresó anteriormente, la plataforma Moodle brinda una serie de recursos, actividades, formas de evaluación, seguimiento, producción de informes de actividad y otras características con distintos niveles de complejidad en su configuración que hizo necesario planificar la capacitación en distintos niveles (cursos introductorio y avanzado)

La unidad ejecutora está capacitando en la gestión de las aulas virtuales, haciendo incipiente en la reflexión sobre el uso didáctico de las herramientas que abarca.

Estos cursos se ofrecieron en un primer momento a los docentes de la FCE, luego se hizo extensiva la invitación a docentes de otras

Facultades de la Universidad Nacional de la Patagonia como así también a docentes de otros organismos. El éxito de la convocatoria fue tan grande que fue necesario reeditarla en cinco oportunidades. Y aún hoy siguen llegando solicitudes. Las autoridades de la FCE han propuesto el curso introductorio como una capacitación continua que se impartirá dos veces al año.

Del análisis de las encuestas tomadas tanto al inicio como en la finalización de los cursos se desprende el impacto positivo que los cursantes se llevaron de la plataforma y más del 50% de los docentes solicitaron la apertura de sus aulas virtuales como complemento a la enseñanza presencial.

#### Cursos de capacitación

##### Nivel introductorio

Este curso ofrece un espacio de formación y acompañamiento para el diseño y creación de propuestas educativas en espacios virtuales. Los cursantes aprenderán a crear y gestionar un curso en Moodle en un contexto crítico y reflexivo acerca de las estrategias de uso de cada una de los recursos y las actividades que se analizan en el curso.

Los objetivos específicos son brindar a los participantes conocimientos y desarrollar habilidades para desempeñarse como administradores y profesores (provistos por la plataforma).

La modalidad de los cursos es a distancia. Los requerimientos técnicos son: Acceso a Internet al menos una hora tres veces por semana, poseer cuenta de correo electrónico, conocimientos de procesador de texto y mínima experiencia de navegación por Internet.

Se determinó un cupo de 35 alumnos, tiene una duración de dos meses. Los alumnos aprueban el curso habiendo realizado correctamente el 80% de las actividades obligatorias en el tiempo establecido. Además los alumnos deben presentar un trabajo de integración el cual podrán implementar en una segunda etapa del curso. Se estableció como escala de evaluación: Muy Bueno, Bueno y Desaprobado; para los alumnos que entreguen la actividad vencido el plazo solo recibirán la calificación de Bueno.

El material básico de lectura fue construido por la unidad ejecutora del proyecto. Las consignas de los foros y las actividades fueron consensuadas. Se utilizó como bibliografía de consulta la que está a disposición en el sitio oficial de la plataforma.

El curso está estructurado en diez unidades didácticas: el perfil, entorno MOODLE, contenidos, foros, calendario, roles, glosario, tareas, resguardo y consulta.

El contenido de cada unidad está estructurado en: descripción de la clase, zona de materiales, zona de comunicación y actividades.

##### Nivel Avanzado



En este nivel se mantiene la metodología y estructura del nivel anterior solo cambian las unidades temáticas: diario, Wiki, cuestionario, lecciones, la Web 2.0, taller, SCORMS y base de datos.

### **Evaluación de estado de situación de conocimientos relativos a los cursos**

Para tomar conocimiento de los saberes previos de los alumnos se toman encuestas referidas a la noción y uso de los distintos servicios que ofrece Internet como la Web, correo electrónico, foros, plataformas educativas. El gran porcentaje de la población solo conoce básicamente el correo electrónico y la Web y no siempre aplicadas a su práctica docente.

Al finalizar los cursos realizamos una encuesta de calidad referida al curso. Aproximadamente el 50% de los alumnos responden la encuesta final.

Todos los aspectos relacionados con el diseño del curso, actividades, contenidos y actividad de tutores fueron muy bien evaluados.

La tecnología, más expresamente la calidad de conexión, fue el ítem con menor puntuación.

El grupo consideró importante al curso en cuanto al aprendizaje logrado y revela la existencia del uso de un pensamiento reflexivo.

### **Conclusiones**

La acogida de la plataforma por parte del plantel docente de la UNPSJB fue muy importante, más del 60% de los cursantes abrieron cursos para complementar sus clases presenciales.

De todas las facilidades que se impartieron en el curso introductorio las más usadas son: enlazar archivos y direcciones web, el calendario, los foros de discusión, el foro de novedades.

Aumenta la necesidad de capacitación tanto en el manejo tecnológico como en las oportunidades pedagógicas de uso de los distintos servicios.

Sin duda en los foros de reflexión se realizan aportes importantes en cuanto a propuestas de uso de las distintas herramientas. En este momento nos encontramos analizando las aulas virtuales en uso y vemos que la mayoría aún no utiliza al máximo su capacidad.

La unidad ejecutora del proyecto ajusta sus cursos de capacitación realizando un segundo nivel en los cursos que consiste en la implementación y acompañamiento de un curso real. De esta forma pensamos que tendremos oportunidad de acompañar a los docentes en la incorporación de esta herramienta tan potente al servicio de la enseñanza y el aprendizaje.

Para aquellos profesores que ya hicieron los cursos en versiones anteriores les ofrecemos reuniones de apoyo para socializar las experiencias y mejorar sus prácticas.

## **5. BIBLIOGRAFIA**

MENA Marta (1996) "La educación a distancia en el sector público, manual para la elaboración de Proyectos", INAP – DNC.

DANS, Marta (2003) "El medio maestro virtual de EaD" en la Facultad de Ciencias Económicas de la UNPSJB Anuario.

MUÑOZ, Miguel Ángel "Nuevos entornos y posibilidades telemáticas en educación"

GARCÍA ARETIO, L. (2001). La educación a distancia. De la teoría a la práctica. Barcelona-España: Editorial Ariel.

ALONSO, C. M., GALLEGU GIL, D.J. (2000) "Aprendizaje y Ordenador". Madrid: Dykinson.

MOODLE DOC: <http://docs.moodle.org/es>

Artículos de congresos publicados en la Web:

La educación a distancia: virtualidad y tecnología Autores: De León Reyes Félix Amado, Esperón Lorenzana Sonia Lorena, García Mendoza Susana, Mora Rojas Héctor Alberto - México, D.F., Marzo 2001

La educación a distancia en México: ¿Quién y cómo la hace? - Luis César Torres Nabel

Blended learning, un modelo pertinente para la educación superior en la sociedad del conocimiento - Julio César González Mariño -jmarino@uat.edu.mx Universidad Autónoma de Tamaulipas

Introducing Tutor Professor to Online Distance Education. Autor: Mtra. Verónica Salinas Urbina. Revista EGE. Julio/Agosto de 2000

Formación de postgrado para capacitación docente en educación y formación abierta y a distancia a través de redes digitales. Autor: Miguel Zapata. Artículo consultado de: <http://www.um.es/ead/red/4/informe.pdf>

Identificación de habilidades para tutores en Educación a Distancia. Autores: Enrique Díaz de León y Antonio Millán A.

MENA Marta La agenda de educación a distancia. Comunicación colectiva por parte de la vicepresidencia del ICDE para América Latina y el Caribe.

DAVINI Cristina (1993) Métodos de enseñanza, Ed. Dantillana.

ASTOLFI Jean –Pierre (1993) "Trois Paradigmes pour les recherches en Didactique", Revue Francaise de Pedagogie, N 103 Avril-junin, 1993.

CAMILLONI Alicia W. "La investigación en la enseñanza, III. Profesores y alumnos. Barcelona, Paidós Educador/MEC.

ACTAS PEDAGÓGICAS de la Universidad de Palermo Vol 1 Nro 1, febrero 2006. Diseños de investigación "Sossier para el seminarios"

FACULTAD DE EDUCACIÓN "La investigación sobre las NTIC en la Educación Superior". Universidad de Palermo (pp. 9-17)

Badía, Antoni y Barberá, Elena (2002). Hacia el aula virtual: actividades de enseñanza y aprendizaje en la red. Web DIM.

BUCKINGHAM, David y Callister, Thomas (2006), Educación: riesgos y promesas de las nuevas tecnologías de la información, Granica, Buenos Aires.

CHAN, María Elena (1999), "Educación a distancia y competencias comunicativas", en Revista La Tarea; núm: 11, junio, disponible en línea: <http://www.latarea.com.mx/articu/articu11/mechan11.htm>  
Fecha consulta 15-05-11

# Metodología de Desarrollo de Herramientas Informáticas Didácticas para el Aprendizaje del Cálculo Diferencial e Integral

Lic. Silvia Liliana Muzachiodi, Ing. Irma Manuela Benitez  
Tecnología Informática Aplicada en Educación/Ciencias Exactas-Sistemas/Facultad de  
Ciencia y Tecnología/Universidad Autónoma de Entre Ríos  
San Martín 1680-Paraná-Entre Ríos  
0343-154 460 900  
[silviamuza@yahoo.com.ar](mailto:silviamuza@yahoo.com.ar)

## Resumen

Este trabajo plantea la conjunción de dos problemáticas importantes en el ámbito educativo universitario, la didáctica de la matemática y la aplicación de la Ingeniería de Software. Surge en el marco del proyecto PIDA: “Metodología de desarrollo de herramientas informáticas didácticas para el aprendizaje del Cálculo Diferencial e Integral”, donde se propone aprovechar los recursos de la Universidad (docentes, alumnos y saberes) de manera de producir nuevos conocimientos y nuevas estrategias de enseñanza que serán usadas en la cátedra de Cálculo Diferencial e Integral de 1er año de las carreras de Licenciatura en Sistemas de Información de la Universidad Autónoma de Entre Ríos. El resultado de la ejecución de la investigación será el desarrollo e implementación de un Aplicativo Web, diseñado sobre un tema específico de la cátedra, el cual será utilizado como material de estudio de la misma.

**Palabras clave:** Nuevas Tecnologías, Educación Matemática, Aplicativo Web.

## Contexto

El presente trabajo surge en el ámbito del proyecto PIDA: “Metodología de desarrollo de

herramientas informáticas didácticas para el aprendizaje del Cálculo Diferencial e Integral”, presentado en la Facultad de Ciencia y Tecnología de la Universidad Autónoma de Entre Ríos. En el mismo se genera una articulación horizontal entre las cátedras de 3er año, Ingeniería de Software II y Taller de Integración, la primera le brinda al proyecto los conocimientos necesarios para el análisis y diseño de un producto de software y la segunda aplicando los conocimientos anteriores de análisis y diseño le permite el desarrollo e implementación del Aplicativo Web, el resultado del mismo será sobre un tema específico de la asignatura Cálculo Diferencial e Integral de 1er año.

## Introducción

La sociedad de la información se caracteriza por sus rasgos esenciales: generalizaciones del uso de la tecnología en el trabajo y en el hogar, disponibilidad de redes de comunicación baratas y globales, posibilidad de acceso instantáneo a grandes fuentes de información y proporción importante de trabajadores dedicados a las tecnologías de la información y la comunicación. La sociedad está en continua evolución y las personas se tienen que integrar, tanto en su cultura, su educación

como en los medios concretos materiales y personales.

El concepto de "sociedad de la información" hace referencia a un paradigma que está produciendo profundos cambios en la humanidad. Esta transformación está impulsada principalmente por los nuevos medios disponibles para crear y divulgar información mediante tecnologías digitales. En esta "sociedad", la creación, distribución y manipulación de la información, forman parte importante de las actividades sociales, culturales y económicas. (CEPAL-2003).<sup>1</sup>

La incorporación de las TICS en el ámbito de la educación universitaria imprime al proceso de enseñanza del Cálculo, dinamismo, interactividad, innovación y tecnología y además motivan a los alumnos en el aprendizaje (Martín, E. – 2007)<sup>2</sup>.

Es necesaria la implementación de este tipo de herramientas tanto en docente como en alumnos para no quedar desfasados con la ciencia y la tecnología que trae aparejado un atraso en la cultura científica y tecnológica, además posibilitan de una manera más efectiva la atención de las diferencia individuales, logrando así aprovechar las capacidades de cada alumno, no sólo pensando en los creativos y talentosos, sino también los que se encuentran con una diversidad funcional. Es de destacar que también es una propuesta de equidad y democratización del conocimiento ya que es posible llegar con el mismo a todas aquellas personas que lo requieran y por algún motivo no puedan asistir a cursar la asignatura Calculo Diferencial e Integral, como son aquellos que están bajo la propuesta presentada por la Facultad

en el proyecto "Educación Universitaria en contextos de encierro", que se enmarca en el convenio firmado por la Facultad de Ciencia y Tecnología y el Servicio Penitenciario.

Desde el punto de vista constructivista (Ausubel D. - 1978)<sup>3</sup> se apunta a un proceso de enseñanza y aprendizaje apoyado en la acción del alumno a quien se estimula a reorganizar y ampliar sus conocimientos, a través de herramientas informáticas. En este proceso de adquirir el conocimiento cada alumno determinará sus propios tiempos, sus desarrollos y aplicará su propia metodología la cual en determinados momentos será de prueba y error, siempre guiados por los docentes con pautas claras de objetivos y evaluación.

Por otra parte, en la enseñanza de la matemática de los alumnos de primer año de las carreras universitarias se afronta distintos problemas con la adaptación y la articulación entre la escuela media y la universidad. Esto incide ampliamente cuando se enseña Cálculo en el primer año ya que para un correcto desenvolvimiento en esta asignatura se necesita un dominio adecuado, conocimientos y habilidades precedentes, para poder afrontar con éxito los nuevos contenidos, como por ejemplo los conceptos de álgebra básica, de trigonometría, etc.

Poco a poco en los últimos años se ha incorporado en las clases de las asignaturas correspondientes al área de matemática las herramientas que ofrecen las nuevas tecnologías, y se ha podido observar que el uso de estas han tenido mucha influencia no solo en el que aprende, sino también



en el que enseña Matemática (Kutzler, 2003)<sup>4</sup>.

El hecho del mayor acceso que se tiene a las computadoras en todos los ámbitos donde se mueven los alumnos, obliga a los docentes a producir cambios en la metodología de enseñanza. Y los alumnos como los docentes tienen acceso a buenos programas en la Web específicamente diseñados para trabajar en distintos temas de matemática.

La evolución y la aplicación de las TIC a la enseñanza de matemática, ofrece nuevas formas de enseñar, aprender, brindando amplias posibilidades didácticas y facilita a los alumnos a construir conocimientos y obtener una visión más amplia de la matemática. Es por ello que se propone desarrollar un prototipo de Aplicativo Web que apoye el proceso de enseñanza/aprendizaje de un tema determinando del Calculo Diferencial e Integral basándose en un modelo pedagógico constructivista. Los docentes como los alumnos interactúan en el desarrollo del tema, construyendo, creando, preguntando, criticando y reflexionando sobre la comprensión del conocimiento adquirido.

Mediante este aplicativo, el alumno aprende haciendo, el docente actúa como un facilitador que contribuye al desarrollo de las capacidades de los estudiantes, para que estos piensen, construyan el aprendizaje participando activamente mediante la interacción con el aplicativo, sus compañeros y reflexionen sobre lo aprendido.

Pero para llegar a un buen resultado es muy importante la capacitación docente en herramientas

informáticas (Hernández, 2000)<sup>5</sup>, que el docente se actualice en dimensiones didácticas, metodológicas y tecnológicas.

Se destaca la incorporación de estrategias de trabajo en equipo de profesionales de distintas disciplinas, como por ejemplo Ingenieros en Construcciones, Eléctricos, Mecánicos y Aeronáuticos, todos estos especializados en el área del Cálculo y Licenciados en Sistemas de Información especializados en el área de Informática Educativa e Ingeniería de Software.

Para la construcción del software, se seleccionará un ciclo de vida, pero el utilizar una metodología de desarrollo u otra depende de varios factores como el equipo de desarrollo, el tipo de proyecto, el cambio previsto desde un inicio de los requisitos, etc.; por lo que la selección de la misma no es una tarea fácil.

Las metodologías ágiles se han desarrollado enfocadas en dar soluciones óptimas al desarrollo de sistemas y varían dependiendo del sistema que se desea implementar, las herramientas, el enfoque y su equipo de trabajo como una contra respuesta a las metodologías tradicionales que caen en la extrema complejidad, su poca flexibilidad y dificultad de uso (Letelier P., Penadés M.C., 2006)<sup>6</sup>.

No existe una metodología que asegure el éxito de un proyecto de desarrollo de software, como también es difícil que una metodología se ajuste completamente a las características de un proyecto de software en particular. Toda metodología debe ser adaptada al contexto del proyecto (recursos

técnicos y humanos, tiempo de desarrollo, tipo de sistema, etc).

En el desarrollo del prototipo del aplicativo se utilizará una metodología ágil de desarrollo, adaptando las ya existentes según surjan las características del problema a resolver.

Se empleará Análisis, Diseño y Desarrollo Orientado a Objeto utilizando UML, Lenguaje de Modelado Unificado. Se realizará la selección de herramientas para el desarrollo del prototipo.

Para la selección de herramientas se efectuará un análisis de contexto en base a criterios de selección del grupo de trabajo siguiendo los objetivos de un rápido desarrollo, documentación básica y de calidad; flexibilidad en el proceso, facilidad en la comprensión y uso de las herramientas; de manera que de soporte y pueda contribuir de forma importante al aprovechamiento de las fortalezas del equipo.

### **Líneas de investigación y desarrollo**

El presente trabajo se enmarca dentro de las siguientes líneas de investigación consideradas prioritarias para la promoción de la actividad de investigación y desarrollo en la Facultad de Ciencia y Tecnología de la UADER:

- Tecnologías innovadoras en la producción y servicios.
- Generación de software aplicaciones y componentes.
- Problemática educativa y didáctica en Matemática, Física, Química, Biología y Educación Tecnológica.

### **Resultados y Objetivos**

El objetivo general de esta investigación es: Facilitar la enseñanza y el aprendizaje del Cálculo Diferencial e Integral a través del desarrollo de un Aplicativo Web. Siendo los objetivos específicos: determinar a partir del trabajo en grupo de los docentes de Cálculo Diferencial e Integral, las dificultades en la enseñanza y establecer prioridades en los temas planteados; elaborar los contenidos conceptuales, procedimentales y actitudinales del tema elegido, y su evaluación; determinar las herramientas informáticas a utilizar en el desarrollo del Aplicativo; determinar la metodología de desarrollo para la elaboración del Aplicativo, Análisis, diseño del prototipo estructural y del prototipo del aplicativo web; desarrollar del prototipo del aplicativo en dos versiones, una con acceso desde Internet con control de ingreso y otra instalable desde un ejecutable.

Los resultados que se han alcanzado hasta este momento de acuerdo a las actividades propuestas y realizadas son: determinación de las dificultades en la enseñanza del Cálculo Diferencial e Integral, lo que permitió establecer el tema sobre el cual se desarrollará el aplicativo; se ha comenzado a elaborar los contenidos conceptuales, procedimentales y actitudinales del tema elegido, así como las actividades a desarrollar en el Programa Descartes, con la consiguiente evaluación; se han seleccionado las Herramientas informáticas para el desarrollo del software; se estableció una metodología de desarrollo para la elaboración del aplicativo.

## **Formación de Recursos Humanos**

Por las características de este proyecto (PIDA), se promueve en primera instancia la formación del director y co-director en el área de investigación bajo la orientación del investigador asesor que los guiará para que el proyecto sea factible de concretar, formando un equipo de investigadores que pueda cumplir los objetivos fijados en el proyecto.

Los participantes del proyecto tienen trayectoria suficiente en el área del Calculo Diferencial e Integral y en el área de Ingeniería de Software e Investigación Educativa y durante el

transcurso del proyecto pueden aportar su experiencia para la realización del trabajo interdisciplinario.

La formación de becarios y estudiantes es sustancial para la consolidación de los futuros profesionales, permitiendo la capacitación en las tareas de investigación. Tres estudiantes de la carrera Licenciatura en Sistemas de Información aprobarán la cátedra Taller de Integración con el desarrollo del software producido en el proyecto de investigación.

## **Referencias**

1 CEPAL(2008). “La sociedad de la información en América Latina y el Caribe: Desarrollo de las tecnologías y tecnologías para el desarrollo”, Santiago de Chile. <http://www.oei.es/tic/cepal.pdf>

2 Martin, E. (2007). “El impacto de las TIC en el aprendizaje”, ponencia presentada en el [seminario internacional “Como las TIC transforman las escuelas”, Buenos Aires.](#)

3 Ausubel, D., Novak, J., Hanesian, H. (2000). “Psicología Educativa: un punto de vista cognocitivo” – Ed. Trillas – México – 2da ed.

4 Kutzler, B. (2003). “CAS as pedagogical tools for teaching and learning mathematics”. en Fey et al. (Eds.), Computer algebra systems in secondary school education.

5 Hernández, L. (2000). “Una vía transdisciplinar sobre las TICS para el desarrollo de habilidades profesionales generales, en cursos de posgrado semi-presenciales” –Tesis de doctorado, Universidad de la Habana.

6 Letelier, P., Penadés, M. C. (2006). “Metodologías ágiles para desarrollo de software: eXtreme Programming (XP)”. ISSN 1666-1680. Bs. As. Argentina.



# UN ABORDAJE DE LA ENSEÑANZA DE LA MATERIA INTELIGENCIA ARTIFICIAL DESDE LOS AMBIENTES DE APRENDIZAJE COLABORATIVOS EN LA CARRERA DE ISI UTN FRC

OLARIAGA, Sandra Monica (1), PAEZ Nancy del Valle (2)  
 CARRIZO, Blanca Rosa (3), ,  
 Dpto. Ingeniería en Sistemas de Información / Facultad Regional Córdoba  
 Universidad Tecnológica Nacional  
 Maestro Marcelo López s/n. Ciudad Universitaria.  
 Teléfono: 0351 – 4684317 - 4684215  
 solariaga@sistemas.frc.utn.edu.ar (1) / npaez@sistemas.frc.utn.edu.ar (2)  
 bcarrizo@tecnicatura.frc.utn.edu.ar (3)

## RESUMEN

Este trabajo pretende aplicar ambientes de aprendizaje colaborativos para la enseñanza de la asignatura Inteligencia Artificial, apoyándose en trabajos realizados anteriormente en las Universidades del Mundo, principalmente en España y Latinoamérica.

El objetivo de este proyecto es evaluar la aplicación de una herramienta que permita un abordaje de las clases de esta asignatura desde el entorno de ambientes colaborativos de aprendizaje a través de:

- Indagar acerca de los modelos de ambientes colaborativos de aprendizaje existentes.
- Realizar un análisis del material obtenido a fin de poder llegar a la selección de la herramienta adecuada.
- Seleccionar la herramienta de acuerdo al modelo colaborativo elegido para el abordaje de la enseñanza de la materia Inteligencia Artificial desde los ambientes colaborativos de aprendizaje.
- Implementar la herramienta seleccionada en la materia Inteligencia Artificial.
- Realizar la evaluación y documentación de los resultados obtenidos con la aplicación de la herramienta seleccionada
- Elaborar el informe general del proyecto.

**Palabras Clave:** Inteligencia Artificial, Ambientes colaborativos, Aprendizaje colaborativo, Tecnologías de Información y Comunicación, blended learning.

## CONTEXTO

Este trabajo de investigación está encuadrado en el seno del Dpto. Ingeniería en Sistemas de Información mediante la Res. N° 311/2011 del Secretario de Ciencia, Tecnología y Posgrado de la UTN con fecha de aprobación del 24/10/2011

Cabe aclarar que, los objetivos de la asignatura Inteligencia Artificial están normados por la Ordenanza 1150, que rige el diseño curricular de la carrera Ingeniería en Sistemas de Información, entre los cuales detallamos:

- Aplicar las metodologías de representación y resolución de problemas utilizadas en Ingeniería Artificial para ser empleadas en el abordaje de situaciones que se presentarán en la actividad profesional.
- Implementar Sistemas Inteligentes utilizando lenguajes y herramientas de Inteligencia Artificial.
- Conocer la aplicabilidad, el desarrollo y la arquitectura de los sistemas inteligentes artificiales.
- Profundizar en el conocimiento de agentes inteligentes y su diseño, los distintos tipos, los ambientes en donde deben desenvolverse y la aplicabilidad en distintas situaciones planteadas.
- Intervenir en el desarrollo de sistemas basados en conocimiento y sistemas expertos.

## 1. INTRODUCCIÓN

El uso de las nuevas tecnologías como apoyo al proceso de enseñanza y

aprendizaje, en nuestros tiempos, más que ser algo novedoso es necesario e imprescindible; sobre todo en aquellas áreas donde el uso de éstas, potenciaría significativamente dicho proceso.

En el caso particular de la asignatura Inteligencia Artificial, se vienen incorporando, progresivamente, herramientas que ayudan a la presentación de los diferentes contenidos curriculares.

En este contexto, las TIC desempeñan un rol estratégico como medio para que el estudiante cree su propio conocimiento, pero en una dinámica grupal y colaborativa, lo que actualmente se denomina ambiente colaborativo de aprendizaje.

En este tipo de espacio de aprendizaje, el estudiante explora, analiza, recaba, sintetiza, información creando su propio conocimiento y a la vez, pone a prueba sus conocimientos llevando a la práctica aquellos lineamientos que da la teoría.

Los contenidos que se desarrollan en la asignatura Inteligencia Artificial, tienen la característica de ser temas que se prestan a la búsqueda bibliográfica, ya sea en forma impresa o digital (en la web), y a la comparación y confrontación de dichas búsquedas, donde predomina el trabajo grupal, la interacción, la colaboración, etc. Haciendo una analogía entre éstas y las características que este modelo presenta, su implementación es totalmente viable.

En el ambiente de aprendizaje colaborativo, cada alumno asume su propio ritmo y potencialidades, impregnando la actividad de autonomía, pero cada uno comprende la necesidad de aportar lo mejor de sí al grupo, para lograr un resultado sinérgico; se logra así una relación de interdependencia y favorece los procesos individuales de crecimiento y desarrolla las relaciones interpersonales y la productividad. Además favorece la capacidad de resolver problemas de forma creativa, a partir de estrategias de negociación y mediación y la búsqueda cooperativa de alternativas.

## 2. ESTADO ACTUAL DEL CONOCIMIENTO DEL TEMA.

En la búsqueda de antecedentes del tema encontramos un trabajo de la Universidad Autónoma de Mexico con el título

“Exploración del aprendizaje de los estudiantes haciendo uso de ambientes colaborativos enseñando inteligencia artificial” de Gustavo de la Cruz Martinez y Fernando Gamboa Rodriguez. En este trabajo se hace un desarrollo de una herramienta cuyo objetivo es proveer un mundo virtual donde los estudiantes pongan en práctica los conocimientos teóricos vistos en clase. Como resultado comprobaron que el uso de una herramienta de este tipo ayuda al estudiante a entender de manera más adecuada los conceptos teóricos ya que le sirve para percibir las ventajas y limitaciones de los mismos, mediante la recolección de información acerca de cómo el estudiante aplica los conocimientos teóricos impartidos.

## 3. LINEAS DE INVESTIGACIÓN Y DESARROLLO.

El proyecto se inscribe dentro de los lineamientos de investigación en innovaciones curriculares en Educación Superior Universitaria dentro del área de Sistemas e interfaces adaptables en la intervención humano-computadora.

Se trata de una investigación descriptiva porque busca profundizar la comprensión del problema y está basada en la observación. La preocupación estará centrada en indagar los hechos o fenómenos educativos en su “realidad natural”. Se aplicarán las técnicas de la investigación acción. Para ello se seguirán los siguientes pasos:

El **primer paso** consiste en la identificación de un problema, en este caso como enriquecer el proceso de enseñanza y de aprendizaje de la materia Inteligencia Artificial, surge entonces el abordaje desde los ambientes colaborativos de aprendizaje.

El **segundo paso** es la investigación bibliográfica y por internet acerca de los diferentes modelos de los ambientes colaborativos de aprendizaje y una exploración acerca de las herramientas informáticas existentes.

El **tercer paso** es la selección de una herramienta adecuada.

El **cuarto paso** es el planteo de una hipótesis: "La aplicación de alguna herramienta desde los ambientes colaborativos de aprendizaje, que permitiría

el abordaje de la enseñanza de la materia Inteligencia Artificial con un mayor aprovechamiento del potencial temático que se trabaja, adaptando tanto, las estrategias de enseñanza de los docentes como las de aprendizaje de los alumnos a nuevos ambientes de trabajo". Esta hipótesis se plantea dado que como docentes de la materia Inteligencia Artificial vemos que la enseñanza de los contenidos de la materia presentan un potencial muy interesante a ser desarrollado si son abordados desde el punto de vista del ambiente de aprendizaje colaborativo

El **quinto paso** es la puesta en marcha de una propuesta para la comprobación de la hipótesis planteada, en donde en función de las dimensiones que nos interesan, desarrollaremos un modelo de enseñanza y de aprendizaje a ser implementado con la ayuda de la herramienta seleccionada.

El **sexto paso** es la implementación de un instrumento de recogida de datos en base a las dimensiones ya establecidas.

El **septimo paso** es el análisis de los datos obtenidos en función de las categorías y dimensiones establecidas.

El **octavo paso** es la documentación de los resultados obtenidos en base a los registros realizados.

#### 4. ESTADO ACTUAL DEL AVANCE DEL TRABAJO

Nuestro trabajo se viene desarrollando a través de varias tareas, las cuales algunas de ellas se han realizado simultáneamente, dando el punto de partida a otras; a saber:

- Elaborando un instrumento de recogida de datos, como es la encuesta, a fin de determinar que softwares actualmente usan nuestros alumnos en relación a su actividad académica y en su vida diaria
- Recabando información acerca de las plataformas que vamos a evaluar a los efectos de cual de ellas se adaptaría mejor a nuestro trabajo, entre ellas moodle, edmodo, Dokeos mind.
- Analizando las bondades que ofrecen las aplicaciones de cloud computing y

la compatibilidad con nuestro proyecto.

- Utilizando distintas herramientas que sirvan para análisis comparativo a los fines de poder determinar la conveniencia del uso de alguna de las plataformas o bien decimos por utilizar la cloud computing.

En este momento estamos:

- Diseñando un sitio web que contiene las distintas actividades que el alumno podrá usar en modo interactivo, tanto para realizar tareas solicitadas por el profesor, como para afianzar a través de la práctica, aquellos conceptos vistos teóricamente.
- Analizando y probando distintos utilitarios a incluir en el sitio web que puedan ser usados por el alumno como herramientas didácticas.
- Compatibilizando las prácticas pedagógicas tradicionales con el nuevo enfoque colaborativo vía Tic.

Los análisis comparativos de las distintas aplicaciones a usar, es acompañado por pruebas piloto en la misma cátedra de Inteligencia Artificial, así probamos con la plataforma Ecaths para el uso de foros, encuestas, publicación de material multimedia, etc., el cual fue accedido por los alumnos. Las clases son semanales, intercalando una clase teórica y una clase práctica, por lo que en el lapso que transcurría de una clase a otra, la comunicación con los alumnos fue a través del aula virtual, a fin de que los estudiantes contaran con actividades que pudieran llevar a cabo en base al material bibliográfico digital, consultar o plantear dudas, inquietudes, como así también sugerir o proponer bibliografía, videos, sitios web, etc., relacionados con el tema en cuestion. Como conclusión de esta experiencia piloto podemos decir que la plataforma Ecaths es muy limitada, ya que no permite por ejemplo subir a la misma archivos compactados, lo cual es muy importante en esta cátedra ya que necesitamos valernos de intérpretes y simuladores para varios temas de la currícula.

En este momento hemos estado trabajando en forma piloto con la plataforma Moodle y estamos en vías de desarrollar un software dentro del entorno de Java que contenga ejercicios prácticos para la primera unidad de la materia los cuales serán implementados durante las clases de este año, haciendo énfasis en que su aplicación será dentro del ámbito del trabajo colaborativo.

## 5. RESULTADOS Y OBJETIVOS.

Los docentes son los actores que mayores cambios pueden generar en las universidades ya que en ellos recae la responsabilidad de la formación y la orientación de los alumnos.

Los resultados obtenidos serán presentados ante el jefe de cátedra de la materia Inteligencia Artificial, para que evalúe la viabilidad de implementar dicho proyecto en todos los cursos que integran dicha cátedra, lo cual implicaría una modificación de la visión y de la forma de trabajo de los docentes y alumnos, es decir un cambio en las prácticas áulicas, a través de la aplicación de una herramienta que aprovecha las ventajas del trabajo grupal y colaborativo, potenciando el proceso cognitivo de aprendizaje tanto en forma individual como grupal.

Este proyecto es, por lo tanto, un aporte para la difusión de nuevas formas, aproximaciones y aspectos particulares en la implementación de las nuevas tecnologías educativas, como una manera de incorporar el trabajo en ambientes colaborativos dentro del ámbito universitario.

## 6. FORMACIÓN DE RECURSOS HUMANOS

Todos los integrantes de este PID son Especialistas en Docencia Universitaria, siendo la Directora Magister.

Con este proyecto se continuará con el proceso de formación en investigación de todos los integrantes, ya que los mismos han participado en proyectos anteriores, con temáticas referidas al ámbito de la educación y de competencias.

Este estudio en particular contribuirá en gran medida a ampliar el conocimiento acerca de

la aplicación de ambientes colaborativos de aprendizaje, de manera de que los resultados obtenidos puedan ser difundidos a fin de que pueda llegar a ser el punto de partida de actividades que favorezcan el mejoramiento de los procesos de enseñanza y de aprendizaje en el aula universitaria.

Esta prevista la incorporación de becarios alumnos, algunos de los cuales ya han participado en proyectos anteriores, de la carrera de Ingeniería en Sistemas de Información.

Además se preve que una de las integrantes del grupo complete su trabajo final en la especialización en Tecnologías Multimedia para Desarrollos Educativos, dentro del marco de esta investigación.

## 7. BIBLIOGRAFÍA

- [1] Aprendizaje colaborativo presencial, aprendizaje colaborativo mediado por computador e interacción. Aclaración aportes y evidencias. Gloria M. Londoño. Revista Q. Revista Educación, Comunicación y Tecnología. Vol. 2 Nro. 4 Enero – Junio 2008 Medellín Colombia ISSN 1909-2814.
- [2] Estrategias docentes para un aprendizaje significativo. Una interpretación constructivista. Díaz Barriga México, McGraw-Hill. (1999).
- [3] The Use of Agents Techniques on Intelligent tutoring System. Giraffa, L. (1998) RIBIE –98. Brazil.
- [4] Entre el trabajo colaborativo y el aprendizaje colaborativo Revista Iberoamericana de Educación. Lucero, M. M: ISSN 1681-5653 (2003).
- [5] El aprendizaje Cooperativo en el Aula. Johnson, D. y Otros. Ed. Paidós Educador. Bs.As. 1999.
- [6] Los siete saberes necesarios para la educación del futuro. Organización de las Naciones Unidas, para la Educación, la Ciencia y la Cultura. Morin, E. Traducción de Mercedes Vallejo-Gómez, Medellín. Colombia. (2000).
- [7] Computer Supported Collaborative Learning (CSCL)", en Computers and Education. Silverman, B. (1995)
- [8] La zona de desarrollo próximo y su colaboración en la práctica de aula. Tudge, Vigotsky. Nueva York, Universidad de Cambridge. (1994)



**[9]** Tecnología y Comunicación Educativas Año 22-23 Nro. 47-48 Julio 2008 – Junio 2009. Publicación integrante de la Red Iberoamericana de Revistas de Comunicación y Cultura y a la Red de Publicaciones de Innovación Educativa y Educación a Distancia.

**[10]** Una revisión de desarrollos inteligentes para aprendizaje colaborativo soportado por computadora. Rossana Costaguta. Dpto de Informática, Facultad de Ciencias Exactas y Tecnológicas. Universidad Nacional de Santiago del Estero. Argentina. Revista Ingeniería Informática Edición N° 13 Noviembre 2006.

## TECNOLOGÍAS DE LA INFORMACIÓN Y LA COMUNICACIÓN APLICADAS A LA FORMACIÓN DE PROFESIONALES EN TURISMO RURAL

Zulma Cataldi<sup>1</sup>, Enrique Bombelli<sup>2</sup>, Sandra Fernández<sup>2</sup>, Valeria di Pierro<sup>2</sup>, María Belén Britos<sup>2</sup>  
y Rodolfo Bertoncello<sup>3</sup>.

<sup>1</sup>Facultad de Ingeniería. Universidad de Buenos Aires (FI-UBA).

<sup>2</sup>Facultad de Agronomía Universidad de Buenos Aires (FA-UBA).

<sup>3</sup>Facultad de Filosofía y Letras. Universidad de Buenos Aires (FILO-UBA).

liema@fi.uba.ar

### RESUMEN

A partir del primer curso de "Técnicas de Diseño y Organización de Actividades", perteneciente a la carrera Técnica en Turismo Rural de la Facultad de Agronomía/UBA, que se dictó en 2009, se detectó la necesidad de brindar conocimientos a los estudiantes sobre instrumentos comunicacionales de promoción y difusión de bajo costo y masivos.

Se decidió que la enseñanza de uso de herramientas para el desarrollo de weblogs reunía estas condiciones. Entonces se incluyó el módulo Tecnologías de la Información y la Comunicación (TIC) aplicadas al Turismo Rural. Luego de perfeccionar esta temática dentro de los contenidos del curso, se presentaron varios blogs, algunos de ellos, muy exitosos y creativos, muchos de los cuales se mantienen aún vigentes. Dicho módulo permitió articular el trabajo en el aula con comunidades rurales, generando al mismo tiempo, un interés por las localidades, traccionada a su vez por esta novedosa forma de comunicación turística. El número de visitantes, la cantidad de entradas, la diversidad informativa y la interacción con otras redes sociales, hizo de algunos casos ejemplos virtuosos para mencionar. En este trabajo se trata de destacar la incorporación de una herramienta innovadora de la Web 2.0 en el trabajo educativo y su impacto en las comunidades.

**Palabras clave:** Turismo Rural, Web 2.0, weblog.

### CONTEXTO

Esta comunicación abarca la dimensión educativa, en relación a las necesidades de formación en TIC que los alumnos y el área del Turismo Rural demandan y requiere, respectivamente. Dicha línea de investigación, se enmarca en el proyecto UBACyT, acreditado y

correspondiente a la programación científica 2012-2015, titulado *El uso de Tic como recursos tecnológicos de competitividad aplicados a emprendimientos de turismo rural* (20020110100223).

### INTRODUCCIÓN

Teniendo en cuenta que para desempeñarse en sociedad, el futuro profesional debe superar la instancia de acumulador o reproductor de conocimientos para ser un usuario inteligente y crítico de la información, y que el turismo y la información están estrechamente vinculados [1], al punto que algunos expertos le suelen denominar "negocio de la información" [2], es preciso aprender a administrar, obtener, procesar y comunicar información y convertirla en conocimiento. Necesita ser consciente además, de sus capacidades intelectuales, emocionales o físicas y disponer del sentido de su competencia personal. Es decir, debe valerse de sus habilidades para iniciarse en el aprendizaje y continuar aprendiendo de manera cada vez más eficaz y autónoma a lo largo de toda la vida, de acuerdo con sus necesidades y objetivos.

Como señalan algunos autores [3], "...estamos viviendo un período de transición y cambio en el sistema de educación superior, en el cual la sociedad de la información ha condicionado el proceso a las exigencias que ésta demanda a mujeres y hombres del presente siglo. Asumir estas exigencias y transformarlas en atributos diferenciadores supone la permanencia de las universidades en el emergente mercado del conocimiento, donde podrán mantenerse vigentes y competitivas en el corto, mediano y largo plazo. Por el contrario, no adaptar las estructuras universitarias al contexto global supone un deterioro del posicionamiento institucional imposible de remediar".

Uno de los desafíos para las TIC en la educación universitaria, es disminuir la brecha de la heterogeneidad digital y las desigualdades en las competencias profesionales y, al mismo tiempo, aprovechar sus ventajas para construir un tejido productivo y social rural más igualitario. Estos elementos permiten detectar un escenario donde un mayor y mejor manejo y uso de las TIC por parte de los estudiantes no se relaciona solo con cuestiones de accesibilidad, sino más bien con la capacidad de elegir, adaptar, incorporar y aprovechar los recursos más adecuados en cada circunstancia, lo que depende en gran medida de los conocimientos adquiridos sobre las tecnologías disponibles.

En este contexto, resulta indispensable formular programas integrados que permitan encarar de forma sistémica las iniciativas orientadas a la incorporación de TIC en las currículas, con el objeto que estas tecnologías puedan ser útiles para alcanzar no solo el incremento de la productividad de las firmas donde los profesionales se desempeñan, sino además favorecer su visibilidad, una menor heterogeneidad en la estructura turística rural y una mayor inclusión social.

La sociedad del conocimiento y de la creatividad, es una sociedad con sobreabundancia de datos y hechos los cuales deben ser ordenados, agrupados, analizados e interpretados y sintetizados. El aumento de la información cambia la percepción del modo de actuar sobre el mundo, el ritmo de nuestros archivos de imágenes, de pensar y sintetizar, de prever las consecuencias de nuestra acción, así como la interrelación de grandes fuerzas causales que generan los sucesos y la actitud de sorpresa del observador mientras no construya los conocimientos previos adecuados y debidamente fortalecidos para “ordenar” esa complejidad [4].

Es en esta cuestión donde la enseñanza superior puede y debe dar respuestas orientadas a generar habilidades para seleccionar, descartar, reagrupar y ordenar con criterio la información, para transformarla en conocimiento. El espacio telemático no es presencial sino representacional, no es proximal sino distal, no es sincrónico sino asincrónico y no se basa en recintos espaciales con interior, frontera y exterior, sino que depende de redes electrónicas cuyos nodos de interacción están dispersos interplanetariamente [5].

El producto final de la telemática es que cada vez más todos estarán afectados por todos y en el aspecto cultural, dicha telemática permitirá que las identidades se resinteticen permanentemente

con las otras identidades culturales e ingresen a un vértigo de transfiguraciones permanentes. Simultáneamente, es este mismo fenómeno el que está definiendo el grado de modernidad de los países y/o de las personas, por la capacidad que tienen de incorporar tecnología y valor intelectual agregado en la producción de fenómenos y bienes [6].

Liderar el capital intelectual, el conocimiento y la tecnología, tiene que ver con formas efectivas de hacer explícitas, localizar, organizar, transferir y usar las ideas, información, experiencia y creatividad, acumuladas en la organización. Tiene que ver con implementar procesos para la distribución de conocimientos entre grupos de trabajo interdisciplinario sobre la generación de nuevos productos, diseñando estrategias para compartir el conocimiento corporativo, transfiriéndola a la organización a través de instancias para llevar a cabo conversaciones estratégicas. Se trata de que el capital intelectual se transforme en un bien organizacional sólido y duradero a través de una forma de introducción adecuada, haciéndolo accesible y reproducible para todos [2].

## LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Esta comunicación enfatiza las necesidades de formación en TIC que los alumnos y el área del Turismo Rural demandan y requiere, respectivamente, las que a su vez se retroalimentan entre sí. Dicha línea de investigación, se enmarca en el proyecto UBACyT, acreditado y subsidiado, correspondiente a la programación científica 2012-2015, titulado *El uso de Tic como recursos tecnológicos de competitividad aplicados a emprendimientos de turismo rural* (20020110100223). Este último es continuación de otro titulado *TIC y Turismo Rural. Utilización de nuevas tecnologías en ámbitos rurales vinculados a la prestación de servicios turísticos circundantes a la Ciudad Autónoma de Buenos Aires*, acreditado por resolución de CS 1004/10, financiado y correspondiente a la programación científica 2010/2012.

Ambos constituyen y han constituido proyectos, que intentan dar cuenta de las vinculaciones existentes entre las Tecnologías de la Información y la Comunicación (TIC) y el Turismo Rural, en el marco de las continuas transformaciones que experimenta el sector. Participan y han participado en la constitución de

los mismos, tres Facultades de la UBA: Ingeniería, Agronomía y Filosofía y Letras.

**Grado de avance:** Los avances hasta aquí presentados, corresponden al uso de instrumentos comunicacionales de bajo costo y masivos (weblogs) para la promoción y difusión de actividades en Turismo Rural. Desde esta perspectiva, la temática curricular está focalizada al logro de las competencias que se desean desarrollar en los futuros profesionales.

Los objetivos de este trabajo se orientan a demostrar la potencia que las herramientas de la Web 2.0 tienen como enlace entre el aula y las comunidades, cuando se las aplica en el entorno formativo a partir de competencias profesionales. La implementación requirió la reconfiguración del rol docente y un cierto grado de adaptación de los estudiantes a la herramienta. En este sentido, la infraestructura tecnológica y el grado de alfabetización digital, tanto de alumnos como docentes, resultaron adecuados para su implementación. El trabajo al interior de los grupos, siempre con la guía docente, facilitó el aprendizaje colaborativo, fortaleciendo las interacciones alumnos/alumno y docente/alumno. La característica más sobresaliente de la implementación, es el desarrollo de una cultura que se extiende más allá de la propia asignatura. Se destacan algunos de los sitios más representativos (Gesell Distinta, Mis Parques Nacionales, Pueblos con memoria) desarrollados por los propios estudiantes, los cuales permitieron articular el trabajo en aula con diversas comunidades rurales, generando un interés por la localidad con un enfoque menos tradicional.

Desde el punto de vista del alumnado, el uso de las TIC en general y del blog en particular, es una atractivo que activa el desarrollo de las clases, fomentando la participación de una manera lúdica, activa y colaborativa, llevándoles a concebir el conocimiento de una forma más aplicada y concreta que si siguiesen una dinámica tradicional. Es por ello que desde la perspectiva docente, como desde la visión del alumnado el blog se convierte en una herramienta fundamental en el proceso de integración y normalización de las TIC en el aula universitaria, llevándose a la práctica por medio de la colaboración, la participación y la entrada en escena de herramientas 2.0.

## RESULTADOS Y OBJETIVOS

Algunas de las preguntas que surgieron en el primer proyecto y continúan siendo rectoras de la

investigación son: *¿Utilizan TIC los actores vinculados a emprendimientos de Turismo Rural?, ¿Cómo juega la disponibilidad o accesibilidad de TICs en el medio rural?, ¿Las consideran una herramienta para ganar competitividad?, ¿Quiénes específicamente hacen uso de ellas?, ¿Cómo las utilizan?, ¿De qué tipo de TICs disponen?, ¿Cómo son los resultados?, ¿Qué diferencias existen en el uso que realizan distintos actores?, ¿Cuáles son los beneficios o ventajas de su utilización a nivel de la competitividad?, ¿Cómo transformar las TICs en herramientas verdaderamente competitivas?*

Los objetivos propios del proyecto actual responden al análisis del uso de TIC en emprendimientos de Turismo Rural y su vinculación con el desempeño competitivo (lo que a su vez permite la actualización de necesidades de formación en TIC por parte de estudiantes de Turismo Rural), a partir de la elaboración de una matriz de recursos TIC, agrupando los casos observados de acuerdo a combinaciones de Tipo y Grado de TIC con el período de adopción, en los emprendimientos relevados. Consecuentemente, se vuelve una necesidad relacionar cada grupo de casos logrado en la fase anterior, con indicadores de competitividad de los emprendimientos.

**Metodología de la investigación:** La singularidad del mismo radica en los escasos antecedentes sobre el tema en nuestro país. Si bien existen trabajos que han aplicado enfoques similares, dicha investigación se justifica en el medio rural y más aún en el turismo, debido al alto potencial de impacto que las TIC ejercen sobre el desarrollo del sector. Por otra parte, la confluencia del andamiaje conceptual que integra TIC, competitividad y turismo rural, resulta un aporte original y promisorio para comprender mejor la situación competitiva de los emprendimientos dedicados a esta actividad. Se resalta también la jerarquía del relevamiento en lo relativo a los aportes procedentes del entorno regional (meso) y su interacción con la aplicación de TIC, que resultan en un determinado desempeño competitivo. Otra fortaleza del proyecto consiste en su vinculación directa con otro grupo de investigación relacionado con el tema *Turismo y Territorio. Política y cultura en la definición de lugares y atractivos turísticos de la Argentina* (UBACyT 20020100100230). A través del mismo se da lugar al enfoque interdisciplinario, integrando otros campos del conocimiento y enriqueciendo el intercambio de ideas y experiencias de abordaje científico entre investigadores.



**Resultados preliminares:** Desde la creación de la carrera Técnica en Turismo Rural/FAUBA, han pasado ya tres cohortes (2010, 2011, 2012) por la asignatura Técnicas de Diseño y Organización de Actividades, con una cantidad promedio de 30 alumnos/cohorte. Ello determinó la producción grupal (3 a 4 integrantes) de 34 weblogs. A continuación se describen las características sobresalientes de los tres más representativos de la última cohorte (2012) y su impacto territorial [5] puede apreciarse directamente a través del sitio respectivo.

#### Gesell Distinta

<http://www.geselldistinta.blogspot.com.ar/>

La idea inspiradora de este sitio, fue el conjunto de experiencias propias de los estudiantes en el lugar. La temática constituyó una excelente elección bajo el lema “Gesell Distinta”, sobre todo, teniendo en cuenta que dicha localidad es una oferta turística de sol y playa, aportando de manera innovadora, el un tipo de turismo no convencional. Han realizado excelentes estrategias de difusión, ya sea por vinculación del sitio con redes sociales como “Facebook” y transmisión interna vía correo electrónico a sus contactos. Esto dio como resultado un muy buen posicionamiento en Internet. En cuanto a las herramientas que utilizaron fueron muy variadas, incluso con elaboración de videos propios, editando audio, imagen y sonido. Otra herramienta interesante y muy en boga en la actualidad, fue la elaboración de Código QR, destacándose como pioneros dentro de la respectiva cohorte. También han creado un subproducto (sitio dentro de sitio) al que llamaron “Tierras del Tuyú”, lo que muestra una visión dinámica que se renueva permanentemente en busca de nuevos visitantes potenciales.

#### Mis Parques Nacionales

<http://misparquesnacionales.blogspot.com.ar/>

Lo interesante de la temática elegida, fue la existencia de muy pocos sitios que traten a los Parques Nacionales en su conjunto dentro de un mismo sitio, motivo por el cual fue muy visitado. El grupo de alumnos ha tenido una muy buena organización del trabajo, autodenominándose “Triple Horizonte”, ya que eran 3 participantes, cada uno de los cuales se destacaba en relación a algún aspecto de la elaboración, investigación y redacción del blog. Han elegido una muy buena URL, de esta manera el blog próximamente estará dentro de los primeros puestos en los buscadores de Internet. También se destaca la excelente interpretación de las estadísticas proporcionadas por Google Analytics. Las propuestas de mejoras a futuro, tales como viajar

a los parques y darle al sitio un estilo “blog viajero”, muestran la continuidad de de los sitios más allá del curso.

#### Pueblos con Memoria

<http://www.pueblosymemoria.blogspot.com.ar/>

La idea novedosa de promocionar y dar a conocer zonas postergadas de nuestro país, constituyó el principal valor del sitio. Asimismo rescata la identidad cultural de la región, a partir de la propia experiencia del estudiante. Posee una estética bien pensada y acorde a la temática, con una excelente exposición de imágenes. Este aporte profesional, se debe al desempeño del alumno en el área del cine y la fotografía.

#### **Producción 2012/2013**

- Cataldi, Z.; Bombelli, E.; Fernandez, S.; Di Pierro, V.; Barberis, G. y Bertoncello; R. (2012). *TICs y Turismo Rural: Las redes sociales de emprendedores y viajeros como oferentes y demandantes de servicios turísticos*. WICC 2012. 26 y 27 de abril. UN Misiones.
- Bombelli, E., Mella, A.; Byorkman, A., Barberis, G. y Cataldi, Z. (2012). *Modelo b-learning para la Enseñanza de la Informática en Ingeniería*. II JEIN 2 y 3 de agosto. San Nicolás, Vol. 1 pag. 33-39.
- Cataldi, Z.; Di Pierro; V.; Bombelli, E.; Fernández; S. Barberis; G.; Bertoncello, R.; Mansilla Maidana; D. y Guzmán, J. (2012). *Conocimiento en red de las TIC para su uso como recursos de competitividad. El caso del turismo rural*. Edutec 2012. Palma Gran Canaria. 9-11 noviembre.
- Bombelli, E.; Fernández, S.; Di Pierro, V.; Barberis, G.; Bertoncello, R. y Cataldi, Z. (2012). *Formación en TIC en carreras de Turismo Rural*. III Jornadas nacionales I Jornadas latinoamericanas de investigadores/as en formación en educación instituto de investigaciones en ciencias de la educación (IICE) 3 y 4 de Diciembre.
- Bombelli, E., Cataldi, Z, y Barberis, G. (2012). *Aprendizaje visual interactivo para formación docente en informática*. II Congreso Metropolitano de Formación Docente. 5, 6 y 7 de diciembre, Facultad de Filosofía y Letras Universidad de Buenos Aires.
- Bombelli, E.; Mella, A.; Byorkman, A.; Barberis, G. y Cataldi, Z. (2013). *Nuevos modelos educativos para la enseñanza de la informática en el nivel superior*. IV Jornadas de Campus Virtuales. Palma de Mallorca 14, y 15 de febrero.

- Cataldi, Z; Fernández, S.; Di Pierro, V.; Bombelli, E.; Barberis, G.; Bertoncello, R.; Dahiana Mansilla Maidana, D. y Guzmán, J. (2013). *Avances de la Investigación en TICs y Turismo Rural*. ICECE 2013. VIII International Conference on Engineering and Computer Education, Luanda, Angola. Marzo 3-6.

## FORMACION DE RECURSOS HUMANOS

La Lic. Di Pierro realizará su Maestría en Docencia Universitaria de la UBA durante 2013 y la estudiante María Belén Britos, de la Tecnicatura en Turismo Rural está próxima a graduarse e inscribirse en un Ciclo de Licenciatura en Turismo por la Universidad de Quilmes. Se prevé la incorporación de dos estudiantes más que estén interesados en formarse en el área de la investigación en Turismo Rural.

## REFERENCIAS

- [1]. Galí, N. y Majó, J. (2002). *Internet en la información turística*. IV Congreso Nacional de Turismo y Tecnologías de la Información y las Comunicaciones. Pp. 397-409.
- [2]. Schertler, W. (1998). *Virtual Enterprises in Tourism: Folklore and Facts: Conceptual Challenges for Academic Research*. En Buhalis, D.; Tjoa, A. M. y Jafari, J. (eds.): *Information and Communication Technologies in Tourism*, Springer-Verlag.
- [3]. Valenzuela Gárate, J.; Palant Bronn, M.; Zúñiga, M.E.; Iriarte Palma, P.; Rojas Muñoz, A.; y Hormazabal, J. (2002). *Hacia la una universidad global: la inserción de las tecnologías de información y comunicación en la educación superior*. Alfagrama. Santiago de Chile. Pág. 150.
- [4]. Yanes Guzmán, J. (2009). *Las TIC y la Crisis de la Educación. Algunas claves para su comprensión*. Biblioteca Digital Virtual Educa.  
<http://www.virtualeduca.org/documentos/yanez.pdf>
- [5]. Echeverría, J. (1998). *La escuela distal*. Comunicar, 10, pp 27-31. Madrid.
- [6]. Hopenhayn, M. (1994). *Ni apocalípticos ni integrados: aventuras de la modernidad en América Latina*. Fondo de Cultura Económica. Chile. Pág. 281.
- [7]. Savy, M. (2000). *Techniques d'information et de communication (TIC) et territoire*. En Brousseau, E.; Rallet, A. *Technologies de l'information et de la communication et performances économiques*. ATOM. Centre d'analyse théorique des organisations et des marchés. Université de Paris. PP 309-350.  
[http://www.brousseau.info/pdf/tic\\_plan/TICOPE Chap4.PDF](http://www.brousseau.info/pdf/tic_plan/TICOPE Chap4.PDF).

## TICs aplicadas a problemas de Gobierno Electrónico y de E-Learning

Depetris B. <sup>(1)</sup>, De Giusti A. <sup>(1)(2)</sup>, Feierherd G. <sup>(1)</sup>, Lanzarini L. <sup>(1)(2)</sup>, Sanz C. <sup>(2)</sup>, Aguil Mallea D. <sup>(1)</sup>

(1) Instituto de Desarrollo Económico e Innovación – Universidad Nacional de Tierra del Fuego, Antártida e Islas del Atlántico Sur  
Hipólito Irigoyen 880 – CP 9410 – Ushuaia – Tierra del Fuego – Argentina TE 54 2901 440853

(2) III LIDI – Facultad de Informática – Universidad Nacional de La Plata  
Calle 50 y 120 – CP 1400 – La Plata – Buenos Aires - Argentina – TE 54 221 4227707

[bdepetris@untdf.edu.ar](mailto:bdepetris@untdf.edu.ar); [degiusti@lidi.info.unlp.edu.ar](mailto:degiusti@lidi.info.unlp.edu.ar); [gfeierherd@untdf.edu.ar](mailto:gfeierherd@untdf.edu.ar);  
[llanzarini@lidi.info.unlp.edu.ar](mailto:llanzarini@lidi.info.unlp.edu.ar); [csanz@lidi.info.unlp.edu.ar](mailto:csanz@lidi.info.unlp.edu.ar); [daguil@untdf.edu.ar](mailto:daguil@untdf.edu.ar)

### Resumen

Desde mediados del siglo XX se ha acelerado el proceso de conformación de una nueva sociedad. Los profundos cambios sociales, económicos y culturales que lo caracterizan son, en gran medida, consecuencia del entorno tecnológico que la soporta y la envuelve [1]. La cara más visible de este entorno tecnológico son las Tecnologías de la Información y la Comunicación (TICs), cuyo uso creciente permite, entre otros beneficios, mejorar la prestación de servicios al ciudadano: E-Government, E-Learning, E-Health, son algunos ejemplos que evidencian esta situación. No obstante, el éxito de estas iniciativas depende de la existencia de un E-Citizen, es decir, un ciudadano en condiciones de utilizar de modo significativo las tecnologías que se emplean para mediar la prestación de los distintos servicios.

El proyecto trabaja sobre dos líneas principales de las aplicaciones de TICs: una en temas de Gobierno Electrónico y la otra en temas de Educación mediada por Tecnología (E-Learning).

Se han seleccionado aspectos particulares de estas dos líneas que son de interés

actual, y para los cuales se cuenta con recursos humanos e infraestructura para sustentar el desarrollo del proyecto.

### Palabras claves

E-Government, E-Democracy, E-Citizen, E-Learning,

### Contexto

Las líneas de investigación descriptas están insertas en el proyecto de investigación “TICs aplicadas a problemas de Gobierno Electrónico y de E-Learning”, que se desarrolla en el Instituto de Desarrollo Económico e Innovación de la UNTDF.

El proyecto fue avalado en todas las instancias previstas en la UNPSJB para luego, al ser transferida la sede académica donde debía desarrollarse a la nueva Universidad, obtener el aval de la Secretaria Académica y de Investigación de la UNTDF. El proceso de transferencia a la nueva institución ha provocado que el inicio real del proyecto se pospusiera hasta principios de este año.

Su financiamiento será operado por la UNTDF.

El proyecto se lleva a cabo en estrecha colaboración con la Facultad de Informática de la Universidad Nacional de La Plata, en particular con los grupos de investigación dedicados a estas temáticas constituidos en el Instituto de Investigación en Informática LIDI (III-LIDI), con quien la ex Sede Ushuaia de la UNPSJB firmó oportunamente un convenio de colaboración.

## Introducción

Las TICs, *“más que medios de comunicación son mediadores culturales que hacen posible la emergencia de prácticas sociales a partir de una nueva forma de comunicarse”* [2]. Es por ello que en la actualidad constituyen uno de los ejes del desarrollo humano [3]. En este proyecto nos enfocamos en dos grandes temáticas a nuestro entender fuertemente vinculadas: Gobierno Electrónico e E-Learning.

### 1. E-government

E – Government es el uso de las TICs para facilitar la operación de gobierno y la distribución de la información y los servicios del mismo. Su objetivo es hacerlo más accesible, efectivo, transparente y responsable [4]. Para lograrlo se recurre a aplicaciones que mejoran la gestión de la información de interés social.

Estas iniciativas contemplan varios ámbitos de aplicación: Government to Government, Government to Business y Government to Citizens [5]. No obstante puede asegurarse que en todos los casos, pero particularmente en el último, el éxito depende de que el concepto clásico (y restringido) de E-Government, se complemente con el de E-Citizen [6], es decir, con ciudadanos preparados para interactuar con el Estado utilizando estas

tecnologías. Se desprende entonces que, a fin de cumplir con los requisitos de igualdad propios de los sistemas democráticos en los que estas experiencias se están llevando a cabo, es imprescindible disminuir la denominada brecha digital. Este concepto, originalmente reducido a un aspecto estrictamente económico (tener o no tener acceso a los dispositivos de la tecnología), incluye hoy la posibilidad de hacer un uso apropiado de esos dispositivos. En referencia a esto último, las Universidades, particularmente las públicas, tienen la obligación de contribuir al logro de este objetivo.

Por otra parte, y entre otras aplicaciones, el Estado puede utilizar las TICs para integrar información (los casos típicos son las fichas de pacientes de Hospitales y los contenidos pedagógicos en Escuelas) y facilitar el proceso de toma de decisiones.

Por su criticidad todos estos procesos requieren una cuidadosa auditoría, que asegure la calidad y la seguridad de los sistemas y los datos, particularmente los relacionados con información sensible.

En síntesis, el E-Government permite abrir varias líneas de Investigación y Desarrollo, asociadas con la integración de las TICs en la sociedad, y que requieren un esfuerzo importante en la formación de recursos humanos.

### 2. E-Learning

La incorporación de la TICs al espacio educativo plantea de inicio la posibilidad de desarrollar importantes cambios en los procesos educativos [7].

Entre otros ha dado lugar a modalidades de Educación a Distancia e híbridas (Extended y Blended Learning), que tienen particular interés en nuestros días, tanto en los distintos niveles de educación



formal como en la educación continua y para todos.

Como consecuencia de ello nos encontramos frente a nuevos escenarios educativos, en los que elementos como la comunicación y los materiales de estudio digitales han incrementado su importancia.

En este contexto los materiales de estudio multimediales e hipermediales se convierten a su vez en objeto de estudio, no sólo por su capacidad para complementar a los materiales impresos si no porque permiten atender a los diferentes estilos cognitivos de los alumnos [8]. Una tendencia actual es el diseño de objetos de aprendizaje y su almacenamiento posterior en repositorios de objetos de aprendizaje que permiten compartir estos recursos entre docentes de diferentes instituciones. El conocimiento almacenado en estos repositorios debe representarse y recuperarse de modo “inteligente”, tanto por su sistematización ordenada como por el acortamiento de los tiempos de búsqueda. Por ello se produce la asociación de sistemas inteligentes a los entornos de aprendizaje y a los repositorios vinculados.

Otro aspecto importante en el área de materiales educativos digitales es el diseño, implementación y uso de software educativo. Este ofrece al alumno la posibilidad de experimentar y descubrir determinadas relaciones y aspectos funcionales, facilitando los procesos de enseñanza y de aprendizaje

Los recursos de simulación por computadora también proveen para el aprendizaje espacios de práctica para representar distintos aspectos de la realidad, indicaciones para actuar en situaciones de ansiedad, estrategias para la resolución de problemas, retroalimentación informativa mediante el conocimiento inmediato y directo de resultados y oportunidad para reproducir

una sucesión de hechos que difícilmente pueden ser repetidamente observados en un contexto real.

Corresponde destacar que, si bien muchas de las aplicaciones actuales están basadas en la potencialidad de las TICs para distribuir información, un aspecto que está cobrando mayor significación es el potencial de estas tecnologías para constituir nuevos escenarios educativos basados en los conceptos de comunidad y colaboración, que potencian la dimensión social del aprendizaje [9].

Se ha venido estudiando y abordando como elemento integrador de todos los aspectos que hemos mencionado hasta aquí, a los entornos virtuales de enseñanza y aprendizaje (EVEA), incorporándolos en diferentes experiencias educativas, ya sea para la articulación con escuela media, y el grado y el posgrado universitario [10].

Por supuesto, también surgen aquí cuestiones vinculadas a la calidad de los materiales y de las ofertas educativas [11].

### 3. Importancia de los temas.

Por lo expuesto se entiende que los temas planteados en este proyecto son de importancia, tanto desde el punto de vista informático como social y, particularmente, en el ámbito educativo.

### Líneas de Investigación y Desarrollo

- Gobierno Electrónico

En esta línea estamos interesados en cuatro temas concretos que entendemos son de interés nacional y regional:

- La incorporación de sistemas de identificación segura en servicios de E- Government que lo requieren (Voto por InterNet,

trámites impositivos, acceso a servicios de salud, pagos)

- Los sistemas de monitoreo de ciudades en tiempo real y la toma de decisiones para los mismos.

- La incorporación de tecnologías móviles para su empleo por los ciudadanos y por el Estado en aplicaciones de gobierno electrónico.

- Los Sistemas Inteligentes aplicados en problemas de toma de decisiones por el Estado, en particular utilizando Minería de Datos.

- E-Learning.

En esta línea nuestro interés pasa por tres temas de incorporación de TICs en ELearning:

- La aplicación de TICs como medio didáctico para facilitar los procesos de enseñanza y de aprendizaje.

- El empleo en el aula. De tecnologías móviles y la migración de entornos de EAD a dispositivos móviles.

- La aplicación de sistemas inteligentes al ámbito educativo, en particular para la representación y recuperación de conocimiento.

## Resultados y Objetivos

El proyecto ha comenzado este año, por lo que a la fecha no se dispone de resultados. Los que esperamos obtener en distintos momentos de su ejecución incluyen:

- A partir del estudio de sistemas inteligentes basados en heurísticas y su aplicación en minería de

datos relacionada con toma de decisiones, sugerir aplicaciones en el gobierno de municipios / provincias. Analizar el caso particular de los sistemas inteligentes aplicados a los centros de supervisión en tiempo real de cámaras en ciudades (en particular en temas de seguridad y sanidad).

- A partir de las experiencias desarrolladas sobre E –Learning definir y mejorar herramientas existentes que complementen los procesos de enseñanza y de aprendizaje.
- Estudiar y desarrollar experiencias en la utilización de tecnologías móviles en aula, así como la migración de aplicaciones de Entornos de EAD a celulares. Estudiar la incorporación de inteligencia a sistemas de representación y recuperación del conocimiento en el ámbito educativo.

## Formación de Recursos Humanos

Uno de los beneficios más valiosos del proyecto será la consolidación de parte del grupo de I/D que se formara inicialmente con los proyectos anteriores (“Evaluación y desarrollo de herramientas multimediales para análisis de competencias y aplicación de una metodología didáctica para mejorar el aprendizaje inicial en Informática”, “Metodologías y herramientas para la educación no presencial utilizando tecnología multimedial” “Experiencias de intervenciones docentes en espacios virtuales” y “E-Citizen E E-Learning. Utilizando las TICs para reducir la brecha digital y mejorar los servicios al ciudadano”), oportunamente apoyados

por la UNPSJB y en los que han participado recursos humanos de ambas instituciones.

La mayoría de los docentes integrantes del proyecto dirigen tesis de grado relacionadas con el proyecto en la UNTDF.

El grupo de trabajo está formado por 13 docentes - investigadores, de los cuales 1 es Doctor en Ciencias, 2 son Magister, 4 son Especialistas, 4 son Licenciados y dos son alumnos con título de pregrado.

Dos trabajos de tesis de licenciatura, relacionados con las áreas de investigación, fueron desarrollados el año anterior. Está en desarrollo una tesis de licenciatura y dos de maestría.

## Referencias

- [1] Zapata Lopez F., Sociedad del Conocimiento y Nuevas Tecnologías. [En línea]. Disponible en <http://www.oei.es/salactsi/zapata.htm> [Último acceso; 10/03/2013]
- [2] Caballero, S. L. (2005). Prácticas emergentes: la ciberdemocracia, las telecomunidades de conocimiento y los telecentros como alternativas para el desarrollo. Cuadernos del CENDES, 22(58) 97-114. [En línea]. Disponible en <http://www.redalyc.org/articulo.oa?id=40305806> [Último acceso: 09 03 2013].
- [3] Crespo Molera E. (2008). Guía para el Análisis del Impacto de las tecnologías de la Información y la Comunicación en el Desarrollo Humano, Madrid: Universidad Politécnica de Madrid.
- [4] InfoDev and The Center for Democracy & Technology (2002). The E-Government Handbook for Developing Countries. [En línea]. <http://unpan1.un.org/intradoc/groups/public/documents/apcity/unpan007462.pdf>. [Último acceso: 09 03 2012].
- [5] Executive Office of the President - Office of Management and Budget (2002). E-Government Strategy: Simplified Delivery of Services to Citizens, 07 02 2002. [En línea]. Disponible en [http://www.usa.gov/Topics/Includes/Reference/egov\\_strategy.pdf](http://www.usa.gov/Topics/Includes/Reference/egov_strategy.pdf). [Último acceso: 08 03 2013].
- [6] Intel y Gov3 (government for the third millenium) (2006). Citizen Centric Government: Global best Practice in Delivering Public Services to Citizens and Business. [En línea]. Disponible en [http://media.centerdigtaled.com/K12\\_Blueprint/Intel\\_Blueprint\\_CitizenCentricGovt.pdf](http://media.centerdigtaled.com/K12_Blueprint/Intel_Blueprint_CitizenCentricGovt.pdf). [Último acceso: 08 03 2012].
- [7] Salinas J. (2004). Innovación docente y uso de las TICs en la enseñanza universitaria. Revista Universidad y Sociedad del Conocimiento – vol I.
- [8] Mayer R. (2007). The Cambridge Handbook of Multimedia Learning, Cambridge University Press.
- [9] Bruckman A. (2002). The Future of E-Learning Communities. *Communications of the ACM*, vol. 45, nº 4, pp. 60-63.
- [10] Sanz C., González A., Zangara A. De Giusti, A. (2004). WebINFO: un entorno de aprendizaje web. *Eduotec 2004*. Barcelona, España.
- [11] Gorga G., Sanz C., Madoz C. (2011) ECALEAD- Evaluación de Calidad en Educación a Distancia. Análisis del modelo propuesto. XVII Congreso Argentino de Ciencias de la Computación (CACiC2011). La Plata, Buenos Aires. Argentina (pág. 519 – 528)

# Aplicaciones de la Robótica en la Integración de Contenidos en Carreras Técnicas

Frittelli, V. – Osimani, C. – Funes, G. – Geremía, W. – Reyes, D. – Aranda, M. – Blanco, M. J. – Oliver, N.

Proyecto *GoGo II* – Carrera de Ingeniería Informática – Universidad Blas Pascal (UBP) – Córdoba

Dirección: Av. Donato Álvarez 380 – Argüello (5147) – Córdoba (Capital)

Teléfono: (0351) 4144444

{ [vfrittelli@gmail.com](mailto:vfrittelli@gmail.com) - [cesarosimani@gmail.com](mailto:cesarosimani@gmail.com) - [waldo.geremia@gmail.com](mailto:waldo.geremia@gmail.com) - [gustavo.funes@develsa.com](mailto:gustavo.funes@develsa.com) - [diegor984@hotmail.com](mailto:diegor984@hotmail.com) - [elmario\\_207@hotmail.com](mailto:elmario_207@hotmail.com) - [majo\\_b16@hotmail.com](mailto:majo_b16@hotmail.com) - [nicolasoliver03@hotmail.com](mailto:nicolasoliver03@hotmail.com) }

## Resumen

Durante 2011 y 2012 se desarrolló y completó con éxito desde la Universidad Blas Pascal (UBP) el proyecto *Tecnología Educativa para Docentes y Alumnos del Nivel Medio (Proyecto GoGo)* (Universidad Blas Pascal, 2012) Su objetivo era proponer aplicaciones de la *robótica elemental* en el ámbito de la escuela media, tomando como base el *Proyecto GoGo Board* de la Universidad de Stanford (GoGo Board - Stanford University, 2011). El proyecto que se presenta con este nuevo escrito, es la segunda parte de aquel proyecto. Se enmarca en las actividades de investigación y transferencia de las carreras tecnológicas de la UBP, y se designa como *Usos y Aplicaciones de la Robótica como Medio para la Integración de Conocimientos en Carreras de Tecnología Aplicada (Proyecto GoGo II)*. Se propone diseñar experiencias prácticas orientadas a la informática y la robótica elemental, que se apliquen en *carreras técnicas universitarias*, comenzando con aquellas ofrecidas por la propia UBP (*Ingeniería Informática e Ingeniería en Telecomunicaciones*) Durante la etapa

de desarrollo, los investigadores de la UBP guían y asesoran a los docentes participantes, y analizan los principales ejes curriculares de los planes de las carreras citadas para identificar áreas integradoras que puedan ser complementadas con actividades de robótica introductiva basadas en el uso de GoGo Board.

**Palabras clave:** Creatividad - Integración – Interacción Docente – Robótica - Tecnología Educativa.

## Contexto

Como una continuación del *Proyecto GoGo* (Frittelli, et al., 2012) desarrollado en 2011 – 2012 y presentado en *WICC 2012*, se plantea ahora la segunda etapa. Se mantiene vigente la necesidad de aportar experiencias que innoven en la metodología de la enseñanza, pero ahora se avanza sobre la base de investigar y desarrollar estrategias basadas en el uso intensivo de la tecnología informática y elementos de robótica aplicada alrededor del kit GoGo Board orientadas a la aplicación de conocimientos y prácticas integradas, pero de carreras universitarias. En



paralelo al Proyecto GoGo original (que se aplicó sobre escuelas del nivel medio), durante los períodos lectivos 2011 y 2012 se trabajó dentro de la UBP en el planteo de proyectos basados en el kit GoGo Board pero generados como Prácticas de Laboratorio para asignaturas de la Carrera de Ingeniería Informática. Estas actividades estuvieron a cargo de un docente especialmente designado para la actividad, quien además actuaba como asesor en el Proyecto GoGo original. El fundamento del *Proyecto GoGo II* es tomar como base esas experiencias, estudiar los resultados y nuevas perspectivas obtenidas, y luego formalizar el planteo de trabajos de tecnología aplicada que integren las diversas asignaturas del plan de estudios, en distintos niveles de carreras técnicas (tales como Informática, Telecomunicaciones, Electrónica, etc.) Esto incluye el trabajo conjunto de docentes de diversas disciplinas y áreas, así como un fuerte trabajo de formalización de procedimientos, algoritmos, desarrollos matemáticos, planteos físicos, etc., de forma que cada desarrollo sea un trabajo dirigido, fundamentado e integrador (y no una práctica de avance por ensayo y error). Los trabajos se usarán para contribuir a la verificación práctica de conocimientos teóricos (experiencias de marcado corte tecnológico con robots dentro del campo de la física o la química), al desarrollo de tecnologías aplicadas en campos que requieran algún tipo de control automático de mecanismos (seguridad industrial, comunicaciones), al testing y el benchmarking (comparación de rendimientos de algoritmos y procesos, control de procesos) y a la implementación de técnicas de inteligencia artificial (comportamiento

autónomo de robots e identificación de patrones)

## Introducción

El proyecto GoGo Board de la Universidad de Stanford (GoGo Board - Stanford University, 2011) es una iniciativa que busca promover y canalizar actividades para proyectar, diseñar, construir y experimentar con elementos accesibles, de muy bajo costo pero de alto valor formativo, como herramienta para contribuir con la motivación de los jóvenes que estudian disciplinas tecnológicas (Sipitakiat, Blikstein, & Cavallo, 2002). El *Proyecto GoGo II* de la UBP aporta un contexto teórico y práctico sistematizado sobre la temática de la creación de material para la realización de actividades de tipo experimental con soporte de robótica básica, apropiadas a las condiciones particulares de las instituciones universitarias. Promueve además la transferencia de información y conocimiento y la conformación de una comunidad para el intercambio y el trabajo colaborativo en este campo. Durante el proceso de exploración de temas y áreas de aplicación para la integración de conocimientos, se pueden identificar nuevos espacios de inserción profesional para egresados de diversas carreras universitarias. Se trata de un proyecto de marcado corte tecnológico, ya que aspira al diseño de prototipos de informática aplicada la exploración de formas innovadoras de aplicación de la robótica elemental (Rusk, Resnick, Berg, & Pezalla-Granlund, 2008) para su uso en tecnología educativa en el nivel universitario. Asimismo, permite la identificación y favorece la relación de los principales actores que intervienen con prácticas y proyectos propios,

instancias de diálogo, de intercambio y de colaboración. En el desarrollo del proyecto se prevé:

- a. La realización de estudios prácticos de la ciencia y la tecnología.
- b. El uso de la tecnología como instrumento de divulgación, comunicación y formación.
- c. El diseño de instrumentos prácticos novedosos relacionados con los procesos de enseñanza y de aprendizaje, y con la transferencia de conocimientos en las disciplinas científicas y tecnológicas.
- d. El aporte de creatividad e innovación para el desarrollo tecnológico, cultural y social.

Se propone entonces, *analizar los principales ejes curriculares de las asignaturas inicialmente propuestas en carreras técnicas de la UBP, a los efectos de identificar áreas que puedan ser integradas con actividades de robótica experimental.* Estas actividades se logran mediante el diseño y la construcción de aplicaciones basadas en el uso de kits *GoGo Board*.

*GoGo Board* es un kit, basado en una placa electrónica, fácilmente conectable a un computador mediante USB y al cual se le pueden conectar diversos componentes externos. La placa *GoGo* puede ser programada en lenguaje Logo y otros lenguajes más adecuados al ámbito universitario. Con entrenamiento y creatividad se pueden plantear con relativa sencillez diversos dispositivos tales como pequeños vehículos que se puedan mover en forma autónoma o controlados desde la computadora o bien dispositivos provistos de distintos

tipos de sensores que realicen alguna tarea en forma automática. Si se aplican correctamente estos recursos, se puede contribuir a facilitar el aprendizaje y el avance, ya sea en el campo de las tecnologías aplicadas, o en forma transversal en otros campos (Bers, New, & Boudreau, 2004).

## Líneas de Investigación y Desarrollo

Las principales actividades de investigación del proyecto incluyen las siguientes líneas:

- ✓ Determinar posibles áreas curriculares de aplicación, en conjunto con los docentes participantes.
- ✓ Analizar el impacto de la tecnología aplicada en dichas áreas, y la mejor forma de implementación de nuevas prácticas con elementos de robótica.
- ✓ Documentar y analizar las actividades y resultados obtenidos por los alumnos en colaboración con sus docentes.

Aun cuando ahora se dirige hacia contenidos y temas prácticos de asignaturas universitarias, el *Proyecto GoGo II* continuará no obstante en la línea de transferencia de conocimientos hacia los colegios del nivel medio. Por otra parte, a medida que la experiencia vaya avanzando, se prevé integrar a nuevas asignaturas de las carreras testigo, y luego avanzar en proyectos colaborativos con otras universidades.

El Ministerio de Educación y el Ministerio de Ciencia y Tecnología de la Provincia de Córdoba y la UBP siguen realizando una campaña de difusión y aplicación del proyecto. En 2012 se dictó en la ciudad de Mina Clavero un

curso para docentes del nivel medio, de gran impacto en la región. En Octubre de 2012 se realizó en la UBP la Decimosexta Olimpíada Nacional de Telecomunicaciones (ONET 2012), con participación de alumnos de colegios de todo el país. La ONET 2012 incluyó un capítulo GoGo, en el cual se presentaron proyectos específicamente diseñados para el kit GoGo Board (Universidad Blas Pascal, 2012). Tanto a nivel ministerial provincial como a nivel de municipios y de empresas interesadas, la UBP ha conseguido apoyo económico y avales para el proyecto y se prevé un aumento de alcances y participación a lo largo de todo 2013.

## Resultados y Objetivos

El objetivo general del proyecto es *integrar conocimientos prácticos sobre aplicaciones de tecnologías informáticas y robótica elemental a las líneas curriculares de carreras técnicas universitarias, generando programas y prototipos reusables; y promoviendo luego la participación de alumnos y docentes en instancias de congresos y jornadas específicas de la temática.*

En cuanto a objetivos específicos, se plantean los siguientes:

- Selección de asignaturas y temas acordes para la integración mediante actividades prácticas basadas en kits GoGo Board.
- Coordinación del trabajo de los docentes de las diversas asignaturas seleccionadas, para elaborar propuestas de intervención práctica integradora.
- Transferir conocimientos prácticos de robótica hacia los alumnos de las asignaturas seleccionadas, en

colaboración directa con los docentes que participantes.

- Desarrollo en conjunto de los programas y prototipos tecnológicos sugeridos.
- Análisis y control de resultados de la aplicación de programas y prototipos sugeridos.
- Documentación de programas y prototipos, para su futura reutilización.
- Publicación de resultados en congresos, jornadas, eventos y área de gestión de investigación de la UBP.

Como resultado del proyecto se ha obtenido (o se espera obtener):

- La adecuación de los programas de las asignaturas de los planes de estudio, para incluir elementos y actividades prácticas surgidas de estas experiencias.
- La inserción de parte del cuerpo docente de las carreras involucradas en el proceso de formación de docentes investigadores.
- Programas y aplicaciones ya desarrolladas, listas para usar a modo de prototipos tecnológicos reusables.
- Resultados documentados de aplicación de prácticas en diversas asignaturas.
- Publicación de artículos en congresos y jornadas específicas.
- Espacios de participación para empresas privadas en calidad de auspiciantes.
- Espacios de colaboración con el Ministerio de Educación y el

Ministerio de Ciencia y Tecnología de la Provincia de Córdoba.

## Formación de Recursos Humanos

*Director del Proyecto:* Valerio Frittelli (Magister en Docencia Universitaria, Ingeniero en Sistemas de Información, Docente de Grado en la UBP, Investigador Senior en la UBP, Director de la Carrera de Licenciatura en Tecnología Educativa de la UTN Córdoba, Docente de grado y posgrado e investigador en la UTN Córdoba)

*Miembro Becado:* César Osimani (Ingeniero en Telecomunicaciones, becado para completar estudios de robótica en el Instituto de Tecnología de Monterrey – México, Docente e Investigador en la UBP)

*Asesor:* Waldo Geremía (Director de la carrera de Ingeniería Informática, director del Centro de Investigación Aplicada y Desarrollos en Informática y Telecomunicaciones de la UBP)

*Asesor:* Gustavo Funes (Licenciado en Informática, responsable de la transferencia *GoGo Board* en asignaturas iniciales de Ingeniería Informática)

*Estudiantes Adscriptos:* Diego Reyes, Mario Aranda, María José Blanco y Nicolás Oliver (estudiantes de Ingeniería Informática, para todos ellos el proyecto constituye su trabajo final de *Práctica Profesional Supervisada*)

## Referencias

[1.] Bers, M., New, R., & Boudreau, L. (2004). *Teaching and learning when no one is expert: Children and parents explore technology*. Retrieved March 16, 2013, from

Early Childhood Research and Practice, 6(2): <http://ecrp.uiuc.edu/v6n2/bers.html>

[2.] Frittelli, V., Osimani, C., Funes, G., Geremía, W., Reyes, D., Aranda, M., et al. (2012). Elementos de Robótica: Transferencia en Tecnología Educativa al Nivel Medio. *Workshop de Investigadores en Ciencias de la Computacion (WICC 2012)* (pp. 927-931). Posadas: RedUNCI.

[3.] GoGo Board - Stanford University. (2011). *Gogo Board - Science and Robotics for the Inventive Mind*. Retrieved March 16, 2013, from GoGo Board - [Official Website]: <http://gogoboard.stanford.edu/>

[4.] Rusk, N., Resnick, M., Berg, R., & Pezalla-Granlund, M. (2008). New Pathways into Robotics: Strategies for Broadening Participation. *Journal of Science Education and Technology*, 17(1), 59-69.

[5.] Sipitakiat, A., Blikstein, P., & Cavallo, D. (2002). The GoGo Board: Moving towards highly available computational tools in learning environments. *Proceedings of the Interactive Computer Aided Learning International Workshop*. Villach: Carinthia Technology Institute.

[6.] Universidad Blas Pascal. (2012, August 1). *Olimpiadas Nacionales de Electrónica y Telecomunicaciones 2012*. (CIADE-IT UBP) Retrieved March 16, 2013, from <http://www.ubp.edu.ar/onet/>

[7.] Universidad Blas Pascal. (2012, August 1). *Proyecto GoGo UBP [Official Website]*. Retrieved March 16, 2013, from CIADE-IT UBP: <http://www.ubp.edu.ar/ciade-it/proyect-gogo/>



# TIC, GC y e-educación: generación de sistemas informáticos.

Sonia I. Mariño<sup>1,2</sup>, María V. Godoy<sup>1,2</sup>, Barrios Walter G.<sup>1,2</sup>, Alderete Romina<sup>1,2,3</sup>,  
Mirta G. Fernandez<sup>1,2</sup>, Escalante Jaquelina E.<sup>1,2,3</sup>

<sup>1</sup>Área de Ingeniería Web. Facultad de Ciencias Exactas y Naturales y Agrimensura

<sup>2</sup>Departamento de Informática. Facultad de Ciencias Exactas y Naturales y Agrimensura.

Universidad Nacional del Nordeste.

9 de Julio 1449. CP: 3400. Corrientes. Argentina.

<sup>3</sup>Facultad de Humanidades. Universidad Nacional del Nordeste.

Av. Las Heras 727. CP: 3500. Chaco. Argentina.

[simarinio@yahoo.com](mailto:simarinio@yahoo.com), [mvgg2001@yahoo.com](mailto:mvgg2001@yahoo.com), [waltergbarrios@yahoo.com.ar](mailto:waltergbarrios@yahoo.com.ar),  
[ary\\_59@hotmail.com](mailto:ary_59@hotmail.com), [mirtagf@hotmail.com](mailto:mirtagf@hotmail.com), [jaqueescalante@yahoo.com.ar](mailto:jaqueescalante@yahoo.com.ar)

## Resumen

Se sintetiza la línea de investigación vinculada con el estudio y la aplicación de metodologías, técnicas y herramientas para el desarrollo del software de calidad en el dominio de la e-educación, a fin de generar soluciones informáticas para responder a necesidades regionales en el NEA. Se enfatiza la formación de recursos humanos en la temática, incorporando la misma en el desarrollo de tesinas de grado y becas otorgadas por la UNNE.

**Palabras clave:** e-educación, técnicas y herramientas, soluciones o productos informáticos, formación de recursos humanos.

## Contexto

En el trabajo se exponen los resultados obtenidos en los proyectos "Tecnologías de la información y comunicación como herramientas del desarrollo local" (2009-2012) y "Sistemas y TIC: técnicas y herramientas" (2012-2016), ambos acreditados por la Secretaría General de Ciencia y Técnica (UNNE) en la línea de e-educación desde la gestión del conocimiento y como aporte a la construcción de la Sociedad del Conocimiento.

## 1 Introducción

En los últimos años se ha observado un incremento en el uso de las TIC y de la multiplicidad de soluciones orientadas para facilitar el manejo, el acceso y la difusión de la

información. Cada vez es mayor el número de personas e instituciones que integran la tecnología (computadoras, servidores, software, programas de computación, conectividad, Internet, computadoras de mano, teléfonos celulares) al trabajo y a las actividades diarias.

En el marco de los proyectos de I+D "Tecnologías de la información y comunicación como herramientas del desarrollo local" y "Sistemas y TIC: técnicas y herramientas" ambos acreditado por la SGCyT (UNNE) se exponen los avances en el área.

Uno de los campos en que las TIC ha evolucionado e influido positivamente es la gestión de la educación basada en estas tecnologías.

La educación electrónica (e-Educación), basada en el uso de las nuevas tecnologías, no significa sólo un nuevo medio que se utiliza, sino un nuevo espacio social que se crea y, por tanto, un nuevo tipo de espacio educativo.

En los mencionados proyectos de I+D, se aborda la línea temática de e-educación. Se introduce el concepto de e-educación en [7] y [21].

El objetivo general de esta línea de trabajo se define como: "Realizar investigación aplicada, abordando métodos y herramientas apropiados para el diseño, desarrollo e implementación de sistemas de información en el dominio de e-educación, incorporando estándares de la Ingeniería del Software y tecnologías emergentes".

Es decir, se fundamenta en el estudio, análisis y aplicación de metodologías y herramientas orientadas a la generación de software de calidad para dominio de e-educación de acuerdo a las premisas de la Ingeniería del Software (IS). Además, se aborda la formación de recursos humanos hacia el grupo y hacia la comunidad transfiriendo los conocimientos adquiridos, desarrollando productos concretos tales como [1], [2], [3], [4] y [5].

En numerosos trabajos se describen los aportes de las TIC para gestionar datos y conocimiento en educación. En [18] vinculando las TIC y la educación, se establecen recomendaciones y lineamientos de políticas para el Estado como prospectiva 2020:

- Asegurar que todos los portales educativos cumplan ciertos criterios.
- Aumentar el intercambio de experiencias y contenidos de portales educativos, incluidas aplicaciones de Web 2.0 y otros canales de distribución.
- Difundir experiencias en el uso de herramientas de las TIC en programas educativos para fomentar la diversidad cultural.

Además, señala que la implementación de cambios en el sector educativo, necesita para la generación de contenidos y material de soporte una cantidad de tiempo sustancial de personal altamente calificado no sólo en los temas técnicos sino en las formas de transmitirlos.

La iniciativa desarrollada por este equipo de trabajo de la UNNE, coincide con los expuestos por otros equipos universitarios como [6] y [7].

## 2 Líneas de Investigación y Desarrollo

En la línea de I+D, con mira a la transferencia de sistemas accesibles se trabaja en:

- El relevamiento, selección y estudio de metodologías propias de la IS y de la Ingeniería Web. En este último caso enfatizando aspectos de esta plataforma como los abordado en NDT o Navigational Development Techniques [9], UWE o UML-Based Web Engineering [11], UWA o Ubiquitous Web Applications [22],
- La elección, análisis y estudio de herramientas informáticas para la incorporación de elementos reusables y que

aporten a la gestión de datos e información. Entre algunas se mencionan: tecnología móvil [4], acceso a redes sociales, aspectos de accesibilidad y usabilidad como medidas de calidad.

## 3 Resultados y Objetivos

En esta sección se mencionan los logros alcanzados en la línea temática vinculados a los mencionados proyectos que facilita la concreción de actividades de investigación aplicada, desarrollo y transferencia dirigida al contexto de influencia de la UNNE.

La implementación de las propuestas tecnológicas desarrolladas en esta línea temática, se plasma en innovaciones, en el sentido de modificar en el hacer actividades educativas y el acceso y difusión de la información, traduciéndose en una mejor atención hacia los potenciales usuarios. Éste concepto se denomina “ubicuidad” [18], es decir, proporcionar disponibilidad de información en cualquier momento y lugar, a su vez seguridad, fiabilidad y confiabilidad.

Los resultados se plasmaron en presentaciones de congresos, revistas y reuniones científicas ([2], [12], [13], [14], [15], [16], [17] y [19]) y formación de recursos humanos de grado [10], [24] y postgrado [8].

Se profundizó en la generación de plataformas educativas basadas en Moodle. Se eligieron y adecuaron herramientas enfatizando el acceso a redes sociales, se incluyó el acceso a notificaciones en dispositivos móviles sin costo ni necesidad de acceso a internet; y desde el punto de vista educativo como una herramienta para el intercambio de información y construcción colaborativa del conocimiento. Se desarrollaron dos plataformas siendo sus funcionalidades difundidas en [4] y [23].

Por otra parte, el desarrollo de contenidos y productos de e-learning necesita de la existencia de especificaciones y normas que le permitan un crecimiento sostenido, en este sentido en [4], se seleccionó la plataforma Moodle, y actualmente se trabaja en el estudio de los estándares de e-learning.

Los estándares de e-learning permiten a los sistemas y contenidos poder interactuar y combinarse con gran independencia. Los estándares abarcan también cuestiones como la accesibilidad y las normas ISO.

Se prevé como línea futura, adicionar un modulo que cumpla con estándares SCORM

[20], para la comunicación con otras plataformas.

En SCORM [20] las principales habilidades identificadas como fundamentales para la concepción de los contenidos y plataformas de e-learning son: interoperabilidad, reusabilidad, manejabilidad, accesibilidad, durabilidad, escalabilidad y efectividad en los costos.

## 4 Formación de Recursos Humanos

En referencia a la formación de RRHH se mencionan los logros de los años 2011 y 2012.

- Continuaron el estudio de metodologías y herramientas de la IS y su incorporación en análisis y diseño de sistemas de información 1 becario de grado [23] y 1 becario de postgrado [8] de la SGCyT – UNNE
- Personal de investigación del proyecto, se desempeñan como profesores orientadores del Trabajo Final de Aplicación en esta temática. Se resalta este logro considerando que es una modalidad de incrementar la masa crítica comprometida con la docencia e investigación.
- Desarrollan el Trabajo Final de Aplicación (TFA) alumnos de la carrera Licenciatura en Sistemas de Información. Los proyectos de TFA incluyeron la selección y estudio de herramientas, la aplicación de metodologías propias de la IS y de la Ingeniería Web adecuadas a la educación.

## 5. Referencias

- [1] Alderete, R., Escalante, E., Mariño S. I., Godoy, M. V. (2011). "Implementación del módulo Verbos del Prototipo educativo Enseñanza para la lengua del 5º grado". Anales del VI Congreso Argentino de Tecnologías en Educación y Educación en Tecnología. TE&ET'11. Aceptado para su exposición. Red UNCI. Argentina.
- [2] Alderete, R. Y. Mariño, S. I., Primorac, C., Escalante, J. E. y Godoy, M. V. (2012). "Migración del Software Educativo Edu TIC a una plataforma FLOSS accesible vía Web". Revista Internacional sobre Tecnología, Conocimiento. Universidad de California, Los Ángeles (UCLA), EEUU. Aceptado para su publicación.
- [3] Alfonzo, P. L., Abellan, J., Mariño, S. I., Ferrari Alve, S. I. (2012). "Aula virtual de la asignatura Laboratorio de Programación para mejorar el rendimiento de los alumnos universitarios. Caracterización de alumnos del ciclo lectivo 2011". Anales del VII Congreso Tecnologías en Educación y Educación en Tecnología. TE&ET'12. Red UNCI. Argentina.
- [4] Barrios, W. Godoy, M. V., Fernández, M., Mariño, S. I. (2012). "De Moodle a Entornos Personales de Aprendizaje (PLE): Introducción de herramientas sociales a una plataforma e-learning". Simposio Sociedad de la Información SSI 2012, 41 JAIIO. Jornadas Argentinas de Informática e Investigación Operativa. Co-organizadas por SADIO y la UNLP. Facultad de Informática de la Universidad Nacional de La Plata. La Plata 27 al 31 de Agosto de 2012.
- [5] Castro Chans, B. Godoy, M. V., Sobol, B. Mariño, S. I. (2012). "Implementación de un EVA para el ingreso a la Universidad: El caso del módulo Estrategias de Aprendizaje en la Universidad para los alumnos de la Licenciatura en Sistemas de Información". Anales del VII Congreso Tecnologías en Educación y Educación en Tecnología. TE&ET'12. Red UNCI. Argentina.
- [6] Cataldi, Z. y Lage, F. J. (2012). "TICs en Educación: Nuevas herramientas y nuevos paradigmas. Entornos de Aprendizaje Personalizados en dispositivos móviles". Anales VII Congreso de Tecnología en Educación y Educación en Tecnología. TE&ET'12. Red UNCI. Pergamino, Argentina.
- [7] Díaz, F. J., Schiavoni A., Amadeo, A. P., Charnelli M. E. (2012). "Diseño y construcción de objetos de aprendizaje. Su integración en repositorios y plataformas virtuales de aprendizaje". XIV Workshop de Investigadores en Ciencia de la Computación. WICC 2012, Posadas. Universidad Nacional de Misiones.
- [8] Escalante, J. E. Becaria de Postgrado – Iniciación. Secretaría General de Ciencia y Técnica. Universidad Nacional del Nordeste. Resolución N° 987/10 C.S. Período: 01/03/2011-30/03/2014. Tema: "TIC, Herramienta y sus Funcionalidades como Mediadora de Modelos Conceptuales de Gestión del Conocimiento y su Integración en los Sistemas de

- Información". Directora: Mariño, S. I. Codirector: Godoy Guglielmone, M. V.
- [9] Escalona, M. (2004). "Modelos y técnicas para la especificación y el análisis de la Navegación en Sistemas Software". Ph. European Thesis. Department of Computer Language and Systems. University of Seville. Seville, Spain.
- [10] Fernández, M. G. Becaria de Pregrado. Secretaría General de Ciencia y Técnica. Universidad Nacional del Nordeste. Resolución N° 141/12 C.S. Período: 01/03/2012-30/03/2013. Tema: "E-salud. Un prototipo de sistema de información Web para la gestión del Dispensario Dermatológico de Corrientes". Directora: Godoy Guglielmone, M. V. Codirector: Mariño, S. I.
- [11] Koch, N. (2001). "Software Engineering for Adaptative Hypermedia Applications". Ph. Thesis, FAST ReiheSoftwaretechnik. Vol. 12, Uni-Druck Publishing Company, Munich. Germany.
- [12] Mariño, S. I., y Godoy, M. V. (2012). "Reflexiones preliminares de la teoría de la actividad y el desarrollo de software educativo". *Revista de Educación de Extremadura*. Vol. 2, No 3. pp. 27-55. ISSN: 2173-9536.
- [13] Mariño S. I., Maria V. Godoy, M. V., Acosta, J., Roa, D., Mendiburu, A. (2011). "Algunas experiencias de vinculación Universidad – Sistema de Educación, mediante la generación de soluciones informáticas variadas". Anales VI Congreso Argentino de Tecnologías en Educación y Educación en Tecnología. TE&ET'11. Red UNCI. Argentina.
- [14] Mariño, S. I., Godoy Guglielmone, M. V., Alfonzo, P., Escalante, J., Barrios, W., Primorac, C. Fernandez, M., Alderete, R. (2012). "Aplicación de procesos de ingeniería de requerimientos en la generación de productos software Área Ingeniería de Software". Anales XIV Workshop de Investigadores en Ciencias de la Computación. WICC 2012 Posadas, Misiones (Argentina).
- [15] Mariño, S. I., Godoy, M. V., Busso, L. E. y Escalante, J. E. (2011). "Construcción de un EVEA basado en tecnología FLOSS para la gestión de contenidos matemáticos". *Revista de la Escuela de Perfeccionamiento e Investigación Operativa - EPIO. Edición 32 (digital)*, pp. 207-230.
- [16] Mariño, S. I., Godoy, M. V., Escalante, J. E., Acosta, J., Roa, D., Sánchez, K., Lezcano, J., Zacarías, G., Bulloni, D., Schaeffer, M. (2011). "Construcción de software educativos. Síntesis de algunas experiencias 2009-2010". *Hologramatica, Año VII, Vol. 14, No 4*, pp. 55-76. ISSN 1668-5024.
- [17] Meza, L. I., Mariño, S. I., Godoy, M. V. (2012). "Desarrollo de un Software Educativo para asistir el proceso de enseñanza - aprendizaje en una escuela profesional. El caso de la asignatura "Tecnología y Práctica Estético Capilar". Anales del VII Congreso Tecnologías en Educación y Educación en Tecnología. TE&ET'12. Red UNCI. Argentina.
- [18] Ministerio de Ciencia, Tecnología e Innovación Productiva. (2009). Libro Blanco de la Prospectiva TIC. Proyecto 2020.
- [19] Primorac, C. R., Mariño, S. I. y Godoy, M. V. (2012). "TIC y Sociedad: Especificación de Requisitos para Apoyar la Gestión de Información". *Revista Digital Sociedad de la Información. Monográfico TIC. No 33*. ISSN: 1578-326x.
- [20] SCORM: Sharable Content Object Reference Model. Disponible en <http://www.adlnet.gov/scorm>.
- [21] Touriñán López J. M. (2007). "Integrar la escuela en la Sociedad de la Información: desmitificar la perspectiva TIC y orientar la educación electrónica". *Tecnología y Comunicación Educativas Año 21, No 45*. Dic. Catedrático de la Universidad de Santiago de Compostela.
- [22] UWA (2001). UWA Requirements Elicitation: Model, Notation, and Tool Architecture. Disponible en: [www.uwaproject.org](http://www.uwaproject.org).
- [23] Vanderland, M. A., Mariño, S. I. y Godoy, M. V. (2011). "Desarrollo de un EVEA Utilizando Herramientas de Software Libre. El Caso de la Asignatura Inteligencia Artificial". IEEE-RITA". *Revista Iberoamericana de Tecnologías del Aprendizaje. Vol. 6, No 4*, pp. 147-154.
- [24] Vilotta, José M. M. Beca de Pregrado: Gestionando el conocimiento: Un sistema de información para administración de la



Enseñanza Privada en la Provincia.  
Otorgada por Resol. 844/12. Director:  
Godoy Guglielmo, M. V. Co-Director:  
Mariño, S. I. Licenciatura en Sistemas de  
Información Corrientes – Argentina 2012.

## Tecnologías de la Información y la Comunicación en ámbitos educativos: Análisis de experiencias y resultados de su aplicación

Sanz Cecilia, Madoz Cristina, Gorga Gladys, Zangara Alejandra, Gonzalez Alejandro, Russo Claudia, Depetris Beatriz<sup>(1)</sup>, Ibáñez Eduardo, Martorelli Sabrina, Artola Verónica, Albanesi Bernarda, Sanchez Mariano.

Instituto de Investigación en Informática LIDI (III-LIDI)  
Facultad de Informática – UNLP

{csanz, cmadoz, ggorga, agonzalez, crusso, eibanez, smartorelli, bartola, balbanesi, msanchez}  
@lidi.info.unlp.edu.ar, alezan@elsitio.net, depetrisb@gmail.com

### CONTEXTO

Este subproyecto forma parte del proyecto “Tecnología y aplicaciones en Sistemas de Software Distribuidos. Experiencias en E-learning, E-government y Sistemas productivos”, del Instituto de Investigación en Informática LIDI, acreditado por la UNLP (período 2010-2013). Es una continuación del proyecto (2006-2009): “Sistemas de Software Distribuidos. Aplicaciones en procesos industriales, E-government y E-learning”, también acreditado por UNLP.

### RESUMEN

Este subproyecto tiene como objetivo realizar investigación en áreas de Tecnología Informática aplicada a Educación. Se presentarán aquí los avances de éste respecto de los objetivos planteados y las líneas de investigación correspondientes.

Como avances respecto del año 2011 se han obtenido títulos de Magister y de Especialización en Tecnología Informática Aplicada en Educación. Se han presentado propuestas de doctorado y se está trabajando con alumnos en varias tesis de grado en temas vinculados al proyecto. De igual modo se han abordado diferentes experiencias de inclusión de tecnología en ámbitos educativos.

Como resumen de lo realizado se abordaron los siguientes temas:

- El desarrollo de metodologías para trabajo colaborativo usando TICs.,
- Avance en el desarrollo de las métricas e indicadores para medir la calidad de los procesos educativos mediados por tecnología. Análisis de las experiencias desarrolladas.
- Análisis y desarrollo de materiales educativos multimediales e hipermediales. Análisis de ambientes educativos 3D.

- Análisis e implementación de laboratorios virtuales y remotos.
- Uso de la tecnología móvil integrada a entornos virtuales.
- Investigación sobre entornos virtuales de enseñanza y aprendizaje y posibilidades que se generan a partir de su utilización.

Se detallan las líneas de investigación y resultados obtenidos, en particular durante el año 2012.

**Palabras claves:** Trabajo y aprendizaje colaborativo, Laboratorios remotos y virtuales, EVEA, Materiales educativos.

### INTRODUCCION

En el presente proyecto se abordan diferentes aspectos vinculados a la incorporación de tecnología en el ámbito educativo. Esto va desde estudios más teóricos a experiencias concretas donde se diseñan metodologías específicas y luego se evalúa el impacto de dichas experiencias. En general, se implementan y/o utilizan herramientas digitales innovadoras en experiencias educativas concretas, aplicando metodologías que permitan su aprovechamiento.

A continuación, y sólo de manera introductoria, se mencionan algunos de los temas vinculados al proyecto, de manera tal de dar contexto a esta presentación.

Con el avance de la Web2.0 y sus posibilidades para lograr la colaboración entre personas, este tema ha adquirido especial interés en el ámbito académico y particularmente en este subproyecto. Como ejemplo, se puede mencionar la Maestría de Tecnología Informática Aplicada a Educación en la que participan varios investigadores de este proyecto, en donde se vienen desarrollando varias

<sup>1</sup> Universidad Nacional de Tierra del Fuego

experiencias de trabajo colaborativo utilizando herramientas informáticas. Es conocido que el planteo de un trabajo colaborativo en el proceso de aprendizaje involucra el diseño de una consigna, el análisis del rol que tendrá cada uno de los docentes y alumnos participantes, el establecimiento de los medios que resulten necesarios para colaborar más allá de la presencialidad, entre otro. En este último caso toman un especial interés aquellas tecnologías que median el proceso de colaboración a distancia. Se han estudiado y analizado diferentes posibilidades en este sentido e incluso se han abordado temas fuertemente vinculados a la interdisciplinariedad en el aprendizaje colaborativo mediado por TICs. En este sentido, las tendencias actuales muestran una evolución de los entornos virtuales de enseñanza y aprendizaje que permiten acompañar este tipo de experiencias [Sanz 2008] [Sanz 2012]

Otra de las líneas de trabajo del subproyecto se vincula con el diseño, desarrollo y aplicación de Objetos de Aprendizaje en los ámbitos educativos. Una de las actividades realizadas por los profesores en propuestas educativas mediadas por TIC, es la creación de materiales digitales, que buscan acompañar el proceso de aprendizaje. Se trata de materiales que los educadores pueden usar y reutilizar en distintos ambientes.

El paradigma de los Objetos de Aprendizaje (OA) centra su atención en los contenidos educativos y propone su descripción mediante metadatos. La gran cantidad de recursos de aprendizaje y en particular OAs, ha provocado el surgimiento de los denominados Repositorios de Objetos de Aprendizaje, que permiten desplegar y gestionar estos recursos mediante sistemas que apoyan el almacenamiento, etiquetado y recuperación de OA. [Menedez 2012]. En el marco del suproyecto, se trabaja sobre metodologías de diseño de OA, el estudio de diferentes repositorios y sus posibilidades, la utilización de OA para diferentes materias, la evaluación de la calidad de estos objetos, entre otros.

Por otra parte se continúa trabajando en la evolución del entorno virtual de enseñanza y aprendizaje, desarrollado en el III LIDI, de manera de ir adaptándolo e integrándolo con otras herramientas tecnológicas. Así por ejemplo, se ha avanzado en su integración con un laboratorio remoto vinculado al tema de Procesamiento Paralelo Multi-Cluster [De Giusti 2012].

Hasta aquí, se han trazado brevemente algunas de las aristas de este subproyecto que buscan responder interrogantes como ¿qué recursos y estrategias

utilizar en cada caso, y con qué metodología?, ¿cómo diseñar y desarrollar los materiales educativos más adecuados?, ¿qué innovación proponen los objetos de aprendizaje?, ¿qué herramientas de la Web2.0 pueden ayudar en los procesos educativos?

Los aspectos teóricos que se abordan en este subproyecto y los diseños que se realizan tratan de responder los cuestionamientos planteados en diferentes experiencias educativas concretas. Por ello, uno de los puntos centrales que hemos trabajado en estos últimos años ha sido la implementación de experiencias tanto en escuelas de nivel medio, como en carreras universitarias de grado y postgrado. Esto nos permite tener una retroalimentación respecto de las hipótesis que planteamos.

En la siguiente sección se presentará concretamente el listado de las líneas de I/D que se abordan.

## **LINEAS DE INVESTIGACION / DESARROLLO**

Se mencionan aquí las principales líneas de investigación y desarrollo abordadas en el marco del proyecto:

- Entornos virtuales de enseñanza y aprendizaje. Proyectos que los incorporan, metodologías, desarrollos, evaluación de su calidad y experiencias.
- Materiales educativos digitales. Objetos de aprendizaje. Mutimedias e hipermedias.
- Innovación en el uso de tecnología en ambientes educativos. Experiencias. Uso de tecnología móvil, classmates pc, OLPC, etc. (Herrera 2012)
- Formación de recursos humanos en el uso de TICs.
- Trabajo colaborativo mediado por TICs. Conceptualización, análisis y desarrollo de software y metodologías.
- Simuladores, laboratorios virtuales y remotos. Ambientes virtuales 3D. Aplicaciones, desarrollos, experiencias (De Giusti 2012).

## **RESULTADOS ESPERADOS/OBTENIDOS**

Se detallan a continuación algunos de los resultados obtenidos en este proyecto, para algunas de las líneas de I/D mencionadas que han tenido mayor desarrollo durante el año 2011.

▪ **Entornos virtuales de enseñanza y aprendizaje. Proyectos que los incorporan, metodologías, desarrollos, evaluación de su calidad y experiencias.**

Vinculado a esta línea de investigación se ha continuado con la realización de experiencias educativas en el entorno virtual de enseñanza y aprendizaje, diseñado y desarrollado en el marco del proyecto.

Se continúa con la participación en un proyecto de evaluación de calidad de este tipo de entornos virtuales en conjunto con la Universidad Complutense de Madrid.

Se desarrolló una tesis de Magister (finalizada y esperando evaluación) dirigida por un integrante del proyecto, vinculada al desarrollo de una propuesta de evaluación de los EVEAs basada en criterios de usabilidad [Ferreira 2012].

Se estudian posibles extensiones y actualizaciones acordes a los avances y análisis del estado del arte de los EVEAs. En particular, se investigan herramientas que puedan ser integradas, para enriquecer su funcionalidad.

Se continúa realizando un taller anual en el marco de la Maestría de Tecnología Informática Aplicada a Educación, cuyos docentes se vinculan con este proyecto, sobre los entornos virtuales más utilizados por el ámbito académico, comparando sus funcionalidades.

Se ha realizado un avance en la evaluación de calidad del curso de preingreso a distancia que se desarrolla en el ámbito de la Facultad de Informática, utilizando el EVEA desarrollado en este proyecto. Este curso ha obtenido la certificación de calidad según la norma ISO 9001 (2012). El modelo de evaluación se aplicó a otro caso de estudio y se han realizado publicaciones [Gorga 2012]

▪ **Materiales educativos digitales. Objetos de aprendizaje. Mutimedias e hipermedias - Simuladores, laboratorios virtuales y remotos. Ambientes virtuales 3D. Aplicaciones, desarrollos, experiencias.**

Se ha terminado un trabajo de doctorado vinculado a temas de calidad de los Objetos de Aprendizaje. La dirección estuvo a cargo de un miembro del proyecto [Massa 2012]

Se ha desarrollado un simulador que permite introducir a los alumnos de primer año conceptos iniciales de concurrencia. Se han realizado publicaciones en relación a este tema [De Giusti 2012].

Se ha finalizado la tesina de grado vinculada al almacenamiento, búsqueda y recuperación de preparados virtuales, desde repositorios de imágenes, de microscopía virtual. Se han realizado experiencias de uso de este sistema y publicaciones al respecto [Martorelli 2012].

Se abordó el desarrollo de una tesina de grado (finalizada y esperando evaluación) que trabaja en el diseño de materiales educativos ad-hoc para diferentes contextos. Se estudian herramientas de autor que soporten estos desarrollos [Gonzalez 2012].

▪ **Innovación en el uso de tecnología en ambientes educativos. Experiencias. Uso de tecnología móvil, Classmates pc, OLPC, etc.**

Se está avanzando en el estudio de las posibilidades de la tecnología móvil integrada a los EVEAS. Un miembro del equipo está participando en la investigación orientada a esta temática.

Se ha presentado en 2012 una propuesta de tesis de doctorado que se focaliza en el uso de dispositivos móviles para el aprendizaje [Herrera 2012].

▪ **Formación de recursos humanos en el uso de TICs.**

Los miembros de este proyecto participan en diferentes instancias de formación de recursos humanos: dirección de trabajos de grado y especialización, tesis de maestría y doctorado, cursos de formación y capacitación.

Se detallan, en la sección de formación de recursos humanos, algunos avances en este sentido.

▪ **Trabajo colaborativo mediado por TICs. Conceptualización, software y metodologías.**

En esta línea de investigación, se viene desarrollando un estudio de herramientas específicas para llevar adelante aprendizaje y trabajo colaborativo, en procesos educativos. Al mismo tiempo, se ha desarrollado una metodología de trabajo colaborativo mediada por TICs en el marco de un curso de postgrado de la Maestría de Tecnología Informática Aplicada a Educación, que se actualiza y se pone en juego año a año. En particular durante el año 2011, se ha avanzado en la definición de un marco conceptual sobre



e-actividades, donde se enmarca el trabajo colaborativo mediado por TIC. Existe una tesis de Maestría y una tesis de doctorado que abordan esta temática. Se realizó un Taller de e-actividades en el marco de la Maestría y que se vincula con parte de las experiencias de trabajo colaborativo desarrolladas [Sanz 2012].

Se abordó el desarrollo de una tesina de grado (finalizada y esperando evaluación) que trabaja en el diseño de materiales educativos en un entorno colaborativo web para llevar adelante la técnica de “Metaplan”. Se ha realizado una experiencia de uso de este sistema. Se han realizado publicaciones al respecto [Gonzalez 2012]

Se ha avanzado en el diseño de una aplicación basada en interacción tangible para llevar adelante un juego didáctico colaborativo, utilizando una tabletop. Se presentó una propuesta de tesina de grado dirigida por un miembro del proyecto vinculada a este proyecto [Artola 2012].

En cuanto a los proyectos vinculados con la temática, el III- LIDI participa en los siguientes:

- “TICs aplicada a problemas de Gobierno Electrónico y de E-Learning” para mejorar las prestaciones de los distintos servicios del ciudadano. Se desarrolla en conjunto con la Universidad Nacional de la Patagonia San Juan Bosco (2012-2013)
- Se ha firmado un acuerdo de colaboración en estos temas con la Universidad de Zaragoza. En este marco una docente de dicha Universidad ha realizado una estadía de investigación en el III LIDI y dictado el curso de postgrado: “Paradigmas actuales de interacción” (Grupo de Informática Gráfica Avanzada).
- “Arquitecturas avanzadas, gestión del conocimiento y calidad: una respuesta coordinada a los retos de los campus virtuales de nueva generación”. Proyecto conjunto con la Universidad Complutense de Madrid y la UNED de España. (2012)
- FRIVIG: Formación de Recursos Humanos e Investigación en el Área de Visión por Computador e Informática Gráfica. Acreditado AECID (Agencia Española de Cooperación). Código 027824/09. Se continua con el proyecto en el período 2012-2013.
- RedAUTI: Red temática en Aplicaciones y Usabilidad de la Televisión digital Interactiva.

Miembros del proyecto participaron de la Organización de jAUTI2012 y del I Seminario anual de encuentro de grupos participantes de la RedAUTI 2012 [Zangara2012].

## FORMACION DE RECURSOS HUMANOS

Dos integrantes de esta línea de investigación están desarrollando su Tesis de Magister en Tecnología Informática Aplicada en Educación.

Se dirigen Tesis de Magister, Trabajos de Especialista y Tesinas de Licenciatura de la Facultad de Informática en temas relacionados con el Proyecto.

En 2012, se han aprobado 7 trabajos de Maestría y 2 de Doctorado, en el área de Tecnología y Educación, dirigidos por miembros de este proyecto.

Hay 2 trabajos de Maestría terminadas y esperando su evaluación y 1 propuesta de Doctorado y 3 propuestas de Especialización que están desarrollándose actualmente.

Hay 2 trabajos de grado terminados y esperando su evaluación y 1 trabajo de grado que se están desarrollando y se vinculan con el proyecto.

En la siguiente sección se presenta: la bibliografía y los trabajos citados aquí, algunos textos de estudio que se utilizan en la investigación, y por otra parte, algunas publicaciones que forman parte de los resultados de este proyecto.

## BIBLIOGRAFIA

- Artola V., Sanz C., Giacomantone J. Tesina de Grado “Interacción Tangible en aplicaciones educativas. Diseño e implementación de un prototipo basado en este paradigma de interacción orientado al aprendizaje colaborativo”. Facultad de Informática. UNLP. 2012.
- Astudillo G., Sanz C., Willging P. “Repositorios de Objetos de Aprendizaje: un espacio de búsqueda centrado en lo educativo”. VII Congreso de Tecnología en Educación y Educación en Tecnología. TE&ET. ISBN: 978-987-28186-0-9. Pág.187-196. Junín/Pergamino, Buenos Aires, Argentina, Junio 2012.
- Baldasarri S., Sanz C., Quinteros E. Propuesta de Especialización TIAE “Interacción tangible, educación y trabajo colaborativo”. Diciembre de 2012.

- Calderon M., Gonzalez A., Sanchez Salvioli A. "Comunicación visual orientada a EAD para docentes de UNLP. Una propuesta de desarrollo de materiales didácticos". VII Congreso de Tecnología en Educación y Educación en Tecnología. TE&ET. ISBN: 978-987-28186-0-9. Pág. 150-158. Junín/Pergamino, Buenos Aires, Argentina, Junio 2012.
- Crespo C., Zangara A. Propuesta de Maestría en Tecnología Informática aplicada en Educación "Análisis de las interacciones entre los actores de un escenario virtual de aprendizaje". Octubre de 2012.
- De Giusti A., Frati F., Sanchez M., De Giusti L. "LIDI Multi Robot Environment: Support software for concurrency learning in CS1" Collaboration Technologies and Systems (CTS), 2012 International Conference on Digital Object Identifier: 10.1109/CTS.2012.6261065. Page(s): 294 – 298. 2012
- De Giusti A., Frati F., Sanchez M., De Giusti L. "LIDI Multi Robot Environment: Support software for concurrency learning in CS1". Collaboration Technologies and Systems (CTS), 2012 International Conference on Digital Object Identifier: 10.1109/CTS.2012.6261065. Publication Year: 2012, Page(s): 294 - 298
- Ferreira Szpiniak A., Sanz C. "MUSA un modelo de evaluación de Entornos virtuales de enseñanza y aprendizaje. Aplicación a un caso de estudio". Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología. TE&ET. ISSN: 1850-9959. Nro 8. Pág: 94-103. 2012.
- Gonzalez A., Madoz C., Saadi F., Dan H. "Virtualización de Métodos de Trabajo Colaborativo. Una Propuesta para la Utilización de la Técnica de Metaplan". VII Congreso de Tecnología en Educación y Educación en Tecnología. TE&ET. ISBN: 978-987-28186-0-9. Pág. 10-18. Junín/Pergamino, Buenos Aires, Argentina, Junio 2012.
- Gonzalez A., Madoz C. Tesina de Grado: "Aplicaciones Hipermedia para el planteo y resolución de problemas. Un caso particular para los ingresantes a la Facultad de Ciencias Naturales de la UNLP". 2012.
- Gonzalez A., Fernandez Moujan I., Lovos E. "El uso de estrategias colaborativas mediadas por tecnología. La enseñanza de programas en Primer Año de la Lic. en Sistemas de la UNRN". Julio de 2012.
- Gonzalez A., Moujan I., Zampar M. Propuesta Especialización TIAE "Análisis de modelos tutoriales virtuales. Revisión de la utilización de herramientas asincrónicas en la educación no formal de enseñanza de informática". Diciembre de 2012.
- Gorga G., Sanz C., Madoz C., Manresa Yee C., Abásolo M. "ECALEAD: Evaluación de Calidad en Educación a Distancia. Aplicación en un caso de estudio". III Congreso Iberoamericano sobre Calidad y Accesibilidad de la Formación Virtual. Universidad de Alcalá. Alcalá de Henares (España). ISBN: 978-84-8138-367-6. Pag: 409-416. Abril 2012.
- Gorga G., Sanz C., Madoz M. "ECALEAD – Quality Assessment in Distance Education. Analysis of the proposed model". Publicado en el libro "Computer Science & Technology Series XVII Argentine Congress of Computer Science Selected Papers". EDULP, Argentina 2012. Pág. 109-119. ISBN 978-950-34-0885-8.
- Herrera S., Fennema M., Sanz C. "Estrategias de m-learning para la formación de posgrado". VII Congreso de Tecnología en Educación y Educación en Tecnología. TE&ET. ISBN: 978-987-28186-0-9. Pág. 363-370. Junín/Pergamino, Buenos Aires, Argentina, Junio 2012.
- Lovos E., Gonzalez A., Mouján I., Bertone R., Madoz M. "Estrategias de Enseñanza Colaborativa para un Curso de Programación de Primer Año de la Lic. en Sistemas". I Workshop de Innovación en Educación en Informática. CACIC2012. ISBN: 978987-1648-34-4. Pág. 1534-1543. Bahía Blanca, Buenos Aires, Argentina, Octubre 2012.
- Martorelli S., Sanz C., Giacomantone J. Martorelli S. "ParasitePics el primer repositorio de imágenes Parasitológicas de Argentina" Revista Argentina de Parasitología (RAP) Asociación Parasitológica Argentina, Volumen 1, No 1. Pag 2006. ISSN 2313-9862
- Massa S., Rico C., Huapaya R. "Generación de requerimientos de un Objeto de Aprendizaje a partir de escenarios: un caso de estudio para un curso de Programación Inicial". XVIII Congreso Argentino de Ciencias de la Computación. Bahía Blanca. Argentina.. Páginas 495-504. ISBN 978-987-1648-34-4. Octubre 2012.
- Massa S., Zangara A. Propuesta de Maestría en tecnología Informática aplicada en Educación "Aula Extendida en la educación superior en Ingeniería. Una propuesta de aplicación en el área tecnológica básica de electrotecnia". Octubre de 2012.
- Menéndez V., Domínguez M., Castellanos-Bolaños, Vidal-Castrob C, Segura A. "Un Modelo de Calidad de Objetos de Aprendizaje

- basado en la Semántica de sus Metadatos”. Séptima Conferencia Latinoamericana de Objetos y Tecnologías de Aprendizaje. ISSN 1982 – 1611. Guayaquil. Ecuador. Octubre 2012.
- Sanz C., De Giusti L. Ponzio C. Propuesta de Especialización TIAE “Herramientas de autor para el desarrollo de material educativo multimedial. Relevamiento y clasificación”. Noviembre de 2012.
  - Sanz C., Nobile C. Propuesta de Maestría en Tecnología Informática aplicada en Educación “Procesos de Integración de Tecnologías de la Información y la comunicación en Instituciones de Educación Superior. El caso de la Facultad de Ciencias Económicas de la UNLP”. Noviembre de 2012.
  - Sanz C., Santacruz Valencia, L. Propuesta de Maestría en Tecnología Informática aplicada en Educación “Estrategias de diseño y ensamblaje de Objetos de aprendizaje”. Octubre 2012.
  - Sanz C., Zangara A. (2012) “El desarrollo de los foros en el marco de las e-actividades en una propuesta de educación mediada”. Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología. TE&ET. ISSN 1850-9959. Número 7. Pág.29-35. Junio 2012.
  - Sanz C., Zangara A. “La escritura colaborativa como una e-actividad”. Reunión: XVIII Congreso Argentino de Ciencias de la Computación. Bahía Blanca. Argentina. Octubre 2012. Páginas 580-588 - ISBN 978-987-1648-34-4.
  - Sanz C., Zangara A., Manresa-Yee C. (UIB-España) “E-activities in teaching processes using icts Collaborative activity as a case study EDULEARN 2012 - 4th International Conference on Education and New Learning Technologies. International Association of Technology, Education and Development (IATED). Pág. 2034 – 2041 - ISBN: 978-84-695-3491-5. Universidad de Alcalá, Alcalá de Henares, España. Julio de 2012.
  - Sanz C., Zangara A., Otero N. (2008). “El trabajo colaborativo como espacio de reflexión teórica y práctica. El proceso de negociación visto desde el punto de vista cognitivo y desde la lógica de cada disciplina.” Conferencia Internacional ICDE 2008 (International Conference on Distance Education). Universidad del Caribe. Santo Domingo. República Dominicana.
  - Sanz C., Zangara A., Escobar Gutierrez M. “” Posibilidades educativas del entorno 3D Second Life para docentes universitarios. Caso de aplicación. Julio de 2012.
  - Toledo G., Gonzalez A., Malbrán M. “Accesibilidad digital para usuarios con limitaciones visuales”. VII Congreso de Tecnología en Educación y Educación en Tecnología. TE&ET. ISBN: 978-987-28186-0-9. Pág.371-380. Junín/Pergamino, Buenos Aires, Argentina, Junio 2012.
  - Zangara A., Llarena M., Ríos L. Maestría en Tecnología Informática aplicada en Educación “La comunidad virtual de práctica: un espacio de colaboración y reflexión para docentes de Matemática”. Expuesta en octubre de 2012.
  - Zangara A., Oltolina Giordano M. Propuesta de Maestría de Tecnología Informática aplicada en Educación “Formación Digital de alumnos de profesorado universitarios. La alternativa de las e-actividades”. Diciembre 2012.

## Las Tecnologías de la Información y la Comunicación en el ámbito de la Educación Especial.

Sanz Cecilia, Guisen Andrea, Moralejo Lucrecia, Artola Verónica, Caseres German, Baldassarri Sandra<sup>(1)</sup>, Pesado Patricia<sup>(2)</sup>

Instituto de Investigación en Informática LIDI (III-LIDI)  
Facultad de Informática – UNLP

{csanz, [aguisen](mailto:aguisen@unlp.edu.ar), [lmoralejo](mailto:lmoralejo@unlp.edu.ar), [varbola](mailto:varbola@unlp.edu.ar), [gcaseres](mailto:gcaseres@unlp.edu.ar), [ppesado](mailto:ppesado@unlp.edu.ar)}@lidi.info.unlp.edu.ar, sandra@unizar.es

### CONTEXTO

Las investigaciones aquí presentadas forman parte del proyecto “Tecnología y aplicaciones en Sistemas de Software Distribuidos. Experiencias en E-learning, E-government y Sistemas productivos”, del Instituto de Investigación en Informática LIDI, acreditado por la UNLP (período 2010-2013).

### RESUMEN

En el marco del proyecto mencionado, existe un subproyecto en el que se investigan, desarrollan y aplican tecnologías de la información y la comunicación de manera tal de favorecer procesos educativos en diversos contextos. Este subproyecto se vincula fuertemente con disciplinas como la Tecnología Educativa, la Informática, la Comunicación, y las Ciencias de la Educación, entre otras. Es por ello, que en él participa un equipo interdisciplinario, para poder considerar diferentes dimensiones de análisis de la temática. Aquí se presentan una serie de líneas de trabajo específicas en las que se viene avanzando, dentro de este subproyecto, vinculadas a favorecer procesos de enseñar y aprender en el ámbito de la Educación Especial. Algunos de estos avances se vinculan, por ejemplo, con personas con necesidades complejas de comunicación y otros con personas con dificultades motoras. Se detallan los objetivos en esta línea y cada uno de los trabajos realizados en pos de alcanzarlos.

**Palabras claves:** Tecnologías de la Información y la Comunicación, Educación especial, Assistive Technology

### INTRODUCCION

La reconceptualización de la Educación Especial a la luz de los procesos de integración, da cabida a la inclusión educativa, entendiéndola como una

interacción que se genera en el respeto hacia las diferencias individuales y las condiciones de participación desde la perspectiva de igualdad y equiparación de oportunidades sociales, así como cualquiera que sean los valores culturales, raza, sexo, edad y “condición” de la persona.

Son varios los autores (Blackhurst, 2000), (Boix, 2005), (Bouck, 2012), (Okolo, 2008), (Nochajski, 1999), que sostienen la importancia de la utilización de las Tecnologías de la Información y la Comunicación en el ámbito de la Educación Especial.

Según Muntaner (Muntaner, 2010), se destacan cuatro grandes ámbitos de actuación en esta perspectiva:

a) Ayudas técnicas para la realización de las actividades de la vida diaria, tales como movilidad, transporte, higiene, vestido o alimentación, entre otras.

b) La incorporación de los ordenadores para facilitar la comunicación y la interacción con los demás. El ordenador permitirá: por un lado, que una persona pueda expresarse verbalmente de forma oral y escrita; por otra parte, convertirse en un potenciador de la comunicación, en cuanto que favorece la interacción y el diálogo espontáneo, siempre y cuando que se utilicen las adaptaciones y el software adecuado.

c) La creación de redes computarizadas que permitan controlar el entorno: aparatos eléctricos o mecánicos del hogar o de cualquier vivienda. Estos sistemas permiten a las personas con discapacidad manejar de forma autónoma los elementos y aparatos de su entorno.

d) Las posibilidades de inserción laboral, tanto en el aspecto de la formación profesional, como en el desempeño de sus funciones laborales, ya sea en entornos normalizados, ya sea a través de experiencias de teletrabajo.

Sin embargo, según este mismo autor, es en el ámbito educativo donde las TIC están aportando

<sup>1</sup> Universidad de Zaragoza, España

<sup>2</sup> Comisión de Investigaciones Científicas de la Pcia. de Bs. As.



nuevas posibilidades para el logro de una escuela inclusiva.

En vinculación con estas temáticas, dentro de este proyecto, se ha puesto énfasis en diversos estilos de interacción persona-ordenador en atención a la diversidad. Se revisaron trabajos como los de (Battocchi, 2010), (Ong, 2011), (Pipper, 2006), (Veen, 2009), en los que se presentan aplicaciones de diferentes estilos y paradigmas de interacción (tales como realidad aumentada, interacción tangible) para el trabajo con alumnos con necesidades especiales.

Se detallarán en este trabajo los avances en relación a las investigaciones y desarrollos realizados, en el proyecto del III LIDI, en relación con las TIC en el ámbito de la Educación Especial. En la siguiente sección se presentan las líneas de I+D+I específicas que se abordan.

## **LINEAS DE INVESTIGACION / DESARROLLO**

Se mencionan aquí los dos ejes que se vienen abordando en relación a la temática de TIC y Educación Especial:

- Estilos y Paradigmas de interacción persona – ordenador atendiendo a la accesibilidad e inclusión:
  - Reconocimiento de voz en pos de lograr una comunicación alternativa con el ordenador
  - Realidad Virtual y Aumentada posibilidades para el ámbito de la discapacidad
  - Interacción tangible y multitáctil
- Herramientas colaborativas para la Educación Especial

En estos ejes se han estudiado y revisado antecedentes, se ha trabajado en el estado del arte de estas temáticas y se han realizado trabajos conjuntos con otras universidades e instituciones. En la siguiente sección se detallan los avances y resultados alcanzados.

## **RESULTADOS ESPERADOS/OBTENIDOS**

Se detallan a continuación algunos de los resultados obtenidos en el área de TIC y Educación Especial.

Herramientas desarrolladas:

- JClícVoice: se desarrolló una adaptación a la herramienta de autor JClíc<sup>2</sup>, que permite la construcción de actividades educativas. En particular, JClícVoice posibilita a los docentes construir actividades de asociación simple y compleja, puzzle de intercambio y puzzle de agujero, y que puedan ser resueltas utilizando comandos por voz. De esta manera, se orienta a alumnos con dificultades motrices y sin graves compromisos en el habla. La herramienta está disponible en <https://projectes.lafarga.cat/projects/jclivoice/downloads>. El software cuenta con un registro de propiedad intelectual. (Moralejo, 2012)

- ACoTI (Augmentative Communication thought Tangible Interaction): es una herramienta basada en el paradigma de interacción tangible, utilizando una superficie horizontal aumentada computacionalmente (tabletop), diseñada como ayuda tecnológica al desarrollo de competencias comunicacionales propias a la adquisición del lenguaje en potenciales alumnos usuario de Sistemas de comunicación Aumentativa y Alternativa (SAAC) de Alta tecnología A través de objetos cotidianos se controla e interactúa con un software educativo con el objetivo antes mencionado (Sanz, 2012)

- ECCA (Entorno Colaborativo de Comunicación Aumentativa): se trata de una red social educativa diseñada como ayuda tecnológica al entrenamiento de prácticas de CAA, mediante el diálogo en dinámicas didácticas colaborativas, orientadas al desarrollo de competencias comunicacionales de potenciales usuarios de SAAC de alta tecnología (Guisen, 2012).

En referencia a estos temas, una de las integrantes de este equipo ha obtenido una beca de la UNLP, en la que se estudian diferentes paradigmas de interacción persona-ordenador que puedan favorecer el uso de un ordenador por parte de alumnos de Educación Especial (Moralejo, 2011) (Moralejo, 2012). Esta becaria también ha presentado su propuesta de tesis de Magister durante 2012, y se trabaja en relación a la construcción de una herramienta de autor para actividades educativas que permitan diferentes estilos de interacción.

Se ha realizado también una tesis doctoral, en la que se estudió el diseño de entornos virtuales colaborativos que permitan el uso de herramientas

<sup>2</sup> JClíc, sitio oficial.

<http://clíc.xtec.cat/es/jclíc/index.htm>. Consultado en 2012

de comunicación alternativa y aumentativa (Guisen, 2011). Esta tesis ya se ha entregado y consta con jurado para su defensa.

También se ha trabajado en conjunto con grupos de investigación de la Universidad de Islas Baleares (Manresa, 2012) y de Zaragoza. Se ha avanzado en el diseño, desarrollo y evaluación de un software basado en interacción tangible para usuarios de Comunicación Aumentativa, en proceso de entrenamiento para la utilización de un Sistema de Comunicación Aumentativa de Alta Tecnología. Se ha trabajado con esta herramienta en instituciones de Educación Especial. (Baldassarri, 2011) (Sanz, 2012).

También, se dirige una tesis doctoral que aborda el uso de TIC para alumnos con deficiencias auditivas. La propuesta de esta tesis ha sido aprobada durante 2011, y actualmente se encuentra en realización (Fachal, 2011).

Las investigaciones y desarrollos realizados se nutren del contacto con profesionales del área de Educación Especial. Estos han colaborado aportando información sobre las necesidades específicas de los alumnos de escenario, opiniones sobre los sistemas diseñados, evaluaciones y contacto con grupos de alumnos que han participado de diferentes tests.

En cuanto a los proyectos vinculados con la temática, el III- LIDI participa en los siguientes:

- Se ha firmado un acuerdo de colaboración en estos temas con la Universidad de Zaragoza. En este marco una docente de dicha Universidad ha realizado una estadía de investigación en el III LIDI y dictado el curso de postgrado: “Paradigmas actuales de interacción” (Grupo de Informática Gráfica Avanzada).
- FRIVIG: Formación de Recursos Humanos e Investigación en el Área de Visión por Computador e Informática Gráfica. Acreditado AECID (Agencia Española de Cooperación). Código 027824/09. 2010-2011 – Se ha aprobado la continuación del proyecto para 2012-2013. En el marco de este proyecto hay una línea específica que se vincula con la evaluación de calidad de procesos educativos mediados por TIC y en la que se han realizado avances en relación a la temática de este trabajo.
- REDAUTI: Red Temática en Aplicaciones y Usabilidad de la Televisión Digital Interactiva. Compuesta por 170 investigadores de 28 grupos de investigación (22 universidades y 6 empresas) de 11 países iberoamericanos. Entidad financiadora: Programa Iberoamericano de Ciencia y Tecnología

para el Desarrollo (CYTED). Duración desde: Enero 2012 hasta: Diciembre 2014. Investigadora principal: Dra. María José Abásolo. En el marco de este proyecto, se aborda la construcción de contenidos para la TVDI que puedan ser de aporte también para el área de Educación Especial.

## FORMACION DE RECURSOS HUMANOS

Hay un becario de UNLP, que además está realizando la tesis de maestría en la temática, un becario de CONICET (con beca tipo II) que es doctorando y está por exponer su tesis de doctorado. Otra tesis de doctorado se está realizando en la temática. Hay 3 becarios alumnos del III-LIDI trabajando en el proyecto.

Se dirigen Tesis de Magister, Trabajos de Especialista y Tesinas de Licenciatura de la Facultad de Informática en temas relacionados con el Proyecto.

Hay 1 trabajo de grado que se ha desarrollado en vinculación con las temáticas aquí presentadas.

En la siguiente sección se presenta: la bibliografía y los trabajos citados aquí, algunos textos de estudio que se utilizan en la investigación, y por otra parte, algunas publicaciones que forman parte de los resultados de este proyecto.

## BIBLIOGRAFIA

- Baldassarri S., Marco J., Sanz C., Guisen A., De Giusti A., Cerezo E. (2011), “Interacción tangible para desarrollar competencias comunicacionales en educación especial“..XII Congreso Internacional de Interacción Persona-Ordenador. Lisboa, Portugal. Publicado en Actas del Congreso - ISBN: 978-84-9281-234-Págs. 341-346
- Battocchi A., Ben-Sasson A., Esposito G., Gal E. Pianesi F., Tomasini D., Venuti P., Weiss P. L. and Zancanaro M. (2010), “Collaborative Puzzle Game: a Tabletop Interface for Fostering Collaborative Skills in Children with Autism Spectrum Disorders”. *Journal of Assistive Technologies*. 4(1): pp. 4-14
- Blackhurst, A. E. & Lahm, E. A. (2000). *Foundations of technology and exceptionality*. En J. Lindsey (Ed.) *Technology and Exceptional Individuals* (3rd ed, pp. 45). Austin, TX: Pro-Ed.
- Boix J., Basil C. (2005). *CAA en atención temprana. Comunicación y pedagogía: NT y*

- recursos didácticos. ISSN: 1136-7733, N° 205, pp. 29-35.
- Bouck E., Maeda Y., Flanagan S. Assistive Technology and Students With High-incidence Disabilities: Understanding the Relationship Through the NLTS2 Remedial and Special Education September 1, 2012 33: 298-308.
  - Fachal A., Abásolo M.J., Sanz C. (2011). Propuesta de tesis doctoral: "Aprendizaje de Lengua de Señas Argentina (LSA). Un software educativo para el acompañamiento del aprendizaje en personas con necesidades especiales de audición". Expediente: 3300-003998-000. Facultad de Informática. UNLP.
  - Foster M. E., Avramides K., Bernardini S., Chen J., Frauenberger C., Lemon O., Porayska-Pomsta K. (2010), "Supporting children's social communication skills through interactive narratives with virtual characters. International Conference on Multimedia". pp. 1111-1114
  - Guisen A., Sanz C., De Giusti A. (2012). "ECCA: Augmentative Communication Collaborative Environment". International Conference on Collaboration Technologies and Systems. Denver, Colorado, USA. Proceedings of the 2012 International Conference on Collaboration Technologies and Systems (produced for IEEE by The Printing House, Inc) - ISBN 978-1-4673-1380-3 - Páginas 282-285. IEEE Catalog Number: CFP1216A-CDR
  - Guisen, A. Sanz, C. De Giusti, A. (2011). "ECCA: Entorno Colaborativo de Comunicación Aumentativa. Avances de Diseño". VI Congreso de Tecnología en Educación y Educación en Tecnología (TEyET 2011). Salta, Argentina. Actas del Congreso: ISBN 978-987-633-072-5. Págs. 54 -63
  - Manresa-Yee C., Muntaner J., Sanz C. (2012). "e-Inclusión Educativa para Alumnos con Graves Dificultades Motoras". III Congreso Iberoamericano sobre Calidad y Accesibilidad de la Formación Virtual CAFVIR 2012. Universidad de Alcalá, Alcalá de Henares, España. Actas del Congreso ISBN: 978-84-8138-367-6 - Páginas 97-104.
  - Marco J., Cerezo E., Baldassarri S. (2010). "Playing with toys on a tabletop active surface". IDC 2010: 296-299
  - Moralejo L., Ostermann E., Sanz C., Pesado P. "Voice command adaptation for Jclie, for the special education context". (2011). Publicado en el Libro "Computer Science & Technology Series – XVI Argentine Congress of Computer Science Selected Papers", Guillermo Simari (ed.), Hugo Padovani (ed.). ISBN 978-950-34-0713-4. Editorial: EDULP, Argentina, 2011. Pág.99-108
  - Moralejo L., Ostermann S., Sanz C., Pesado P. (2011), "Adaptación a Jclie para alumnos con deficiencia motriz, mediante comandos por voz". VI Congreso Iberoamericano de Tecnologías de Apoyo a la Discapacidad. Palma de Mallorca, España. Publicado en actas del congreso. Págs. 236 a 243. ISBN O.C: 978-84-8384-186-X. Volumen I: 978-84-8384-187-8
  - Moralejo L., Sanz C., Pesado P. (2012) "El reconocimiento de voz como paradigma de interacción para personas con dificultades motoras". XVIII Congreso Argentino de Ciencias de la Computación. Universidad Nacional del Sur, Bahía Blanca. Argentina. Anales del Congreso. Páginas 677-685 - ISBN 978-987-1648-34-4
  - Muntaner J. (2010). De la Integración a la Inclusión: un nuevo modelo educativo. V Congreso Internacional de Tecnología Educativa y Atención a la Diversidad.
  - Nochajski, S. M., Oddo, C., & Beaver, K. (1999). Technology and transition: Tools for success. Technology and Disability, 11, 93–101.
  - Okolo, C. M. (2008). Technology and individuals with mild disabilities. In J. D. Lindsey (Ed.), Technology and exceptional individuals (pp. 325–376). Austin, TX: PROED.
  - O'Malley, C., Fraser D.S. (2004), "Literature Review in Learning with Tangible Technologies". NESTA Futurelab.
  - Piper A. M., O'Brien E., Morris M. R., Winograd T. (2006), "SIDES: a cooperative tabletop computer game for social skills development". 20th Conference on Computer Supported Cooperative Work.
  - Ong S.K., Shen Y., Zhang J., and Nee A.Y.C.. (2011) "Augmented Reality in Assistive Technology and Rehabilitation Engineering". 2011. ISBN 978-1-4614-0063-9, pages 603 - 630
  - Salinas J. (2004). "Innovación docente y uso de las TIC en la enseñanza universitaria". Revista Universidad y Sociedad del Conocimiento. Vol.1, Nro.1- Noviembre de 2004.
  - Sanz C., Baldassarri S., Guisen A., Marco J., Cerezo E. De Giusti A. (2012). "ACoTI: herramienta de interacción tangible para el desarrollo de competencias comunicacionales en usuarios de comunicación alternativa. Primeros resultados de su evaluación". VII Congreso de Tecnología en Educación y

Educación en Tecnología (TEyET 2012). Junín, Buenos Aires, Argentina. Actas del Congreso - ISBN 978-987-28186-0-9. Págs. 226-233

- Veen, M. van. (2009), "Improving collaboration with raketeer: development of a serious game with multi-touch interaction for teaching children with PDD-NOS collaboration". Doctoral thesis. Rijksuniversiteit Groningen



## Plataforma de hardware de bajo costo para robótica educativa

### Resultados de la primera etapa

Gonzalo Zabala, Sebastián Blanco, Ricardo Morán, Matías Teragni

Centro de Altos Estudios en Tecnología Informática – Universidad Abierta  
Interamericana

Montes de Oca 745, CABA

(5411)43015323

[gonzalo.zabala@uai.edu.ar](mailto:gonzalo.zabala@uai.edu.ar), [sebastian.blanco@uai.edu.ar](mailto:sebastian.blanco@uai.edu.ar), [ricardo.moran@uai.edu.ar](mailto:ricardo.moran@uai.edu.ar), [matias.teragni@uai.edu.ar](mailto:matias.teragni@uai.edu.ar)

#### Resumen

El objetivo general del presente proyecto es el desarrollo de una plataforma de hardware de bajo costo que pueda ser utilizada en escuelas primarias y secundarias para el aprendizaje de robótica. Se presentan los resultados de la primera etapa, donde se tomaron las decisiones del hardware a utilizar y se desarrollaron los primeros prototipos del adquisidor de datos.

#### Contexto

El presente proyecto será radicado en el Centro de Altos Estudios en Tecnología Informática (CAETI), dependiente de la Facultad de Tecnología Informática de la Universidad Abierta Interamericana. El mismo se encuentra inserto en la línea de investigación “Sociedad del conocimiento y Tecnologías aplicadas a la Educación”. El financiamiento está dado por la misma Universidad Abierta Interamericana.

#### Introducción

En los últimos años, la inserción de recursos tecnológicos en las aulas de nuestro país ha crecido exponencialmente. Un ejemplo de esto es el proyecto “Conectar Igualdad”, que entregará en el transcurso de

su ejecución, 3 millones de netbooks a los estudiantes de nivel medio de todo el país. La presencia masiva de estos recursos abren las puertas a diversas formas de aprovechamiento. Nos encontramos con un dispositivo que puede servir como procesador de sistemas de adquisición de datos, de nodo en una estructura de red de comunicaciones o como controlador en un artefacto robótico, entre otros ejemplos. El objetivo de este proyecto es el desarrollo de un controlador de motores y un hub de sensores, ambos de bajo costo, junto con la plataforma de software para la programación, de manera tal que se puede convertir a la netbook en un adquisidor de datos y un pequeño robot móvil.

#### Líneas de investigación y desarrollo

Nuestra propuesta está dividida en dos etapas:

##### a) Adquisidor de datos

En esta etapa el objetivo es el desarrollo de un hub que permita conectar en forma sencilla diferentes sensores a la netbook, sirviendo de interfaz entre los sensores y el puerto usb. Además debemos seleccionar sensores de bajo costo y crear la electrónica y el encapsulado necesario para su uso en ámbitos educativos. Por último, con respecto

al software tendremos que realizar el firmware del hub para la interpretación de los datos de los sensores, el software para la lectura, graficación, bajada de datos (que será una especialización de nuestro trabajo anterior, Physical Etoys) y un sistema de reconocimiento de imágenes para el uso de la cámara incorporada en las netbooks.

#### b) Kit de robótica

En este caso, desarrollaremos un hub para la conexión de hasta dos motores de corriente continua y de una alimentación externa de batería. Con respecto al software, tendremos que realizar el firmware del hub y agregar a Physical Etoys el control de los motores. También tendremos que diseñar un chasis para poder insertar la netbook, los motores, las ruedas, los sensores y la alimentación.

### Resultados y Objetivos

En esta primera etapa hemos realizado un análisis de los tipos de microcontroladores presentes en el mercado que pudieran ofrecer una solución a nuestro desarrollo. Finalmente hemos seleccionado el PIC 18F4550. Entre las características que posee, las que más nos interesaron fueron su memoria flash de 32K, las 13 entradas analógicas y la conexión USB en forma directa (para conectarlo con la netbook sin electrónica adicional).

A partir de este micro desarrollamos la electrónica adicional para la adquisición de datos. A la fecha podemos leer diferentes sensores sencillos desde las entradas analógicas, como potenciómetros, CNY70, ldr y switches. Estamos programando la interfaz para la lectura de los datos desde Physical Etoys, la plataforma de software que usaremos en educación que también fue desarrollada por nuestro equipo.

Con respecto al uso de las entradas que tienen las netbooks, desarrollamos una pequeña librería de análisis de imágenes dedicada a segmentar una foto para ubicar zonas de interés dentro de la misma basándose principalmente en los colores presentes en ellas. Existe actualmente un prototipo funcional dentro de Physical Etoys que permite acceder esta funcionalidad.

Contamos con un objeto de forma rectangular al cual se le puede asignar mediante una propiedad al origen de la información a rastrear, y este rectángulo tomará la posición y tamaño necesarios para englobar la zona de interés dentro de la imagen, permitiendo fácilmente actuar en consecuencia a la posición de objetos gráficos dentro de la pantalla, o de objetos físicos mediante el uso de una cámara.

Finalizada esta primera etapa, queda por delante el desarrollo del hub de motores para controlarlos desde la netbook, la elección de las baterías a utilizar y la incorporación de las herramientas necesarias a Physical Etoys para poder leer los sensores y programar el comportamiento de los motores de una forma didáctica y sencilla.

### Formación de Recursos humanos

El equipo de trabajo estará conformado por un investigador adjunto del Centro de Altos Estudios en Tecnología Informática (CAETI) quien ejercerá el rol de director del proyecto y tres ayudantes alumnos de la Facultad de Tecnología Informática de la Universidad Abierta Interamericana.

### Referencias principales

- Builes, J.A.J., Ovalle, D.A. & Ochoa, J.F., Propuesta de una Plataforma para la Difusión de la Robótica Móvil: E-SMART

Proposal for a Platform to the Diffusion of Mobile Robotics: E-SMART.

- Muñoz, N.D., Andrade, C.A. & Londoño Ospina, N., 2006. Diseño y construcción de un robot móvil orientado a la enseñanza e investigación. Ingeniería y desarrollo, (19), págs.114–127.

- Valera, A. et al., DESARROLLO Y CONTROL DE ROBOTS MÓVILES DE BAJO COSTE. IEEE Robotics and Automation Magazine, 11, pág.13.

## Difusión de la actividad académica a través de la integración de entornos virtuales de aprendizaje con redes sociales, sistemas académicos y repositorios digitales

Javier Díaz, Alejandra Schiavoni, Alejandra Osorio,  
Paola Amadeo, M. Emilia Charnelli  
Laboratorio de Investigación en Nuevas Tecnologías Informáticas.  
Facultad de Informática. Universidad Nacional de La Plata  
Calle 50 esq. 120, 2do Piso  
+54 221 4223528  
jdiaz@unlp.edu.ar, {ales, aosorio}@info.unlp.edu.ar,  
{pamadeo, mcharnelli}@linti.unlp.edu.ar

### Resumen

El siguiente artículo presenta las líneas de investigación que se llevan a cabo y los resultados obtenidos en un proyecto global que se desarrolla desde el año 2011 en el Laboratorio de Investigación en Nuevas Tecnologías Informática, LINTI, de la Facultad de Informática de la UNLP. Este proyecto comprende la conformación de un Repositorio Digital de Acceso Abierto a partir de material académico; en relación a esto la estandarización del contenido siguiendo normas preestablecidas y, como objetivo global la integración de la plataforma virtual de aprendizaje (Learning Management System-LMS) Moodle con otras plataformas y aplicaciones.

El uso de una plataforma virtual de aprendizaje favorece el proceso de e-learning en el ámbito educativo. Sin embargo, dadas las nuevas tendencias en la organización de contenido y los nuevos hábitos sociales, resulta útil expandir su funcionalidad integrándola con otras plataformas y aplicaciones. De esta manera, es que se implementó la comunicación de la plataforma Moodle con un conjunto de herramientas diversas: el repositorio DSpace para consultar y acceder a los recursos en forma directa, la red social Twitter en pos de mejorar la interacción con los alumnos usuarios de la plataforma educativa y el sistema de administración de alumnos SIU-Guaraní.

Es importante destacar que en el campo de e-learning, los objetos de aprendizaje (OA) representan una base fundamental de la interoperabilidad, razón por la cual la integración de un LMS con Repositorios de Objetos de Aprendizaje (ROA's) permite incrementar el uso de los OAs en la planificación y construcción de contenido educativo por parte de los docentes [1]

**Palabras clave:** Objetos de Aprendizaje, Entornos Virtuales de Aprendizaje Repositorios, Redes Sociales

### Introducción

Los estándares para el desarrollo del e-learning están marcando la pauta para crear sistemas que integren las aplicaciones para los procesos de enseñanza y aprendizaje en línea, en las que los contenidos puedan ser reutilizados y compartidos, entre personas y entre sistemas. A estos contenidos se les conoce como Objetos de Aprendizaje, OAs, que representan unidades de aprendizaje, independientes y autónomas, descritas por metadatos. Los OAs tienen características particulares que dan capacidades y funcionalidades a los sistemas de gestión de aprendizaje, principalmente desde el punto de vista de la organización y reutilización de recursos. Una vez creados los objetos de aprendizaje, éstos se almacenan en un Repositorio de Objetos de Aprendizaje. Actualmente se constituyen como



importantes herramientas que pueden almacenar y distribuir los OAs, para los usuarios y las aplicaciones que los requieran. Para lograr una mayor integración y reuso de estos objetos se está implementando la integración entre un entorno virtual de aprendizaje y el repositorio.

Por su parte las redes sociales forman parte de la vida cotidiana de las personas y de los estudiantes de la Facultad en particular. Como menciona Buckingham et.al (2012) *Internet ya no es una mera herramienta para distribuir y recuperar información sino un instrumento para dialogar y compartir, para la comunicación interpersonal y para el entretenimiento.* [2] Esto sumado a los teléfonos celulares con acceso a Internet al alcance de todo público han colocado a las redes sociales al alcance de muchas personas, con la publicación de blogs y la distribución digital.

Lograr la interoperabilidad entre todos estos sistemas junto con distintas aplicaciones de gestión académica es uno de los objetivos del proyecto que se presenta en este artículo.

### Contexto

La Facultad de Informática de la Universidad Nacional de La Plata cuenta con más de 4.151 alumnos, 3 carreras de grado, con un promedio de 800 ingresantes anuales [3].

En la Facultad se utilizan distintos entornos virtuales de aprendizaje como complemento de las clases presenciales, como Moodle [4], WebUNLP [5], GoogleGroups [6] entre otros. En particular Moodle, entorno virtual de aprendizaje de software libre, se utiliza en numerosas materias que se dictan en distintos años de las carreras, desde hace más de 7 años, <http://catedras.info.unlp.edu.ar>. Involucra más de 5000 alumnos y más de 180 cursos. Este entorno virtual cuenta con una gran cantidad de material educativo digitalizado almacenado como recursos. En general, el material de cada asignatura está compuesto de unidades teóricas y prácticas que

incluyen simuladores, sesiones de chats, videos, audio, archivos ejecutables, entre otros. Las actividades utilizadas en la plataforma son variadas, como fotos, autoevaluaciones, wikis y en particular la entrega de trabajos a través del módulo Tareas. Es decir, no sólo almacena material generado por los docentes sino también material generado por los propios alumnos, muchas veces con las correspondientes calificaciones y comentarios de los docentes.

La Facultad de Informática de la UNLP está haciendo uso también de las redes sociales para poder llegar a los distintos sectores de la sociedad que están vinculados con ella y de esta manera realizar una mejor difusión de las actividades que se llevan a cabo. En particular dispone de un perfil en Facebook,

<http://www.facebook.com/facultad.informatica.unlp> y distintas cuentas de difusión académica como @InformaticaUNLP, @infounlp entre otras. Además ciertas cátedras utilizan estos canales para comunicarse con sus estudiantes, como la cátedra de Algoritmos y Estructuras de Datos @ayed\_fi, Introducción a los Sistemas Operativos @iso\_info\_unlp, Sistemas Operativos, Tecnologías Aplicadas a Business Intelligence @catedraBI, entre otros.

En la UNLP se usa el sistema de gestión académica SIU-Guaraní, que permite la gestión de alumnos dentro de cada facultad, desde que ingresan hasta que se diplomán, y provee una serie de controles y pistas de auditoría para los organismos de gestión de títulos. Fue desarrollado por el Ministerio de Educación de la Nación [7] para todas las universidades nacionales. Por otra parte, el creciente uso de los sistemas informáticos en las organizaciones para llevar adelante la gestión en general, el creciente uso de las redes sociales por parte de los jóvenes y la necesidad de complementar la información que se origina en sistemas heterogéneos, cuyo desarrollo aislado representa un problema de acoplamiento, nos lleva a trabajar en cuestiones de

interoperabilidad entre los sistemas, de acuerdo a los protocolos y APIs disponibles para alcanzarla.

### **Líneas de Investigación y Desarrollo**

La tarea de investigación que estamos llevando a cabo comprende varias líneas de trabajo todas ellas en el área de e-learning. Por un lado continuamos con el análisis del proceso de diseño y construcción de objetos de aprendizaje, y su aplicación en casos concretos. A partir de las metodologías de diseño que veníamos estudiando [8], se avanzó a la siguiente instancia de evaluación de herramientas para la construcción de los OAs. Con el objetivo de obtener OAs reusables y compatibles con diferentes plataformas de gestión de aprendizaje, se eligió el estándar SCORM y se analizaron dos herramientas muy populares y que ofrecen ventajas y funcionalidades específicas: Reload Editor [9] y eXe Learning [10].

Otra de las líneas de investigación está basada en la integración de Moodle con repositorios digitales, sistemas académicos y redes sociales.

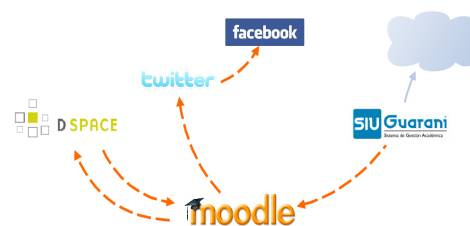
Respecto a la integración de Moodle con repositorios, en principio se comenzó a trabajar en la comunicación de Dspace hacia Moodle. Se integró la versión 2.0 del LMS con la versión 1.8. del repositorio a través de la API Rest para poder consultar e importar recursos almacenados en este último y poder incorporarlos en un curso de la plataforma.

Teniendo en cuenta el otro canal de la comunicación, se están analizando otros protocolos de comunicación con el objetivo de implementar la publicación de contenidos desde la plataforma Moodle hacia distintos repositorios. Esta nueva funcionalidad resulta muy interesante para exportar toda clase de material publicado en los cursos, tales como clases teóricas y prácticas, documentación relacionada, trabajos realizados y entregados por alumnos, etc. Además, es posible la exportación de contenido estandarizado como objetos de aprendizaje.

En particular se está analizando el protocolo SWORD [11] para el depósito de recursos en los repositorios digitales y del estándar METS [12] para codificación de los metadatos. Esta transferencia de contenido requiere del estudio de la estructura interna de la información dentro de la plataforma de aprendizaje para su recuperación y adaptación para el envío respetando los estándares establecidos.

Como se mencionó previamente, los jóvenes utilizan asiduamente las redes sociales y es allí donde se desenvuelven y establecen y refuerzan sus vínculos sociales. Esta realidad motivó el desarrollo de un módulo para integrar la plataforma virtual Moodle con las redes sociales, extendiendo la funcionalidad propia de la plataforma en pos de mejorar la comunicación con los estudiantes a través de sus hábitos sociales.

Durante el año 2012, se realizó la integración de este sistema con la versión 2.0 de la plataforma Moodle recientemente instalada, permitiendo hacer una sincronización entre el sistema administrativo y la herramienta educativa con el objetivo de compartir cursos y usuarios. Las plataformas y sistemas que forman parte del proyecto de integración con Moodle se resumen en la Fig. 1.



**Fig. 1.** Plataformas y sistemas que se integran con Moodle

### **Resultados y Objetivos**

Este proyecto de investigación persigue distintos objetivos que responden a las diferentes líneas de trabajo.

Por un lado destacamos la estandarización del material educativo que se genera a través de la creación de objetos de aprendizaje que resultan portables y pueden ser utilizados en varios contextos.

Un objetivo de gran importancia es la integración de distintas plataformas del ámbito académico lo que permite aumentar la disponibilidad de los recursos y mejorar la interacción entre los distintos actores que intervienen.

Teniendo en cuenta estos objetivos se han alcanzado los siguientes resultados:

- 1) Construcción de objetos de aprendizaje aplicando dos herramientas específicas: Reload Editor y eXe Learning. Se analizaron y evaluaron las ventajas de cada una de ellas en un caso concreto. Este análisis fue expuesto en detalle en el VII Congreso TE&ET Tecnología en Educación y Educación en Tecnología [13]
- 2) Evaluación de los objetos de aprendizaje del Curso sobre Accesibilidad Web, siguiendo una metodología de evaluación específica [14] y cuyos resultados fueron descriptos en 5<sup>th</sup> Annual International Symposium Emerging Technologies for Online Learning [15]
- 3) Desarrollo de un módulo para la integración de la plataforma Moodle con la red social Twitter. Este módulo está disponible en la comunidad de Moodle y puede ser accedido en: <http://tracker.moodle.org/browse/CONTRIB-3124> [16]. Se desarrolló un Taller en el evento MoodleMoot 2012 en que se explicó la forma de instalación, configuración y uso del módulo [17].
- 4) Diseño de un prototipo de módulo utilizando el protocolo SWORD , para la comunicación desde Moodle hacia DSpace, con el objetivo de transferir contenido y almacenarlo en el repositorio [18]
- 5) Actualización de la plataforma Moodle a la versión 2.2, para su integración con el resto de las plataformas
- 6) Instalación y configuración de la versión 3.1 de DSpace, para

utilizarla como repositorio de la Facultad.

- 7) Adaptación y configuración de la interfaz de DSpace 3.1 con un nuevo diseño para representar a la Facultad

Para este año se plantea continuar con los objetivos propuestos en el proyecto, extendiendo el módulo de comunicación de la plataforma de aprendizaje con el repositorio de manera de ampliar su funcionalidad respecto al contenido a transferir. Además, se prevé estudiar la posibilidad de comunicación con otros repositorios.

También está previsto realizar modificaciones en el módulo desarrollado para la integración con Twitter, en pos de mejorar la difusión y recuperación de la información dentro de la red social.

### **Formación de Recursos Humanos**

El equipo de trabajo se encuentra formado inicialmente por dos profesoras, y una estudiante avanzada de la carrera de Licenciatura en Informática, quien a partir de octubre de 2012 obtuvo una Beca de Entrenamiento en la Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC). El plan de trabajo para el desarrollo de la beca comprende algunas de las líneas de este proyecto. Desde septiembre de 2012 se incorporaron dos alumnos avanzados de la carrera Licenciatura en Sistemas.

La participación en eventos de la especialidad, como TE&ET, desarrollado en junio de 2012 en Pergamino y MoodleMoot desarrollado en Montevideo en noviembre del año 2012, ha permitido seguir estableciendo canales de comunicación con otros investigadores que trabajan en las mismas áreas.

### **Referencias**

- [1] A. P. Rodrigues, L. M. R. Tarouco, M. A. R. Schmitt, "The use of Web Services as a strategy for integration of LORs to LMSs",

CACIC 2011, XVII Congreso Argentino de Ciencias de la Computación, La Plata, 2011

[2]

<http://arsenopirita.boumort.cesca.cat/index.php/comunicar/article/viewFile/20712/11257>

[3] Anuario 2012 UNLP <http://unlp.edu.ar>

[4] <http://moodle.org>

[5] <http://webunlp.unlp.edu.ar/>

[6] <http://google.com/groups>

[7] Ministerio de Educación de la Nación Argentina <http://me.gov.ar>. Última consulta mayo 2012.

[8] J. Díaz, A. Schiavoni, P. Amadeo, E. Charnelli, “Diseño y construcción de objetos de aprendizaje. Su integración en repositorios y plataformas virtuales de aprendizaje”, WICC 2012, XIV Workshop de Investigadores en Ciencias de la Computación, Posadas, Misiones, 26 y 27 de Abril, 2012.

[9] Reload Editor,

<http://www.reload.ac.uk/editor.html>

[10] eXe Learning, <http://exelearning.org>

[11] SWORD: Simple Web-service Offering Repository Deposit, <http://swordapp.org/>

[12] METS,

<http://www.loc.gov/standards/mets/>

[13] J. Díaz, A. Schiavoni, P. Amadeo, E. Charnelli, “Construyendo Objetos de Aprendizaje utilizando estándares abiertos y sistemas open source. Una experiencia sobre un curso de Accesibilidad Web”, VII Congreso TE&ET Tecnología en Educación y Educación en Tecnología, Pergamino, 11 y 12 de Junio, 2012.

[14] Ruiz González, R.E., Muñoz Arteaga, J., Álvarez Rodríguez, F. Formato para la Determinación de la Calidad en los Objetos de Aprendizaje. Primera Conferencia Latinoamericana de Objetos de Aprendizaje, LACLO. Coordinan: Repositorio de Conocimiento Europeo (ARIADNE), Corporación Latinoamericana de Redes Avanzadas (CLARA). Guayaquil, Ecuador, 2006.

[15] J. Díaz, A. Schiavoni, P. Amadeo, E. Charnelli, “Web Accessibility Learning Objects Development and Evaluation: a Case Study”, 5th Annual International Symposium Emerging Technologies for Online Learning. A Joint Symposium of the Sloan Consortium and Merlot, Las Vegas, NV, United States, 25-27 Julio, 2012.

[16] J. Díaz, A. Schiavoni, A. Osorio, P. Amadeo, E. Charnelli. “Integración de Plataformas Virtuales de Aprendizaje, Redes Sociales y Sistemas Académicos Basados en

Software Libre. Una Experiencia en la Facultad de Informática de la UNLP”. SSI 2012 Simposio sobre la Sociedad de la Información, JAIIO 2012, 27 – 31 Agosto 2012.

[http://www.41jaiio.org.ar/SSI\\_Contribuciones](http://www.41jaiio.org.ar/SSI_Contribuciones)

[17] J. Díaz, A. Schiavoni, A. Osorio, P. Amadeo, E. Charnelli. “Integración de Moodle con las redes sociales. Presentación del módulo Twitter Activity Module y su uso dentro de la plataforma”. MoodleMoot Uruguay, Montevideo, 22 y 23 de Noviembre de 2012.

[18] J. Díaz, A. Schiavoni, P. Amadeo, E. Charnelli. “Extending a Virtual Learning Platform through its Integration with a Digital Repository”. 7<sup>th</sup> International Technology, Education and Development Conference, Valencia, España, 4-6 Marzo 2013.



# Metodologías de enseñanza interactiva para entornos virtuales

Graciela Santos<sup>1</sup>, María José Bouciguez<sup>14</sup>, Andrea Miranda<sup>1</sup>, Gabriela Cenich<sup>1</sup>, Sebastián Barbieri<sup>1</sup>, María José Abásolo<sup>23</sup>

<sup>1</sup>Núcleo de Educación en Ciencias con Tecnologías (ECienTec)

Facultad de Ciencias Exactas – Universidad Nacional del Centro de la Provincia de Buenos Aires (UNICEN)

Campus Universitario. Paraje Arroyo Seco (7000) Tandil, Buenos Aires, Argentina  
{nsantos, mjbouci, amiranda, gabcen, sbarbier}@exa.unicen.edu.ar

<sup>2</sup>Instituto de Investigación en Informática LIDI (III-LIDI)

Facultad de Informática – Universidad Nacional de La Plata (UNLP)

calle 50 y 120 (1900) La Plata, Buenos Aires, Argentina

[mjabasolo@lidi.info.unlp.edu.ar](mailto:mjabasolo@lidi.info.unlp.edu.ar)

<sup>3</sup>Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CICPBA)

<sup>4</sup>Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)

## Resumen

Esta línea de investigación se orienta al estudio de las posibilidades de las tecnologías informáticas para definir metodologías de enseñanza interactivas y desarrollar aplicaciones informáticas basadas en tecnologías como videojuegos, simulaciones, mundos 3D, Televisión Digital Interactiva (TVDi), etc. tendientes a mejorar la calidad de la educación y la formación docente para ciencias.

**Palabras Clave:** Aprendizaje constructivista, Innovación pedagógica, Videojuegos, Simulación, TVDi

## Contexto

Este proyecto de investigación e innovación pedagógica se encuadra en el Proyecto “Enseñanza de las ciencias y Tecnologías. La influencia de las interacciones sociales, cognitivas y digitales en las prácticas educativas” del Núcleo de Educación en Ciencias con Tecnología acreditado por la

UNICEN en el marco del Programa de Incentivos. Además, se integra a la *Red temática en Aplicaciones y Usabilidad de la Televisión digital Interactiva* (RedAUTI), financiada por la Programa Iberoamericano de Ciencia y Tecnología para el Desarrollo (CYTED), a fin de realizar aportes sobre nuevas metodologías para TVDi.

## Introducción

Esta línea de investigación se inscribe en el campo de la tecnología educativa y el interés primario se orienta al estudio de las posibilidades o *affordance* [1] de las tecnologías informáticas para definir metodologías de enseñanza interactivas y desarrollar aplicaciones informáticas basadas en tecnologías (videojuegos, simulaciones, mundos 3D, TVDi, etc.) tendientes a mejorar la calidad de la educación en ciencias.

## Aprendizaje mediado por tecnología

El estudio de los factores que caracterizan la usabilidad de las aplicaciones interactivas se

analizarán desde el enfoque de la génesis instrumental [2].

Las estrategias de enseñanza se enmarcan en los principios del aprendizaje constructivista conjugando los conocimientos disciplinar, pedagógico y tecnológico siguiendo los lineamientos las teorías de *Enseñanza para la comprensión* [3] [4], *Enseñanza por competencias* [5] y *Conocimiento tecnológico pedagógico disciplinar* (TPACK acrónimo para Technological pedagogical content knowledge) [6].

El aprendizaje mediado por una aplicación informática y en un ambiente posibilitador de intercambios de ideas promueve el desarrollo de capacidades cognitivas superiores. Al interactuar “con otros o con artefactos culturales” se pone en juego un conjunto de cogniciones distribuidas de los sujetos, los instrumentos y la comunidad de aprendizaje que incentivan las competencias del individuo [7]. El enfoque de la génesis instrumental [2] permite considerar la integración de los objetos en la estructura de las actividades humanas y la apropiación de los artefactos como resultado de las transformaciones en el desarrollo de los objetos, las personas y las interacciones sociales. Un proceso de acomodación y asimilación por el que una persona convierte un artefacto en instrumento.

La manipulación de objetos virtuales o de representaciones digitales de la realidad favorece el desarrollo del pensamiento analítico al instar a la reflexión y el debate con otros para comprender la situación representada. Los instrumentos informáticos permiten la visualización de situaciones complejas para la comprensión de conceptos y propiedades de física o matemática; el procesamiento de representaciones; capacidad equivalente a la exteriorización de las funciones cognitivas; la construcción y verificación de modelos, etc. Pero más allá de las potencialidades interactivas de las nuevas tecnologías, el problema central en el desarrollo de aplicaciones para educación es la brecha entre la práctica real y la potencial o esperada [8].

En la actualidad en nuestro país el Programa “Conectar Igualdad.com.ar” ha modificado el

escenario tecnológico en el aula. Las potencialidades más relevantes de las TIC tales como su capacidad para diseñar nuevas formas de enseñanza [9], usar herramientas de pensamiento, posibilitar el acceso e interacción en la TVDi, ejercitar la creatividad y recurrir a almacenes de información, saberes y datos [10] plantean nuevos desafíos para los profesores. Sin embargo, existen evidencias acerca de las dificultades para modificar la práctica docente [11] [12]. La incorporación de novedosos recursos como en este caso la inmersión de la computadora al aula no conlleva, necesariamente, innovaciones o cambios relevantes en la práctica educativa [13] [10]. En este marco de dificultad que presenta la modificación de la práctica y los desafíos educativos se hace necesario explorar las actividades de enseñanza y aprendizaje con TIC que desarrollan los alumnos y docentes.

#### **Entornos virtuales interactivos educativos: Videojuegos y simulaciones**

Las simulaciones y los videojuegos son aplicaciones interactivas [14]. Al hablar de interactividad o interacción se refiere a cuando el software acepta y responde a las entradas de los usuarios finales, tales como datos o comandos [15].

La función pedagógica de la simulación tiene mucho en común con el rol de la experimentación, cuyo valor es muy apreciado en campos como el de la enseñanza de la Física, la Química, etc. En todos estos casos, la simulación permite “experimentar” en condiciones que, muchas veces, resultan difíciles de generar en el laboratorio. Se trata de programas que permiten un alto grado de interacción del alumno con el fenómeno representado.

Una de las funciones educativas de las simulaciones es la representación dinámica del funcionamiento de un sistema [16], mediante la visualización de procesos, mostrando la evolución del sistema representado y la interacción entre los componentes o consecuencias de tales interacciones dinámica. La posibilidad de manipular y transformar objetos en el espacio de la interfaz, comprometiendo un conjunto diferente de

competencias cognitivas, le otorga a las actividades educativas cierto grado de fortalecimiento pedagógico [17].

Un juego se ha definido como "... un entorno de aprendizaje interactivo participativo que cautiva a un jugador ofreciendo desafíos que requieren mayores niveles de dominio ..." [18].

Las simulaciones comparten características con los juegos, incluyendo el uso de un mundo virtual y el enfoque en un objetivo concreto, pero utilizan un más refinado conjunto de reglas, retos y estrategias para guiar a los participantes en el desarrollo de determinadas conductas y competencias que son altamente transferibles [19].

### **Nuevos medios: TVDi**

En este contexto, se pretende trabajar en el desarrollo de aplicaciones interactivas para TVDi enfocadas principalmente a apoyar los procesos de enseñanza y aprendizaje. En contraste con la televisión tradicional, que codifica los datos de manera analógica, la televisión digital codifica imagen y sonido de forma digital permitiendo entre otras cosas la posibilidad de crear una comunicación bidireccional productor de contenidos y el consumidor. La Televisión Digital abre la oportunidad de crear aplicaciones interactivas, de ahí la denominación de Televisión Digital Interactiva (TVDi) - en inglés Interactive Digital TV (IDTV). El consumidor puede pasar de ser un televidente o espectador pasivo a convertirse en un participante activo. La interactividad permite un diálogo entre el ordenador y el participante que tiene la capacidad de intervenir en los programas o servicios que recibe. A través del televisor será posible acceder a un conjunto de servicios públicos o privados. En particular, el término t-learning se utiliza con el significado de aprendizaje interactivo a través de un televisor. Este nuevo medio exige un estudio especial dado que no puede tomarse al t-learning como una mera adaptación de e-learning. Esto es debido a que por muchos años la actitud del consumidor frente al televisor ha sido pasiva, y más que buscar en la TVDi una educación formal el participante deberá involucrarse en

una experiencia educativa a partir del entretenimiento.

En Argentina, se creó en el año 2009 el Sistema Argentino de Televisión Digital Terrestre, y el Consejo Asesor de la Televisión Digital Terrestre con el objetivo de guiar el proceso de migración tecnológica hacia la televisión digital en la República Argentina hasta el 'apagón analógico' previsto en el 2019, contemplando la inclusión de todos los habitantes de la Nación a las nuevas tecnologías de la Sociedad de la Información. El Estado puso en marcha un conjunto de iniciativas para la promoción y producción de contenidos con el propósito de establecer un nuevo paradigma que incluya un sentido federal geográficamente y un sentido democrático desde la participación de distintos actores de la producción como organizaciones sociales sin fines de lucro, pequeñas productoras, cooperativas de trabajo, universidades y otros centros educativos [20].

La aparición de servicios de televisión digital interactivos conduce a un clima de mayor competencia, en el que los proveedores de contenidos ofrecen servicios cada vez más innovadores. Muchos de estos servicios requieren capacidades de inserción de texto. Surge un nuevo problema que emana del hecho de que el principal dispositivo de interacción en estos servicios es el control remoto, elemento que no ha sido creado originalmente para la escritura de texto. En colaboración con la Universidad de Oviedo en el marco de la RedAUTI, se persigue encontrar métodos efectivos de escritura de texto para entornos de Televisión Digital Interactiva mediante un control remoto convencional [21]

### **Líneas de investigación y desarrollo**

- Ambientes virtuales altamente interactivos (HIVEs, highly interactive virtual environments) para el aprendizaje en contextos educativos
- Las interacciones con tecnología: aspectos mediacionales y semióticos
- Interactividad en TVDi
- Estrategias de enseñanza interactiva para TVDi

- La formación docente inicial y continua en ambientes de enseñanza y aprendizaje mediados por tecnologías

### Resultados obtenidos y esperados

Los aportes tienden a colaborar en el diseño de aplicaciones educativas transparentes, para ser asimilados prontamente por los usuarios como instrumentos de mediación en el aprendizaje.

Se ha realizado:

- Descripción de las interacciones sociales, cognitivas y digitales en contexto de aprendizaje colaborativo y situado que permita comprender la construcción de conocimiento en procesos de prácticas socialmente significativas.
- Construcción de instrumentos de análisis de protocolos de interactividad (digital, social y de contenido escolar)
- Elaboración de estrategias de formación docente (en servicio y pre-servicio) en el diseño y seguimiento de cursos colaborativos online en base a los resultados de investigaciones.
- Modelado de Laboratorios de Informática Escolar, maximizando los recursos de hardware y software disponibles.

Entre los resultados esperados se mencionan:

- Desarrollo un Framework para el desarrollo de simulaciones y videojuegos para la enseñanza de física.
- Realización de pruebas con grupos poblacionales diferentes (estudiantes universitarios y personas sin estudios técnicos) de los diferentes modelos de teclados virtuales para ingreso de texto en aplicaciones de TVDi
- Desarrollo de aplicación de TVDI interactiva para la prueba de metodologías de enseñanza y diversas formas de interacción
- Desarrollo de propuestas de enseñanza basadas en modelos a la vez que elaborar recomendaciones para el diseño de planes de formación de docentes en TIC

### Formación de recursos humanos

Dado que este proyecto se inscribe en un campo de estudio interdisciplinario en la que

confluyen las áreas de la Computación, las Ciencias y la Educación se propone como uno de los objetivos prioritarios la formación de recursos humanos con perfiles acordes tanto por la especificidad de la formación como de la problemática de investigación que se aborde. Para dar cumplimiento a este objetivo actualmente hay en desarrollo:

1 tesis de doctorado en Ciencias Informáticas con beca CONICET de posgrado tipo II

- Ambientes Virtuales Altamente Interactivos para la enseñanza de ciencias

2 tesis de doctorado en Ciencias de la Educación

- uso de simulaciones en clases de ciencias
- formación docente en TIC

1 tesina del Profesorado de informática

- t-learning

1 estadia de investigación en universidades españolas en el marco de la RedAUTI: Universidad de Córdoba de España (UCO) y Universidad de Oviedo (UNIOVI).

### Referencias

- [1] Norman, D. (2010). El diseño de los objetos del futuro. La interacción entre el hombre y la maquina. Barcelona: Editorial PAIDOS.
- [2] Rabardel P., People and Technology, 2002, <http://ergoserv.psy.univ-paris8.fr/>
- [3] Wiske Stone, M.; Hammerness, K.; Gray Wilson, D. (1999). La enseñanza para la comprensión. Buenos Aires: Paidós
- [4] Sitio Andes Aprendizajes Nuevos y Dinámicos para Escuelas y Sociedades <http://learnweb.harvard.edu/andes/home/index.cfm>
- [5] Perrenoud, P. (2006). Construir competencias desde la escuela. Ediciones Noreste, J. C. Sáez Editor. Santiago: Ediciones Noreste
- [6] Mishra, P., & Koehler, M. J. (2006). Technological Pedagogical Content Knowledge: A new framework for teacher knowledge. Teachers College Record. 108(6), 1017-1054.
- [8] Conole, G. (2008) Capturing practice: the role of mediating artefacts in learning design, in *Handbook of Research on Learning Design and Learning Objects: Issues, Applications and*



- Technologies*, in L. Lockyer, S. Bennett, S. Agostinho, and B. Harper (Eds), 187-207, Hersey PA: IGI Global
- [9] De Pablos Pons, J. (2009). Tecnología Educativa. La formación del profesorado en la era de Internet. España: Ediciones Aljibe.
- [10] Dussel, Inés (2010). VI Foro Latinoamericano de Educación; Educación y nuevas tecnologías: los desafíos pedagógicos ante el mundo digital. Inés Dussel y Luis Alberto Quevedo, Buenos Aires: Santillana.
- [11] Carbonell, J. (2001). La aventura de innovar. El cambio en la escuela. España: Ediciones Morata.
- [12] Coll, C., Mauri, T. y Onrubia, J. (2008a). La utilización de las tecnologías de la información y la comunicación en la educación: Del diseño tecno-pedagógico a las prácticas de uso. En C. Coll y C. Monereo (Eds.). Psicología de la educación virtual. España: Ediciones Morata, 74-103.
- [13] Coll, C., Mauri, T. y Onrubia, J. (2009). Hacia una modelización del proceso de enseñanza-aprendizaje mediado por las TIC. Algunas teorías y enfoques centrados en la actividad constructiva del alumnado. En J. de Pablos Pons (coord.). Tecnología educativa. La formación del profesorado en la era de Internet. España: Ediciones Aljibe, 145-161.
- [14] Squire, K.D. (2008). Game-based learning: An emerging paradigm for learning. *Performance Improvement Quarterly*, 21 (2), 7-36.  
<http://www3.interscience.wiley.com/journal/120835177/issue>
- [15] Prensky, M. (2001). Digital game-based learning. New York, NY: McGraw-Hill.
- [16] Pontes Pedrajas, A. (2005). Aplicaciones de las tecnologías de la información y de la comunicación en la educación científica. Primera parte: funciones y recursos. *Revista Eureka sobre Enseñanza y Divulgación de las Ciencias*, 2 (1), 2-18, <http://www.apac-eureka.org/revista/>
- [17] Jonassen, D. y Carr, C. (2000). Mindtools: Affording Multiple Knowledge Representations for Learning. En S. P. Lajoie (Ed.). *Computer as cognitive tools*. Mahwah, NJ: Lawrence Erlbaum Associates.
- [18] Balasubramanian, N., & Wilson, B.G. (2006). Games and simulations. In C. Crawford et al., (Eds.), *ForeSITE*, Volume One, (p. 2). 2005, Proceedings of Society for Information Technology and Teacher Education International Conference 2006. Chesapeake, VA: AACE. [Online] Available: <http://site.aace.org/pubs/foresite/> (November, 2011).
- [19] Aldrich, C. (2009). Virtual worlds, simulations, and games for education: A unifying view. *Innovate* 5 (5), [http://www.innovateonline.info/pdf/vol5\\_issue5/Virtual\\_Worlds\\_Simulations\\_and\\_Games\\_for\\_Education- A\\_Unifying\\_View.pdf](http://www.innovateonline.info/pdf/vol5_issue5/Virtual_Worlds_Simulations_and_Games_for_Education- A_Unifying_View.pdf)
- [20] N. González. Contenidos en la TV digital argentina. Estrategias y actores  
A. Barrero, D. Melendi, X. Pañeda, R. García, S. Cabrero. Sistemas de introducción de texto en aplicaciones de TV interactiva jAUTI 2012 | I Jornadas de Difusión y Capacitación de Aplicaciones y Usabilidad de la Televisión Digital Interactiva, Disponible on-line en <http://sedici.unlp.edu.ar/handle/10915/25926>
- [21] A. Barrero, D. Melendi, X. Pañeda, R. García, S. Cabrero. Sistemas de introducción de texto en aplicaciones de TV interactiva jAUTI 2012 | I Jornadas de Difusión y Capacitación de Aplicaciones y Usabilidad de la Televisión Digital Interactiva Disponible on-line en <http://sedici.unlp.edu.ar/handle/10915/25926>