

LIBRO DE ACTAS



CACIC 2017

XXIII CONGRESO ARGENTINO DE CIENCIAS DE LA COMPUTACIÓN



XXIII Congreso Argentino de Ciencias de la Computación (CACIC 2017)

La Plata (Buenos Aires), 9 al 13 de octubre 2017

Organizadores

Red de Universidades con Carreras en Informática – RedUNCI.

Facultad de Informática - Universidad Nacional de La Plata

XXIII Congreso Argentino de Ciencias de la Computación CACIC 2017: libro de Actas.
Editado por Armando De Giusti; Patricia Pesado. - 1a ed.

La Plata: Universidad Nacional de La Plata. Facultad de Informática, 2017.

Libro digital, PDF

Archivo Digital: descarga y online

ISBN 978-950-34-1539-9



1. Computación. 2. Actas de Congresos. I. De Giusti, Armando, ed. II. Pesado, Patricia, ed.
CDD 004



Autoridades de CACIC 2017

Comité de Honor

Mg. Tavela, Danya – Secretaria de Políticas Universitarias

Lic. Perdomo, Raúl – Presidente de la Universidad Nacional de La Plata

Dr. Ceccatto, Alejandro – Presidente CONICET

Ing. Carmona, Aníbal –Presidente CESSI



Comité Científico

Coordinador Ing. Armando De Giusti – UNLP – CONICET – Argentina

- | | | |
|-------------------------------------|------------------------------------|------------------------------------|
| Abásolo, María José (Argentina) | Falappa, Marcelo (Argentina) | Olkotz Katzalin (España) |
| Aciti, Claudio (Argentina) | Feierherd Guillermo (Argentina) | Pardo, Álvaro (Uruguay) |
| Alfonso, Hugo (Argentina) | Fernandez Aurelio (España) | Pesado, Patricia (Argentina) |
| Ardenghi, Jorge (Argentina) | Fillottrani, Pablo (Argentina) | Piattini, Mario (España) |
| Arroyo, Marcelo (Argentina) | Finocchietto, Jorge (Argentina) | Piccoli, María Fabiana (Argentina) |
| Astudillo Hernán (Chile) | Fрати Emmanuel (Argentina) | Piñuel Luis (España) |
| Baldasarrí, Sandra (España) | García Garino, Carlos (Argentina) | Printista, Marcela (Argentina) |
| Balladini, Javier (Argentina) | García Villalba, Javier (España) | Puppo, Enrico (Italia) |
| Barbosa, Luiz (Portugal) | Género, Marcela (España) | Ramón, Hugo (Argentina) |
| Bertone, Rodolfo (Argentina) | Gomez, Sergio (Argentina) | Rexachs, Dolores (España) |
| Bría, Oscar (Argentina) | Gröller, Eduard (Austria) | Reyes, Nora (Argentina) |
| Brisaboa, Nieves (España) | Guerrero, Roberto (Argentina) | Roig Vila, Rosabel (España) |
| Buckle, Carlos (Argentina) | Ierache, Jorge (Argentina) | Rosete Alejandro (Cuba) |
| Cañas, Alberto (EE.UU) | Janowski, Tomasz (Naciones Unidas) | Rossi, Gustavo (Argentina) |
| Casali, Ana (Argentina) | Jordan Ramiro (USA) | Rosso, Paolo (España) |
| Castro, Silvia (Argentina) | Kuna Horacio (Argentina) | Rueda, Sonia (Argentina) |
| Castrolehtaler, Antonio (Argentina) | Lanzarini, Laura (Argentina) | Ruiz Francisco (España) |
| Cechich, Alejandra (Argentina) | Leguizamón, Guillermo (Argentina) | Russo, Claudia (Argentina) |
| Chavez, Edgar (México) | Lopez Gil, Fernando (España) | Salto, Carolina (Argentina) |
| Coello Coello, Carlos (México) | Loui, Ronald Prescott (EEUU) | Sanz, Cecilia (Argentina) |
| Cuckierman, Uriel (Argentina) | Luque, Emilio (España) | Simari, Guillermo (Argentina) |
| Cuevas, Alfredo Simón (Cuba) | Madoz, Cristina (Argentina) | Simon Alfredo (Cuba) |
| De Giusti, Laura (Argentina) | Malberti, Alejandra (Argentina) | Sposito, Osvaldo (Argentina) |
| De Vincenzi Marcelo (Argentina) | Malbran, Maria (Argentina) | Steinmetz, Ralf (Alemania) |
| Deco, Claudia (Argentina) | Manresa Yee, Cristina (España) | Suppi, Remo (España) |
| Denham, Monica (Argentina) | Marco Javier (España) | Tarouco, Liane (Brasil) |
| Depetris, Beatriz (Argentina) | Marín, Mauricio (Chile) | Tirado, Francisco (España) |
| Díaz, Javier (Argentina) | Mas Sansó, Ramón (España) | Velho, Luiz (Brasil) |
| Dix, Juerguen (Alemania) | Micolini, Orlando (Argentina) | Vendrell, Eduardo (España) |
| Doallo, Ramón (España) | Motz, Regina (Uruguay) | Vénere, Marcelo (Argentina) |
| Docampo, Domingo (España) | Naiouf, Marcelo (Argentina) | Villagarcía, Horacio (Argentina) |
| Dujmovic Jozo (USA) | Navarro Martín, Antonio (España) | Zanarini, Dante (Argentina) |
| Estayno, Marcelo (Argentina) | Oktava Hanna (Mexico) | |
| Estevez, Elsa (Argentina) | Olivas Varela, José Angel (España) | |



Comité Académico

- | | |
|---------------------------------------|---------------------------------|
| UBA Cs. Exactas - Garbervetsky, Diego | UNVM - Prato, Laura |
| UBA Ingeniería - Echeverría, Adriana | UNLu - Panessi, Wálter |
| UNLP - Pesado, Patricia | UNCa - Poliche, María Valeria |
| UNS - Rueda, Sonia | UNLaR - Campazzo, Eduardo |
| UNSL - Piccoli, Fabiana | UNTREF - Oliveros, Alejandro |
| UNCPBA - Aciti, Claudio | UNT - Luccioni, Griselda María |
| UNCOMA - Vaucheret, Claudio | UNAJ - Morales, Martín |
| UNLaM - Sposito, Osvaldo | UNCAUS - Zachman, Patricia |
| UNLPam - Alfonso, Hugo | UNO - Foti, Antonio |
| UNLZ - Estayno, Marcelo | UNCuyo - Forradelas, Raymundo |
| UNTdF - Feierherd, Guillermo | UNMdP - Doumecq, Julio Cesar |
| UNSa - Gil, Gustavo | UNQuilmes |
| UNPA - Lasso, Marta | UNAHUR - Medrano, Gustavo |
| UNSJ - Rodríguez, Nelson | UM - Padovani, Hugo |
| UADER - Vivas, Patricia | UAI - De Vincenzi, Marcelo |
| UNPSJB - Buckle, Carlos | UB - Guerci, Alberto |
| UNER - Tugnarelli, Mónica | UK - Panizzi, Marisa |
| UNNE - Dapozo, Gladys | UAP - Bournissen, Juan |
| UNR - Kantor, Raúl | UCAECE - Finocchietto, Jorge |
| UNaM - Kuna, Horacio | UP - Álvarez, Adriana |
| UNNOBA - Russo, Claudia | UCA Rosario - Grieco, Sebastián |
| UNdeC - Carmona, Fernanda | USAL - Zanitti, Marcelo |
| UNLa - Rodríguez, Darío | UDA - Giménez, Rosa |
| UNSE - Duran, Elena | UGA - Beyersdorf, Carlos |
| UNDEF - Arroyo Arzubi, Alejandro | USEMA - Guglianone, Ariadna |
| UNL - Loyarte, Horacio | UAustral - Cassol, Ignacio |
| UNRC - Arroyo, Marcelo | UAA - Rathmann, Liliana |
| UNC - Fridlender, Daniel | UCA La Plata - Bertone, Rodolfo |
| UNJu - Herrera Cognetta, Analía | ITBA - Valles, Santiago |
| UNRN - Vivas, Luis | UCH - Pinciroli, Fernando |



Comité Organizador

El XXIII Congreso Argentino de Ciencias de la Computación es organizado por la Facultad de Informática de la UNLP en nombre de la Red de Universidades con Carreras de Informática (RedUNCI).

El equipo de trabajo para la organización del Congreso está coordinado por el Instituto de Investigación en Informática LIDI (III-LIDI) de la Facultad de Informática de la UNLP.





Autoridades de la RedUNCI

Coordinador Titular

Pesado, Patricia (UNLP)

Coordinador Alterno

Estayno, Marcelo (UNLZ)

Junta Directiva

Kuna, Horacio (UNM)

Finocchietto, Jorge (UCAECE)

Sposito, Osvaldo (UNLaM)

Russo, Claudia (UNNOBA)

Feierherd, Guillermo (UNTDF)

Carmona, Fernanda (UNdeC)

Piccoli, Fabiana (UNSL)

Padovani, Hugo (UM)

Secretarías

Secretaría Administrativa: Tugnarelli, Mónica (UNER)

Secretaría Académica: Lasso, Marta (UNPA)

Secretaría de Ciencia y Técnica: Balladini, Javier (UNCOMA)

Secretaría de Asuntos Reglamentarios: De Vicenzi, Marcelo (UAI)

Secretaría de Vinculación Tecnológica y Profesional: Gil, Gustavo (UNSa)

Secretaría de Congresos, Publicaciones y Difusión: De Giusti, Armando (UNLP)



Índice

XVIII Workshop de Agentes y Sistemas Inteligentes.	1
A Metaheuristic Compendium for Scheduling Problems.	2
Carlos Bermúdez (UNLPAM), Gabriela Minetti (UNLPAM), Carolina Salto (UNLPAM)	
A study of Convolutional Architectures for Handshape Recognition applied to Sign Language.	13
Facundo Quiroga (UNLP), Ramiro Antonio (UNLP), Franco Ronchetti (UNLP), Laura Lanzarini (UNLP), Alejandro Rosete (ISPJAE)	
Active Learning to Reduce Cold Start in Recommender Systems.	23
Luciano Silvi (UNC)	
An agent-based simulation model using decoupled learning rules to (re)schedule multiple projects.	33
Laura Tosselli (UTN-FRVM), Verónica Bogado (UTN-FRVM), Ernesto Martínez (UTN-FRSF)	
Deep Neural Networks for Shimmer Approximation in Synthesized Audio Signal.	43
Mario Alejandro García (UTN-FRC), Eduardo Destefanis (UTN-FRC)	
ESSIM-EA applied to Wildfire Prediction using Heterogeneous Configuration for Evolutionary Parameters.	53
Miguel Méndez Garabetti (UTN-FRM), Germán Bianchini (UTN-FRM), Paola Caymes-Scutari (UTN-FRM), María Laura Tardivo (UNRC), Verónica Gil Costa (UNSL)	
Fuzzy Harmonic Systems for Traffic Risk.	63
Walter Bel (UADER), Daniela López De Luise (UADER)	
Hacia un Formalismo Basado en Sistemas Multiagente para Evaluar Modelos de Organización Humanos. Caso de Estudio (I): Holacracy.	73
Alexandre Ribas Hortal (UTN-FRM), Facundo Bromberg (UTN-FRM)	
Image recognition for x-ray luggage scanners using free and open source software.	83
Pablo Lázaro (PSA), Ariel Maiorano (PSA)	
Improving Artificial Bee Colony Algorithm with Evolutionary Operators.	93
Gabriela Minetti (UNLPAM), Carolina Salto (UNLPAM)	
Learning When to Classify for Early Text Classification.	103
Juan Martín Loyola (UNSL), Marcelo Luis Errecalde (UNSL), Hugo Jair Escalante (INAOE), Manuel Montes y Gomez (INAOE)	
Optimización de un Método de Reducción de Incertidumbre Aplicado a Incendios Forestales.	113
Laura Tardivo (UNRC), Paola Caymes-Scutari (UTN-FRM), Miguel Méndez Garabetti (UTN-FRM), Germán Bianchini (UTN-FRM)	
Survival of the mutable: architecture of adaptive reactive agents.	123
Renata Luiza Stange (USP), Paulo Roberto Massa Cereda (POLI-USB), João Jose Neto (POLI-USB)	
Un formalismo para persuasión con DeLP.	133
Guillermo Aguirre (UNSL), Marcelo Errecalde (UNSL), Guillermo Simari (UNS)	



XVIII Workshop de Procesamiento Distribuido y Paralelo.	143
Análisis de la precisión de la predicción energética de RAPL en una aplicación de multiplicación de matrices en memoria compartida.	144
Juan Manuel Paniego (UNLP), Silvana Gallo (UNLP), Martin Pi Puig (UNLP), Franco Chichizola (UNLP), Laura De Giusti (UNLP), Armando De Giusti (UNLP), Javier Balladini (UNCOMA)	
Blocked All-Pairs Shortest Paths Algorithm on Intel Xeon Phi KNL Processor: A Case Study.	154
Enzo Rucci (UNLP), Armando De Giusti (UNLP), Marcelo Naiouf (UNLP)	
Cloud Robotics: Auto Rover 4WD y Cuadricóptero controlados remotamente desde AWS.	165
Armando De Giusti (UNLP), Ismael Rodríguez (UNLP), Manuel Costanzo (UNLP), Marcos Boggia (UNLP)	
Despliegue de un Sistema Web Vertical Rígido como Sistema Web Elástico en la Nube.	175
Sebastián Invernizzi (UNA), Sebastián Ortiz (UNA), Carlos Núñez (UNA), Cristian Cabanellas (UNA)	
GPU Performance and Power Consumption Analysis: A DCT based denoising application.	185
Martin Pi Puig (UNLP), Laura De Giusti (UNLP), Marcelo Naiouf (UNLP), Armando De Giusti (UNLP)	
High Performance Computing for Tumor Propagation Agent-based Model.	196
Ghazal Tashakor (UAB), Emilio Luque (UAB), Remo Suppi (UAB)	
Performance Analysis of ABM Distributed Simulation for Real Crowd Evacuation Scenarios.	206
Mohammen J. Alghazzawi (UAB), Ghazal Tashakor (UAB), Francisco Borges (IFBa), Remo Suppi (UAB)	
Procesamiento Distribuido y Paralelo de Bajo Costo Basado en Cloud&Movil.	216
David Petrocelli (UNLu), Armando De Giusti (UNLP), Marcelo Naiouf (UNLP)	
Using an Improved Data Structure in Hybrid Memory for Agent-Based Simulation.	226
Silvana Gallo (UNLP), Francisco Borges (IFBa), Laura De Giusti (UNLP), Marcelo Naiouf (UNLP), Remo Suppi (UAB)	
XVI Workshop de Tecnología Informática Aplicada en Educación.	236
A Concurrent Programming Language for Arduino and Educational Robotics.	237
Ricardo Moran (UAI), Matías Teragni (UAI), Gonzalo Zabala (UAI)	
Analyzing the desing of Student Play, an educational module of Agent Socialmetric.	247
Antonietta Kuz (UTN), Roxana Giandini (UNLP)	
Applying fuzziness to enemy behavior in a serious game.	257
Esteban Aitor Zapirain (UNMdP), Stella Maris Massa (UNMdP)	
Caso de aplicación de representación del conocimiento utilizando grafos conceptuales en un sistema de corrección automatizado de exámenes.	265
María Alejandra Paz Menvielle (UTN-FRC), Mario Alberto Groppo (UTN-FRC), Marcelo Martín Marciszack (UTN-FRC), Analía Guzmán (UTN-FRC), Karina Ligorria (UTN-FRC), Martín Casatti (UTN-FRC)	
Construcción de modelos y simuladores con Squeak- Etoys como recurso de aprendizaje en la escuela media.	275
Ricardo P. Salvador (UNR), Claudia Pons (UNLP), Guillermo L. Rodríguez (UNR)	



Desafiate: Juego Serio para la Autoevaluacion de los Alumnos y su Integracion con un Entorno Virtual de Ensenanza y Aprendizaje.	285
Federico Héctor Archuby (UNLP), Cecilia Sanz (UNLP), Patricia Pesado (UNLP)	
Estimación del nivel de habilidad en sistemas tutores inteligentes utilizando una metodología multiatributo.	295
Sonia G. Sosa-Leon (USON), Julio Waissman (USON), Jose A. Olivas Varela (UCLM), Manuel E. Prieto (UCLM)	
Experiencia de enseñanza-aprendizaje de Cloud Computing y Cloud Robotics en la UNLP.	305
Armando De Giusti (UNLP), Ismael Rodríguez (UNLP), Sebastián Rodríguez Eguren (UNLP)	
Extensión, Docencia e Investigación destinada a los Comedores Barriales: un caso testigo de cómo se combinan los pilares de la Universidad en forma multidisciplinaria.	312
Viviana Harari (UNLP), Mariela Stávale (UNLP), Ivana Harari (UNLP)	
Formulación de un Examen Óptimo.	322
Enrique E. Tarifa (UNJu), Sergio L Martínez (UNJu), Samuel Franco Dominguez (UNJu), Jorgelina F. Argañaraz (UNJu)	
Guitarreando: Estudio sobre a utilização de modelagem e tablatura no ensino de violão mediado por tecnologia.	332
Ana Elisa Bonifacio Barrios (UFES), Damián Keller (UFES), Leandro L. Costalonga (UFES)	
Impact of information and communication technologies on teaching-learning processes in children with special needs autism spectrum disorder.	342
Monica R. Romero (UISEK), Javier Diaz (UNLP), Ivana Harari (UNLP)	
Material educativo hipermedial con Realidad Aumentada para la enseñanza y aprendizaje de triángulos a alumnos hipoacúsicos.	354
Silvia Valeiras (UNLP), Natalí Salazar Mesía (UNLP)	
Método para generar recomendaciones personalizadas para integrar grupos de aprendizaje ubicuo y colaborativo.	363
Margarita Alvarez(UNSE), Silvina Unzaga (UNSE), Elena Durán (UNSE)	
Nuevas Tecnologías Aplicadas a la Enseñanza de la Ingeniería. Análisis Comparativo de Ambientes Virtuales 3D.	373
Francisco Lizarralde (UNMdP), Constanza Huapaya (UNMdP)	
Propuesta de un Framework para la creación de Objetos de Aprendizaje.	383
María Lucía Violini (UNLP), Cecilia Verónica Sanz (UNLP), Patricia Mabel Pesado (UNLP)	
ReadMe. Complemento de Aprendizaje Móvil.	393
José Luis Filippi (UNLPAM), Héctor Daniel Pérez (UNLPAM), Sofia Aguirre (UNLPAM), Rodolfo Bertone (UNLP)	
Recta Numérica: Una Actividad de Códimo.	400
Luciano Graziani (UNRN), Gabriela Anahí Cayú (UNRN), Matías Emanuel Sanhueza (UNRN), Enrique Molinari (UNRN)	
Revisión de herramientas para la creación de modelos 3D orientados a la enseñanza de la matemática con realidad aumentada.	408
Tatiana Gibelli (UNRN), Alfredo Graziani (UNRN), Cecilia Sanz (UNLP)	



Selector de actividades de AulasWeb: la utilización de personajes virtuales para acompañar las decisiones de los docentes.	418
Alejandro Hector Gonzalez (UNLP), Cristina Madoz (UNLP), Leandro Matías Romanaut (UNLP)	
SocialSchool: una plataforma web para instituciones educativas.	428
Franco Brusatti (UNRC), Sonia Permigiani (UNRC), María Marta Novaira (UNRC), Rodrigo Ariño (PlumLabs), Eduardo Depetris (PlumLabs), Saul Coria (PlumLabs)	
Una Propuesta de Evaluación de Herramientas CASE para la Enseñanza.	435
Nicolás Battaglia (UAI), Roxana Martínez (UAI), Carlos Neil (UAI), Marcelo De Vincenzi (UAI)	
Uso de Controladores Difusos en el Proceso de Evaluación en Matemática.	445
Graciela C. Lombardo (UNaM), Marcelo J. Marinelli (UNaM)	
Utilización de Técnicas de Análisis Automático para la Asistencia en Entornos de Introducción a la Programación.	455
Maximiliano Chaves (UNRC), María Marta Novaira (UNRC), Sonia Permigiani (UNRC), Nazareno Aguirre (UNRC)	
Virtualización en la Educación: Laboratorio Portátil de Redes.	465
Mercedes Barrionuevo (UNSL), Cristian Gil (UNSL), Matías Garibaldi (UNSL), Christopher Suarez (UNSL), Carlos Taffermaberry (UNSL)	
Visualización del proceso colaborativo como metaconocimiento. Descripción de una estrategia de mirroring y sus resultados.	475
María Alejandra Zangara (UNLP), Cecilia Sanz (UNLP)	
XV Workshop de Computación Gráfica, Imágenes y Visualización.	487
A tutorial on the implementations of linear image filters in CPU and GPU.	488
Álvaro Pardo (UCU)	
An Open-Source Spanish Video Game as a Case Study for the Development of an Interface for Users with Visual Impairment.	498
Estefanía Güimil (UNS), Yesica Sacristan (UNS), Dana K. Urribarri (UNS), Martín Larrea (UNS)	
Fusión de información de geometría e intensidad para segmentación de imágenes TOF.	508
Luciano Lorenti (UNLP), Javier Giacomantone (UNLP), Oscar Bria (UNLP), Armando De Giusti (UNLP)	
Immersive Platform for Neuroscience Experimental Studies.	518
Florencia Rodriguez (UNCPBA), Marcos Lazo (UNCPBA), Virginia Cifuentes (UNCPBA), Juan Pablo D'Amato (UNCPBA), Manuel Serodio (UBA), Fabricio Ballarini (UBA), Pedro Bekinschtein (FAVALORO), Cristian Garcia Bauza (UNCPBA)	
Separación de sombras a los objetos detectados con sustracción de fondo en video.	528
Rosana Barbuza (UNCPBA), Leonardo Fernández Esteberena (UNCPBA), Leonardo Domínguez (UNCPBA), Alejandro Perez (UNCPBA), Aldo Rubiales (UNCPBA), Juan P. D'Amato (UNCPBA)	
Virtual Trainer for Physical Activities using a Natural User Interface.	538
Nicolás Jofré (UNSL), Graciela Rodríguez (UNSL), Yoselie Alvarado (UNSL), Jaqueline Fernández (UNSL), Roberto Guerrero (UNSL)	



XIV Workshop Bases de Datos y Minería de Datos	548
Automatización de la extracción de características en tareas de análisis de sentimiento. Juan M. Rodríguez (UNLA), Hernan D. Merlino (UNLA), Patricia Pesado (UNLP), Ramón García Martínez (UNLA)	549
Cálculo del exponente de Hurst utilizando Spark Streaming: enfoque experimental sobre un flujo de transacciones de criptomonedas. María José Basgall (UNLP), Waldo Hasperué (UNLP), Marcelo Naiouf (UNLP), Aurelio F. Bariviera (Universidad de Rovira i Virgili)	558
El Problema de la Dispersión Máxima en un entorno Multi-Objetivo. Mauricio Machuca Cabral (UNA), Benjamín Barán (UNA), Fernando Sandoya (EPDE)	568
Extensión de una base de datos Objeto-Relacional para el soporte de datos de iris. Carlos Alvez (UNER), Graciela Etchart (UNER), Silvia Ruiz (UNER), Ernesto Miranda (UNER), Juan Aguirre (UNER)	578
Framework para el procesamiento lingüístico de artículos científicos. Caso de estudio: Universidad Nacional de Chilecito. José Texier (UNdeC), Jusmeidy Zambrano (UNdeC), Emmanuel Frati (UNdeC)	588
MetaCLAS: A Prototype Evolutionary Proposal to Automatically Suggest Clustering Methods and their Parameters. Macarena A. Latini (UNS), Rocio L. Cecchini (UNS), Jessica A. Carballido (UNS)	599
Prototipo de sistema de captura y monitoreo de datos OBD-II de vehículos. Claudio Aciti (UNCPBA), Mauricio Urraco (UNCPBA), Elias Todorovich (UNCPBA)	609
Sistema recomendador basado en tópicos latentes. María Emilia Charnelli (UNLP), Laura Lanzarini (UNLP), Javier Diaz (UNLP)	619
Uso de Bases de Datos Espacio-Temporales para la Atención de Eventos de Emergencia. Marcela Aballay (UTN-FRC), Anabella De Battista (UTN-FRCU), Edilma Olinda Gagliardi (UNSL)	629
XIV Workshop de Ingeniería de Software.	639
Análisis de estrategias para clasificar contenidos en foros de discusión. Valeria Zoratto (UNCOMA), Facundo Nehuén Otermin Sánchez (UNCOMA), Nadina Martínez Carod (UNCOMA), Gabriela Aranda (UNCOMA)	640
Análisis de métricas de calidad en foros de discusión técnicos. Nadina Martínez Carod (UNCOMA), Pamela Faraci (UNCOMA), Gabriela Aranda (UNCOMA)	650
Asistente para la evaluación de características de calidad de producto de software propuestas por ISO/IEC 25010 basado en métricas definidas usando el enfoque GQM. Julieta Calabrese (UNLP), Rocío Muñoz (UNLP), Ariel Pasini (UNLP), Silvia Esponda (UNLP), Marcos Boracchia (UNLP), Patricia Pesado (UNLP)	660
Dificultades de los "recién llegados" a proyectos software en ejecución. Gerardo Matturro (ORT), Karina Barrella (ORT), Patricia Benitez (ORT), María Inés Lund (UNSJ)	672



Diseño de Aplicaciones Móviles basadas en Posicionamiento: un Framework Conceptual. Cecilia Challiol (UNLP), Alejandra B. Lliteras (UNLP), Silvia E. Gordillo (UNLP)	682
Early aspects in “Aspect-Oriented Process for a Smooth Transition”. Fernando Pincirolo (UCH), José L. Barros-Justo (UVIGO)	692
Elicitación de requerimientos de stakeholders distribuidos. Fernanda Cuadrado (OTRA), Alejandro Oliveros (UNTREF)	702
Evaluación de performance de engine 3D para dispositivos móviles. Federico Cristina (UNLP), Sebastián Dapoto (UNLP), Pablo Thomas (UNLP), Patricia Pesado (UNLP)	712
Evaluating Quality in Agile Developments. A first validation experience with NEA Software SMEs. Noelia Pinto (UTN-FRRe), Cesar Acuña (UTN-FRRe), Nicolás Tortosa (UTN-FRRe), Blas Cabas Geat (UTN-FRRe)	722
Exploring Architectural Model Checking with Declarative Specifications. Fernando Asteasuain (UNdAv), Francisco Tarulla (UBA)	732
Hacia una Meta-Herramienta de Análisis Automático de Modelos de Variabilidad. Esteban Ruiz de Galarreta (UNCOMA), Matías Pol'la (UNCOMA), Agustina Buccella (UNCOMA), Alejandra Cechich (UNCOMA)	742
Ingeniería de Requisitos centrada en la Estrategia CRM. Gladys Kaplan (UNLaM), Mirian Taboada (UNLaM), Gabriel Pousada (UNLaM), Claudia Litvak (UNLaM), Andrea Vera (UNLaM), Gabriel Blanco (UNLaM)	752
Modelo de Proceso de Gestión para Proyectos de Ingeniería del Conocimiento. Hernán Amatriain (UNLA), Hernán Merlino (UNLA), Sebastian Martins (UNLA), Santiago Bianco (UNLA)	762
Process Variability: concepts, approaches and its application on a model of Cloud BPM. José Martínez Garro (UNLP), Patricia Bazán (UNLP), Javier Diaz (UNLP)	772
Propuesta de Protocolo de Formación de Pares Experimentales de Programadores. Mauricio Dávila (UTN-FRBA), Marisa Panizzi (UTN-FRBA), Dario Rodriguez (UNLA)	782
Strategy for Comparing and Adopting Strengths. Guido Tebes (UNLPAM), Denis Peppino (UNLPAM), Pablo Becker (UNLPAM), M. F. Papa (UNLPAM), M. B. Rivera (UNLPAM), Luis Olsina (UNLPAM)	792
Una Plataforma para el Desarrollo de LPS en el Dominio Geográfico: Caso de Estudio en el Subdominio de Paleontología. Fiorella Pesce (UNCOMA), Sofía Caballero (UNCOMA), Agustina Buccella (UNCOMA), Alejandra Cechich (UNCOMA)	802
Una Revisión Sistemática de la Literatura en Pruebas de Compatibilidad Web. Leandro N. Sabaren (UNNE), Maximiliano A. Mascheroni (UNNE), Cristina L. Greiner (UNNE), Emanuel Irrazábal (UNNE)	812
Utilización de SysML y Papyrus para el desarrollo del software de vuelo del Cubesat FS2017 haciendo uso del enfoque MBSE. Emmanuel Arias (UNLaM), Eduardo Sufan (UNLaM), Javier Uranga (UNLaM), Carlos Barrientos (UNLaM)	822
Wide Evaluation Processes. Leonardo Anibal Fernandez (UNLP), Rodolfo Bertone (UNLP)	832



XII Workshop de Arquitecturas, Redes y Sistemas Operativos.	843
A novel performance metric for Virtual Network Embedding combining aspects of Blocking Probability and Embedding Cost.	844
Enrique Dávalos (UNA), Benjamín Barán (UNA)	
An IPC Software Layer for Building a Distributed Virtualization System.	854
Pablo Pessolani (UTN-FRSF), Toni Cortes (UPC), Fernando G. Tinetti (UNLP), Silvio Gonnet (UTN-FRSF)	
Caracterización de Tráfico-Distribución del Johnson SB.	864
Luis Marrone (UNLP)	
Consideraciones sobre el comportamiento del protocolo TCP en sus variantes Vegas, Reno, Cubic y Westwood ante errores en ráfaga en una topología híbrida.	874
Diego R. Rodríguez Herlein (UNPA), Carlos A. Talay (UNPA), Claudia N. Gonzalez (UNPA), Franco A. Trinidad (UNPA)	
Experimental study of performance and QoS of multicast and unicast video traffic in WAN links.	884
Higinio Facchini (UTN-FRM), Santiago Pérez (UNdeC), Luis Marrone (UNLP), Fabian Hidalgo (UTN-FRM), Adrián Cárdenas (UTN-FRM)	
Modelado y simulación de arquitecturas de Cloud Computing con CloudSim: comunicación entre entidades.	894
Daniel Rosatto (UNAJ), Román Bond (UNAJ), Maximiliano Belizán (UNAJ), Martin Morales (UNAJ), Diego Encinas (UNLP)	
Procesador de Petri Modular para Sistemas Embebidos.	904
Orlando Micolini (UNC), Emiliano N. Daniele (UNC), Luis O. Ventre (UNC), Marcelo Cebollada y Verdaguer (UNC), Maximiliano Eschoyez (UNC)	
Requerimientos de VoIP en Redes Malladas Inalámbricas con Relaying.	914
Aristides Silvestris (UNRC), Federico Aguirre (UNRC), Noelia Veglia (UNRC), Ezequiel Tardivo (UNRC)	
Un Modelo de Detección de Anomalías en una LAN usando K-NN y Técnicas de Computación de Alto Desempeño.	924
Mercedes Barrionuevo (UNSL), Mariela Lopresti (UNSL), Natalia Miranda (UNSL), Fabiana Piccoli (UNSL)	
IX Workshop Innovación en Sistemas de Software.	934
"Informática UNLP" la App de la Facultad de Informática.	935
Juan Fernández Sosa (UNLP), Alfonso Cuitiño (UNLP), Pablo Thomas (UNLP), Lisandro Delía (UNLP), Germán Cáseres (UNLP), Leonardo Corbalán (UNLP), Patricia Pesado (UNLP)	
Desarrollo de un Planificador Embebido para la Gestion de Riego Automatico aplicado a invernáculos. Implementacion en INTA El Sombrero.	944
José Manuel Alonso (UNNE), Hugo Ferrari (UNNE), Iván Sambrana (UNNE), Emanuel Irrazábal (UNNE)	
Desarrollo de un Prototipo detector de caídas utilizando la placa Intel Galileo Generación I y el sensor MPU6050.	954
Esteban Carnuccio (UNLaM), Waldo Valiente (UNLaM), Mariano Volker (UNLaM), Graciela De Luca (UNLaM), Gerardo García (UNLaM), Daniel Giulianelli (UNLaM), Sebastián Barillaro (UNLaM)	



Ejecución de comandos de voz mediante Web Speech API. Javier Perez (UNLP), Javier Diaz (UNLP), Ivana Harari (UNLP)	964
Estación de monitoreo en tiempo-real de parámetros agrometeorológicos para determinar la necesidad de riego en plantaciones agrícolas. Alberto Eduardo Riba (UNdeC), Emmanuel Alejandro Portugal Murcia (UNdeC), Fernanda Beatriz Carmona (UNdeC), Jorge Damián Tejada (UNdeC), Matías Agustín Perez (UNdeC)	971
Generation and Use of a Digest System by Integrating OCR and Smart Searches. Germán Cáseres (UNLP), Lisandro Delía (UNLP), Pablo Thomas (UNLP), Verónica Aguirre (UNLP)	980
Implementación con Raspberry PI de un Servidor Portátil de Contenidos. Rocío Rodríguez (UNLaM), Pablo Vera (UNLaM), Daniel Giulianelli (UNLaM), Pablo Cammarano (UNLaM)	990
Integración de Sistemas ERP para el Monitoreo de Planes y Detección de Eventos Disruptivos en Cadenas de Suministros. Érica S. Fernández (UTN), Pedro L. Querini (UTN), Carlos M. Toledo (UTN)	1000
Modelo de Análisis para el Dominio de Tránsito Ciclista. Leo Ordinez (UNPSJB), Damián Barry (UNPSJB), Alex Torrico (UNPSJB), Emanuel Mallon (UNPSJB), José Luis Devia (UNPSJB), Lucas Abella (UNPSJB), Nahuel Alamo (UNPSJB)	1010
Monitoring System for Elderly Care with Smartwatch and Smartphone. Pablo Rosales (UNPSJB), Alejandro Vega (UNPSJB), Carlos De Marziani (UNPSJB), José Gallardo (UNPSJB), Jorge Pires (UNPSJB), Rómulo Alcoleas (UNPSJB)	1020
Multimodal biometric recording architecture for the exploitation of applications in the context of affective computing. Enrique P. Calot (UBA), Jorge S. Ierache (UBA)	1030
Reconocimiento de gestos aplicado al control de dispositivos. Una experiencia en control de TV. Carla Luna Gennari (UNLP), César Estrebow (UNLP), Laura Lanzarini (UNLP)	1040
VIII Workshop Procesamiento de Señales y Sistemas de Tiempo Real.	1050
Análisis de Experimentos de Comunicaciones de Tiempo Real en el Bus CAN. Fernando G. Tinetti (UNLP), Fernando L. Romero (UNLP), Alejandro D. Pérez (UNLP)	1051
Análisis de Performance y Consumo sobre Sistemas Embebidos Multinucleo. Matías Dell'Oso (UNLP), Santiago Medina (UNLP), Fernando Romero (UNLP), Armando De Giusti (UNLP)	1061
Control PID Robusto de Sistemas de Fase no Mínima Combinando QFT y GSA: Aplicación al Control Carga-Frecuencia. Patricia N. Baldini (UTN-FRBB)	1071
Excluding Ionosphericly Unsafe Satellite Geometries in GBAS CAT-I. Oscar Bria (UNLP), Javier Giacomantone (UNLP), Luciano Lorenti (UNLP)	1081
IoT for smart home energy planning. Emilio Orsi (UdelaR), Sergio Nesmachnow (UdelaR)	1091
Sistema de Control de una cámara de germinación hidropónica con IoT. Marcelo Marinelli (UNaM), Ruben R. Urquijo (UNaM)	1101



VI Workshop de Innovación en Educación en Informática.	1108
"Programar Jugando" con Scratch en escuelas rurales de la provincia de Catamarca.	
Cintia Stella Maris Sosa Chasampi (UNCa), Carlos Elias Jais (UNCa), Marcelo González (UNCa), Eduardo Adrián Toloza (UNCa)	1109
¿Invirtiendo la clase o Invirtiendo en la clase? Enseñanza de tecnología para la producción multimedia mediante Aula invertida.	
Mirta Fernández (UNNE), María V. Godoy Guglielmone (UNNE), Sonia I. Mariño (UNNE), Walter Barrios (UNNE)	1119
ECMRE: Extended Concurrent Multi Robot Environment.	
J. Castro (UNLP), L. De Giusti (UNLP), G. Gorga (UNLP), M. Sánchez (UNLP), M. Naiouf (UNLP), A. De Giusti (UNLP)	1133
Educational Experiences in Informatics: Innovations in Internet Searches in Academic Environments.	
Pablo Argañaras (UNRN), Paola Britos (UNRN), Martín Viligrón (UNRN)	1143
El Pensamiento Computacional: experiencia de su aplicación en el aprendizaje de la resolución de problemas.	
María V. Rosas (UNSL), Mariela E. Zúñiga (UNSL), Jacqueline M. Fernandez (UNSL), Roberto A. Guerrero (UNSL)	1152
Explotación de Información para el análisis de los procesos de enseñanza. Caso de estudio: el rol del Ingeniero y las percepciones de los alumnos.	
Luciano Straccia (UTN-FRBA), Ariel Deroche (UTN-FRBA), Pablo Pytel (UTN-FRBA), María Florencia Pollo Cattaneo (UTN-FRBA)	1162
Incidencia de los estilos de aprendizaje de los estudiantes de informática en la valoración de un Objeto de Aprendizaje sobre Puertas Lógicas.	
Natalí Salazar Mesía (UNLP), Cecilia Sanz (UNLP), César Estrebou (UNLP)	1172
LEGV8, Raspberry Pi 3 y una vieja fórmula.	
Pablo Ferreyra (UNC), Agustín Laprovitta (UNC), Delfina Velez Ibarra (UNC), Gonzalo Vodanovic (UNC), Nicolás Wolovick (UNC)	1182
Milleannizando el CMMI: El "Mito" de los Estilos de Aprendizaje Hecho Realidad.	
Marcelo Campo (UNCPBA), Analía Amandi (UNCPBA)	1193
O Aluno monitor potencializando o uso das salas de Informatica nas escolas da rede municipal de ensino de Santa Maria.	
Eunice Pereira Azenha (UFMS), Marite Medianeira Moro Neocatto (UFMS), Karla Marques da Rocha (UFMS)	1203
Proactive Independent Learning Approach: A case study in computer arithmetic.	
Javier Giacomantone (UNLP), Oscar Bria (UNLP)	1213
Student-Centered Learning in a University Programming Course.	
Carlos Fontela (UBA), Pablo Suárez (UBA)	1222



VI Workshop de Seguridad Informática.	1229
Análisis de metodologías de recolección de datos digitales en servidores web.	
Mónica D. Tugnarelli (UNER), Mauro F. Fornaroli (UNER), Sonia R. Santana (UNER), Eduardo Jacobo (UNER), Javier Díaz (UNLP)	1230
Aspectos de Seguridad en Internet de las Cosas.	1239
Alicia Castro (UNSL), Eduardo Cassanovas (IUA), Verónica Gil Costa (UNSL)	
Buenas prácticas para la recolección de la evidencia digital en la Argentina.	1249
Nicolás Armilla (UM), Marisa Panizzi (UM), Jorge Eterovic (UM), Luis Torres (UM)	
Efficient Votes Storage in a Non-Interactive Dining Cryptographers (NIDC) Protocol.	1259
Pablo García (UNLPAM), Silvia Bast (UNLPAM), Germán Montejano (UNSL)	
Improving the Generation of Labeled Network Traffic Datasets Through Machine Learning Techniques.	1269
Jorge Guerra (UNCUYO), Carlos Catania (UNCUYO)	
Observer effect: How Intercepting HTTPS traffic forces malware to change their behavior.	1279
María José Erquiaga (UNCUYO), Sebastián García (CVUT), Carlos García Garino (UNCUYO)	
Posibles problemas en Qlink.it y librería CryptoJS.	1289
Antonio Castro Lechtaler (IESE), Marcelo Cipriano (IUE), Edith García (IUE), Pablo Lázaro (PSA), Julio Liporace (IUE), Eduardo Malvacio (IUE), Ariel Maiorano (PSA)	
Post-Quantum Cryptography Using Hyper-Complex Numbers.	1299
Jorge Alejandro Kamlofsky (UAI), Juan Pedro Hecht (UBA)	
Two-Way Continual Authentication Model.	1309
Carlos Fabian Lugani (UNRN)	
Verificabilidad "End to End" para OTP - Vote.	1317
Pablo García (UNLPAM), Silvia Bast (UNLPAM), Germán Montejano (UNSL)	

XVIII Workshop Agentes y Sistemas Inteligentes (WASI)

Coordinadores

Gómez, Sergio A. - Leguizamón, Guillermo - Salto, Carolina

Comité Científico

Aguirre, Guillermo (Argentina)	Lanzarini, Laura (Argentina)
Alfonso, Hugo (Argentina)	Maguitman, Ana (Argentina)
Alonso Alemany, Laura (Argentina)	Martínez, Diego (Argentina)
Apolloni, Javier M. (Argentina)	Martínez, María Vanina (Argentina)
Aragón, Victoria (Argentina)	Mezura, Efrén (México)
Bavera, Francisco (Argentina)	Minetti, Gabriela (Argentina)
Carballido, Jessica (Argentina)	Olivera, Ana Carolina (Argentina)
Casali, Ana (Argentina)	Salto, Carolina (Argentina)
Cassano, Valentín (Canadá)	Schiaffino, Silvia (Argentina)
Cecchi, Laura (Argentina)	Simari, Gerardo Ignacio (Argentina)
Cobo, María Laura (Argentina)	Simari, Guillermo R. (Argentina)
Edera, Alejandro (Argentina)	Tamargo, Luciano (Argentina)
Falappa, Marcelo A. (Argentina)	Toscano Pulido, Gregorio (México)
García, Alejandro (Argentina)	Villagra, Andrea (Argentina)
Godoy, Daniela (Argentina)	Zunino, Alejandro (Argentina)
Gottifredi, Sebastián (Argentina)	

A Metaheuristic Compendium for Scheduling Problems

Carlos Bermudez¹, Gabriela Minetti¹, and Carolina Salto^{1,2}

¹ Facultad de Ingeniería, Universidad Nacional de La Pampa

² CONICET, Argentina

{minettig,saltoc}@ing.unlpam.edu.ar

Abstract. The flexible job shop scheduling problem (FJSSP) is a difficult and complex problem, proved to be NP-hard, in manufacturing environments, because it has to assign each operation to the appropriate machine besides sequencing operations on machines. Due to that complexity, metaheuristics became the best choice to solve in practice this kind of problem. Therefore, the aim of this paper is to offer a reliable compendium in order to cover a wide algorithmic spectrum of different techniques. Further, a study of their accuracy and computational effort is carried out in order to achieve a behavior comparison. This paper shows different algorithmic trends that can be observed through this analysis.

1 Introduction

Scheduling operations is one of the most critical issues in manufacturing processes, being the Job Shop Scheduling Problem (JSSP) one of most important and difficult problems [1]. Each job has to undergo multiple operations on the various machines and each job has its own set of processing times and routing characteristics. The decision concerns how to sequence the operations on the machines, so that the time needed to complete all the jobs (makespan) is minimized. The possibility of selecting alternative routes among the machines is useful in production environments where multiple machines are able to perform the same operation (possibly with different processing times), as it allows the system to absorb changes in the demand of work or in the performance of the machines. When this factor is considered, the problem is known as Flexible Job Shop Scheduling Problem (FJSSP).

Generally, the FJSSP is a more realistic production environment and then has more practical applicability. However, the FJSSP is more complex (NP-hard problem [2]) than the JSSP because of its additional decision to assign each operation to the appropriate machine (routing) besides sequencing operations on machines. Due to the mentioned complexity of FJSSP, the adoption of heuristic methods is suggested because they produce reasonably good schedules in a reasonable time, instead of looking for an optimum solution, also for small instances. In recent years, the adoption of metaheuristics [3] has led to better results than classical dispatching or greedy heuristic algorithms [4–6].

The main purpose of this paper is to offer a base reference for FJSSP comparison, using different metaheuristics in order to cover a wide algorithmic spectrum,

from trajectory methods to population-based methods. In this way, we obtain a valuable compendium of approaches for FJSSP by specifying their qualities from two points of view: first, adapting the metaheuristics to solve the problem, and, second, finding good quality schedules with a low computational effort. Consequently, the methodology followed in this empirical study consists in adapting five metaheuristics to solve the FJSSP. We evaluate and compare each of them on the well-known set of the problem instances. The planned analysis and comparisons will help us to explain the result qualities and efficiency of these metaheuristics to solve the FJSSP and also, to compare the best studied metaheuristics with different proposals found in the literature.

The paper is organized as follows. In Section 2, we introduce the problem formulation. In Section 3, we explain the adaptations of the selected metaheuristics to solve the FJSSP. In the following section, we introduce the experimental design and in Section 5, we evaluate the results. Further, in Section 6 we make a comparison between these metaheuristics and the ones present in the literature. Some final remarks and future research directions are given in Section 7.

2 The Flexible Job Shop Scheduling Problem

The flexible job shop can be described as follows. Let a set $J = \{J_1, J_2, \dots, J_n\}$ of independent jobs. A job J_i is formed by a sequence of $O_{i1}, O_{i2}, \dots, O_{in_i}$ operations to be performed one after the other according to the given sequence. Given a set $U = \{M_1, M_2, \dots, M_m\}$ of machines, each operation O_{ij} can be processed on a subset $U_{ij} \subseteq U$ of compatible machines. We have partial flexibility if there exists a proper subset $U_{ij} \subset U$, for at least one operation O_{ij} , while we have $U_{ij} = U$ for each operation O_{ij} in the case of total flexibility. The processing time of each operation is machine-dependent. We denote with d_{ijk} the processing time of operation O_{ij} when executed on machine M_k . Pre-emption is not allowed and the machines cannot perform more than one operation at a time. All jobs and machines are available at time 0. The problem is to assign each operation to an appropriate machine (routing problem), and to sequence the operations on the machines (sequencing problem) in order to minimize the makespan. This measure is the time needed to complete all the jobs, which is defined as $C_{max} = \max_i\{C_i\}$, where C_i is the completion time of job J_i .

3 Metaheuristics for FJSSP

In this section, we describe how the five metaheuristics used in this work are adapted to solve the FJSSP. Two of them are trajectory-based metaheuristics (Simulated Annealing and Iterated Local Search) and the remaining ones are population-based metaheuristics (Genetic Algorithm, Cuckoo Search, and Imperialist Competitive Algorithm). All of them share two main common design points: the representation of solutions handled by algorithms and the definition of the objective function that will guide the search.

We use the encoding proposed by Bierwirth in [7], which is based on permutation with repetitions. A solution, S , is a permutation of the set of operations

that represents a tentative ordering to schedule them, each one being represented by its job number. For example, $S = [2, 1, 1, 3, 2, 3, 1, 2]$ is a valid solution, which corresponding operation sequence is $O_{21}, O_{11}, O_{12}, O_{31}, O_{22}, O_{32}, O_{13}$, and O_{23} . In order to evaluate S , the fitness value is the makespan (C_{max}) of this solution. To compute this value, each operation O_{ij} in S is assigned to a feasible machine M_k in the subset U_{ij} with the shortest completion time, and then the load of M_k must be updated. As the initial solution we use a random procedure, mainly because high performing construction heuristics for the FJSSP are unknown.

Finally, in order to make a fair comparison among these algorithms, they must do the same computational effort in each run. This can be reached if they are executed during the same time for each problem instance. As suggested in the literature for JSSP, the total executed time is calculated as $ExecTime = \#O \times (\frac{\#O}{2}) \times 30$, where $\#O$ is the total number of operations for a given instance.

3.1 Simulated Annealing

Simulated Annealing (SA) is a simple and general purpose Monte-Carlo method which was developed for combinatorial optimization [8]. This version of SA generates an initial solution S_0 in a random way, and a neighbor, S_1 , from S_0 using the exchange operator. This operator randomly selects two positions and their respective operations are swapped if they belong to different jobs, since the encoding is a permutation with repetitions. If S_1 is worse than S_0 , S_1 can be accepted under the Boltzmann probability. In this way, at high temperatures (T) the exploration of the search space is allowed. In contrast, at low temperatures the algorithm only exploits a promising region of the search space. In order to update T , the proportional cooling process [8] is used and it is applied after a certain number of iterations given by the Markov Chain Length (MCL). Finally, SA ends the search when the total executed time ($ExecTime$) is reached.

3.2 Iterated Local Search

The Iterated Local Search (ILS) [9] is a simple but very effective metaheuristic. To solve the FJSSP with ILS, we use the ideas proposed in [10]. The algorithm 1 begins the search from an initial solution x (line 2). After that, the main loop of ILS starts, which consists in the application of local search and perturbation procedures (lines 4-21). A local search procedure (LS) is applied to the current solution x (line 5). This search stops as soon as a better solution x' is found. When no improvements are found, a counter (*count*) is increased (line 12). This counter is set to zero each time the local search improves the current solution (line 14). If the counter exceeds a certain threshold (*threshold*), a perturbation mechanism is applied, with the aim of redirecting the search to more promising regions of the solution space (lines 13-19). This mechanism is applied to the current solution x a certain number of times (defined by *nu_move*) generating a set $\tilde{s}(r)$ of candidate perturbation solutions. The best of them is selected as the new current solution x .

The local search uses iteratively the exchange operator and consists in the following steps. The operation O_{ij} in the first position of the current solution is exchanged by an operation located in a different randomly selected position in the permutation. If this new permutation x' has a better makespan ($f(x') < f(x)$), then it will replace the solution x and the local search procedure ends.

Algorithm 1 ILS to solve the FJSSP

```

1:  $count = 0$ ;
2: Initialize  $x$ ; {ILS generates a random solution}
3:  $x_{best} = x$ ;
4: while stop criterion not met do
5:    $x' = LS(x)$ ; {Interchange operator}
6:   if  $f(x') < f(x)$  then
7:      $x = x'$ ;
8:     if  $f(x') < f(x_{best})$  then
9:        $x_{best} = x'$ ;
10:    end if
11:  else
12:     $count = count + 1$ ;
13:    if  $count > threshold$  then
14:       $count = 0$ ;
15:      for  $r = 1$  to  $nu\_move$  do
16:         $\tilde{s}(r) = perturbation(x)$ ;
17:      end for
18:       $x = \tilde{s}_{best}$ ;
19:    end if
20:  end if
21: end while
22: return  $x_{best}$ ;

```

Algorithm 2 CS to solve the FJSSP

```

1:  $t = 0$ ; {current generation}
2: initialize( $H(t)_{1,N}$ );
3: evaluate( $H(t)$ );
4: sort( $H(t)$ ); {Rank the solutions}
5: while (stop criterion not met) do
6:    $pos = random\_pos(1, N \times p_{best})$ ;
7:    $h = newCuckoo(H(t)_{pos})$ ;
8:    $pos1 = random\_pos(1, N)$ ;
9:   if  $f(h) < f(H(t)_{pos1})$  then
10:      $H(t)_{pos1} = h$ ;
11:   end if
12:   initialize( $H(t)_{N \times p_a, N}$ );
13:   LS( $H(t)$ );
14:   sort( $H(t)$ ); {Rank the solutions}
15:    $t = t + 1$ ;
16: end while
17: return best solution

```

On the contrary, the procedure continues with the operation in the second position of the permutation, and so on. The search iterates in the majority of the operations, without repetition. The perturbation mechanism consists in the insertion operator. Basically, one job is randomly selected and inserted in a different position of the permutation, which is also randomly selected.

3.3 Cuckoo Search Algorithm

Cuckoo search (CS) algorithm is a novel metaheuristic [11], which is inspired on the obligate brood parasitic behavior of cuckoo birds in combination with the Lévy flight behavior. As the first step of the CS to solve the FJSSP (see Algorithm 2), a population $H(0)$ of N eggs (solutions) is randomly generated (line 2). This initial population is evaluated and then sorted regarding the quality of each solution (lines 3 and 4, respectively). After these steps, the main loop follows. A cuckoo selects randomly a position pos from the p_{best} best solutions in the current population $H(t)$ (line 6). A new candidate solution h is generated by perturbing the current $H(t)_{pos}$ solution following Lévy flights (line 7). The new solution h can replace a randomly selected solution from $H(t)$, following an elitist selection strategy (lines 8-11). After that, a fraction p_a of worse nests is abandoned and new nests are built at new locations (solutions created following a random process) (line 12). Afterwards, a local search step is applied to each individual in $H(t)$ with a p_{LS} probability in order to improve the solution (line 13). At this step the elitist selection strategy is also used. Finally, the loop ends with a sorting process which arranges the individuals in decreasing fitness value (line 14), being the best solution the one present in position 0 of $H(t)$.

To move from a current solution to an other one, the concept of Lévy flight is used. A Lévy flight can be described as a random walk in which the step length (distance between two solutions) is decided by certain probability distribution functions which are heavily tailed. Following the ideas of Ouaarab et al.[12], three different moves or operators, controlled directly by the value generated by Lévy

distribution, are used: exchange, insertion, and inversion. These operators are traditional for the problem at hand and generate different move steps. Exchange move is used for small perturbations and large ones are made by inversion move. Insertion move is considered to introduce median perturbations.

3.4 Genetic Algorithm

Genetic Algorithms (GA) [13] simulate the evolution of individual structures via the Darwinian natural selection process. The GA proposed to solve FJSSP begins with the creation of an initial population of μ solutions in a random way, which are then evaluated. After that, the population goes into a cycle (evolution), which consists of the application of genetic operators, to create λ offspring, obtaining a new population. Each parent for mating is selected using binary tournament selection and they are recombined using a specially designed crossover known as the Job-based Order Crossover (JOX) [14], under a certain probability (p_c). Each individual in the new population is mutated, with a certain probability (p_m), by using the exchange operator. Finally, each iteration ends by selecting μ individuals to build up the new population from the set of $(\mu + \lambda)$ existing ones by using proportional selection (a typical selection method in this step).

3.5 Imperialist Competitive Algorithm

The Imperialist Competitive Algorithm (ICA) [15], is inspired by the imperialistic competition. The Algorithm 3 shows the ICA to solve FJSSP which is based on the ICA version presented in [16]. An initial set of countries (solutions) of size N_C is randomly created (line 1). Then the imperialistic countries are determined (lines 2-4). For that and after calculating the fitness function $c_i = 1/C_{max}$ for each country i , the best N_{imp} of them are selected as imperialists. The rest $N_{col} = N_C - N_{imp}$ countries are the colonies. These colonies are divided among imperialists (lines 5-7), based on their power, in order to form the initial empires. The power p_j of each imperialistic country is first calculated according to $p_j = c_j / \sum_{i=1}^{N_{imp}} c_i$. The number of colonies of each imperialistic country (N_{col_j}) is proportional to its power, and it is determined by $N_{col_j} = p_j \times N_{col}$. A total of N_{col_j} colonies are randomly assigned to an imperialist j .

After this initialization process, the colonies in each of the empires start moving toward their relevant imperialist country. In this work, those movements are accomplished by variation operators specially designed by the adopted representation. One of them is the JOX operator [14], which is applied in such a way that a colony and its imperialist country are considered as parents. The other one is the exchange operator; it is applied to each colony in order to simulate a randomly deviated direction, as the original proposal of ICA. At this point and following the improvement done in [16], ICA applies a local search to the imperialist countries with a certain probability (p_{LS}), using the same procedure described in Section 3.2. When moving toward the imperialist country, a colony might reach a position with higher fitness than its imperialist country (lines 13-15). In this case, the imperialist and the colony should change their positions. If there are several colonies better than the imperialist country, then the imperialist will be replaced with the best colony.

Algorithm 3 ICA to solve the FJSSP

```

1: initialize  $N_C$  countries;
2: evaluate  $c_i$  for each country;
3: select the best  $N_{imp}$  countries as imperialists;
4: choose the remaining  $N_{col}$  countries as colonies;
5: compute  $p_j$  for each empire;
6: determine  $N_{col_j}$  for each empire  $j$ ;
7: select the colonies for assigning to each imperialist in a random way;
8: while (stop criterion is not meet) do
9:   apply JOX between each colony and its imperialist;
10:  apply the exchange operator to each colony;
11:  evaluate  $c_i$  for all colonies;
12:  apply a local search to  $j^{th}$  imperialist;
13:  if  $c_i$  is better than  $c_j$  then
14:    exchange the positions between the imperialist and colony;
15:  end if
16:  compute  $tp_j$  for each empire;
17:  select the weakest colony from the weakest empire and give it to the strongest empire;
18:  update  $N_{col_j}$  of the weakest and strongest empires;
19:  eliminate the empire with no colonies and  $N_{imp} = N_{imp} - 1$ ;
20: end while
21: return the strongest imperialist;

```

Following, the total power of each empire (tp_j) is computed including the power of the imperialist country and the power of its colonies, according to $tp_j = c_j + \sigma \times \sum_{i=1}^{N_{col_j}} c_i$ [15] (line 16). The parameter $\sigma \in [0, 1]$ causes the total power of the empire to be determined by just the imperialist country ($\sigma = 0$) or by the colonies increasing the σ value ($\sigma > 0$). We adopt the value of $\sigma = 0.1$ in our implementation as suggested in [15]. In line 17, ICA selects the empire with the highest tp_j as the best one and increase its number of colonies by one ($N_{col_j} + 1$). The colony of the weakest empire, k , with the lowest c_i is considered as the weakest one, and its number of colonies is reduced by one ($N_{col_k} - 1$). The number of empires also decreases by one ($N_{imp} - 1$) if the weakest empire k has $N_{col_k} = 0$ (line 19). Finally, if more than one empire remained or the difference between the previous best imperialist and the current one is greater than $1e^{-6}$, and the execution time is less than the total time then go to line 8.

4 Experimental Design

In this section, we describe the experimental design used in this approach. We have selected a wide range of instances used in the literature taking into account their complexity, which is given by the number of jobs and machines, and the wide variation of flexibility in the amount of available machines per operation. In this sense, we considered the data set proposed by Brandimarte [17], since the number of jobs ranges from 10 to 20, the number of machines belongs to the set $\{4,15\}$ and the number of operations for each job ranges from 5 to 15, consequently the total number of operations ranges from 55 to 240. Taking into account the flexibility varies between 1.43 and 4.10.

The parameter values of the proposed algorithms are selected based on some preliminary trials. The selected parameters are those values that gave the best results concerning both the solution quality and the computational effort. In

Table 1. Parametric configuration of the SA, ILS, GA, CS and ICA algorithms.

SA	ILS	GA	CS	ICA
MCL 100	$threshold$ 15	μ 50	N 50	N_C 50
	un_move 30	λ 50	p_{LS} 0.1	p_{LS} 0.1
	p_{LS} 0.1	p_c 0.8	p_{best} 60%	N_{imp} 5
		p_m 0.3	p_a 20%	σ 0.1
$stop\ criterion \ \#O \times (\frac{\#O}{2}) \times 30$				

order to make a fair comparison among these algorithms, they are executed during the same time for each problem instance. The parametric configuration used in each algorithm is shown in Table 1.

Because of the stochastic nature of the algorithms, we performed 30 independent runs of each test to gather meaningful experimental data and apply statistical confidence metrics to validate our conclusions. As a no normal distribution is followed by the data, we used the Kruskal-Wallis (KW) test. This statistical study allows us to assess whether or not there were meaningful differences between the compared algorithms with a confidence level of 99%.

5 Experimental Results

In this section, we analyze the quality of results considering the C_{max} values obtained for each algorithm described in Section 3 to solve the FSSSP instances. Additionally, we study the hit rate and the distribution of the normalized gap between the best solution found by each proposed metaheuristic and the best known C_{max} for each instance and for all algorithms.

Analyzing the algorithm performance from the hit rate obtained for each of them (see Table 2), the general trend is that trajectory-based algorithms are more efficient than population-based algorithms. SA obtains the highest hit rate values in the majority of the instances. Regarding the population-based algorithms, GA presents the highest mean hit values, followed by CS and ICA, respectively.

Table 2. Hit rate obtained by SA, ILS, GA, CS, and ICA for all FJSSP instances.

Inst.	SA	ILS	GA	CS	ICA
1	100.00	100.00	100.00	96.67	63.33
2	90.00	53.33	10.00	23.33	0.00
3	100.00	100.00	100.00	100.00	100.00
4	100.00	100.00	40.00	16.67	0.00
5	30.00	3.33	0.00	3.33	0.00
6	33.33	10.00	0.00	0.00	0.00
7	26.67	3.33	0.00	0.00	0.00
8	96.67	100.00	100.00	100.00	100.00
9	93.33	100.00	70.00	6.67	0.00
10	0.00	0.00	0.00	0.00	0.00
Mean	67.00	56.99	42.00	34.66	26.33

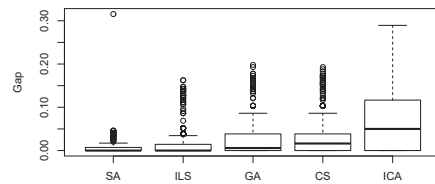
**Fig. 1.** Gap values obtained by SA, ILS, GA, CS, and ICA considering all FJSSP instances.

Table 3. Mean execution time (sec.) to find the best C_{max}

Inst.	SA	ILS	GA	CS	ICA	KW
1	2.27	0.23	0.73	3.20	0.23	+
2	8.23	6.97	8.60	7.33	0.13	+
3	0.00	0.00	0.10	0.03	0.00	+
4	3.50	13.83	58.37	56.93	0.37	+
5	19.40	7.60	37.97	25.80	0.10	+
6	121.13	196.40	164.13	160.83	1.20	+
7	24.23	42.53	65.27	66.17	0.97	+
8	0.00	1.00	2.03	8.83	0.20	+
9	16.03	148.00	421.13	519.27	1.60	+
10	711.87	610.30	658.17	605.93	9.83	+
Mean	90.67	102.69	141.65	145.43	1.46	+

Complementary information to the previous analysis is the distribution of the gap values shown by the algorithms. For that purpose, Figure 1 illustrates these results by means of a box-plot graph. These results support the previous analysis. This allows to remark that the trajectory-based algorithms achieve schedules with the lowest C_{max} values, solving the FJSSP more accurately than the population-based ones.

Regarding the mean execution time to find the best C_{max} , the results presented in Table 3 show that ICA is the fastest algorithm to find its best solutions but their qualities are very poor, as observed in previous analysis. In general, this algorithm ends because one empire conquers the remaining empires with its colonies at the first iterations. As a consequence, ICA does not consume the total execution time computed for each instance. Instead, the longest processing time corresponds to CS and GA. In order to check the confidence of the results a non-parametric test, the KW test, was applied (normality conditions not met). In the last column of the Table 3, the results of this test are shown where the symbol “+” indicates significant differences between the algorithms. Consequently, a post-hoc statistic analysis reveals that CS, GA, and ILS present a similar execution time to find the best solution, but it is noticeable that ILS is able to find better quality solutions than the remaining ones (see Table 2). On the other hand, ILS and SA have statistically similar execution times, but exhibiting significant differences with the population-based algorithms.

Figure 2 shows the distribution of the total number of evaluations carried out by each algorithm. We observe that the algorithms with more evaluations are SA and CS, although SA outperforms CS from the result quality point of view (see Table 2). This means that the SA procedure has a better balance between exploration and exploitation than CS. Instead, ILS and GA do a statistically similar total number of evaluations. However, ILS presents the best behavior, obtaining good solutions in a low number of evaluations. These suggest that ILS does a better exploration and exploitation during the search than GA. Finally, ICA carried out the minimal total number of evaluations and this justifies its poor performance to solve FJSSP.

Considering the quality of results and the computational effort of the algorithms, we can observe two main groups of them and this division corresponds to the classification of the metaheuristics: the trajectory-based algorithms (SA

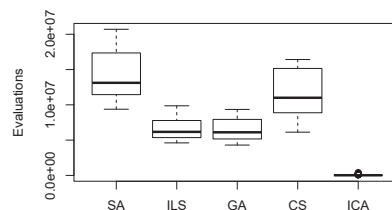
**Fig. 2.** Total number of evaluations done by running SA, ILS, GA, CS, and ICA considering all FJSSP instances.

Table 4. Comparison between SA and metaheuristics from the literature

inst	SA	PVNS	TSPCB	hGA	BEDA	IACO
1	40	40	40	40	40	40
2	26	26	26	26	26	26
3	204	204	204	204	204	204
4	60	60	62	60	60	60
5	172	173	172	173	172	173
6	58	60	65	60	60	60
7	139	141	140	140	139	140
8	523	523	523	523	523	523
9	307	307	310	307	307	307
10	200	208	214	205	206	208

and ILS) and the population-based ones (GA, CS, and ICA). The algorithms in the first group outperforms the algorithms in the second one, showing a relatively good trade-off between solution quality and computational effort. This is highly related with the algorithmic procedure carried out by each one of these metaheuristics and by the number of evaluations per iteration. Particularly, SA obtains the best solution quality (the highest hit rate values and the lowest gap values) in a low mean execution time, exhibiting between exploration and exploitation.

6 Comparison with other algorithms

To determine the goodness of the metaheuristics considered in this work, this section presents a comparison of the results from the best performing algorithm (SA) with several competitive algorithms present in the literature. This allows a comparative assessment of the algorithms for the FJSSP. In this comparison different metaheuristics to solve the FJSSP are considered: *i*) a variable neighborhood search (PVNS) [18], *ii*) a tabu search (TSPCB) [19], *iii*) a hybrid algorithm combining chaos particle swarm optimization with genetic algorithm (hGA) [20], *iv*) a bi-population based estimation of distribution algorithm (BEDA) [21], and finally *v*) an ant colony optimization (IACO) [22].

The C_{max} values of SA and the algorithms included in the comparison are listed in Table 4. From the results, the C_{max} values of SA are equal or lower than the values of remaining algorithms for dealing with almost all ten instances. This observation suggests that the SA developed in this work is a competitive algorithm to solve the FJSSP. Comparisons regarding computational effort are hard to be carried out because the majority of the works do not report the number of evaluations and also, the hardware used for the experimentation has different configurations. Consequently, the relative efficiency of the referred algorithms is difficult to contrast in order to obtain meaningful comparisons.

7 Conclusion

In this article, we have presented a compendium of different metaheuristics in order to solve the FJSSP. In this study two trajectory-based metaheuristics and three population-based ones are included to give a wide spectrum of possible

solutions to the mentioned scheduling problem. The results indicate that the trajectory-based metaheuristics were accurate and efficient for the FJSSP. In the study carried out in this work, SA is the best performing algorithm to solve the FJSSP. Moreover, when SA is contrasted with algorithms in the literature, it also becomes in the best approach. As a consequence, SA gives accurate solutions to this NP-hard problem in an efficient and competitive way.

As future research activities, we plan to widen the scope of the study by including another set of instances with high dimensionality. Furthermore, variants of the FJSSP with more constraints will be evaluated considering the approaches developed in this article.

Acknowledgements

The authors acknowledge the support of Universidad Nacional de La Pampa, and ANPCYT under contract PICTO 2011-0278 and the Incentive Program from MINCYT. The third author is also funded by CONICET.

References

1. M. L. Pinedo, *Scheduling: Theory, Algorithms, and Systems*, 3rd ed. Springer Publishing Company, Incorporated, 2008.
2. M. R. Garey, D. S. Johnson, and R. Sethi, "The complexity of flowshop and jobshop scheduling," *Math. Oper. Res.*, vol. 1, no. 2, pp. 117–129, May 1976.
3. E.-G. Talbi, *Metaheuristics: From Design to Implementation*. Wiley, 2009.
4. G. Vilcot and J.-C. Billaut, "A tabu search and a genetic algorithm for solving a bicriteria general job shop scheduling problem," *European Journal of Operational Research*, vol. 190, no. 2, pp. 398 – 411, 2008.
5. G. Zhang, X. Shao, P. Li, and L. Gao, "An effective hybrid particle swarm optimization algorithm for multi-objective flexible job-shop scheduling problem," *Computers & Industrial Engineering*, vol. 56, no. 4, pp. 1309 – 1318, 2009.
6. G. Zhang, L. Gao, and Y. Shi, "An effective genetic algorithm for the flexible job-shop scheduling problem," *Expert Syst. Appl.*, vol. 38, pp. 3563–3573, 2011.
7. C. Bierwirth, "A generalized permutation approach to job shop scheduling with genetic algorithms," *Operations-Research-Spektrum*, vol. 17, pp. 87–92, 1995.
8. S. Kirkpatrick, C. G. Jr, and M. Vecchi, "Optimization by simulated annealing," *Science*, no. 220, pp. 671–680, 1983.
9. T. Sttzle, "Local search algorithms for combinatorial problems analysis, algorithms and new applications," DISKI Dissertationen zur Kunstliken Intelligenz, Sankt Augustin, Germany, Tech. Rep., 1999.
10. B. Naderi, R. Ruiz, and M. Zandieh, "Algorithms for a realistic variant of flowshop scheduling," *Computers & Operations Research*, vol. 37, no. 2, pp. 236 – 246, 2010.
11. D. S. Yang X-S, "Engineering optimization by cuckoo search," *Int. Journal of Mathematical Modelling and Numerical Optimisation*, vol. 1, pp. 330–343, 2010.
12. A. Ouabarab, X.-S. Y. B. Ahiod, and M. Abbad, "Discrete cuckoo search algorithm for job shop scheduling problem," in *2014 IEEE International Symposium on Intelligent Control (ISIC) Part of 2014 IEEE Multi-conference on Systems and Control*, 2014, pp. 1872–1876.
13. D. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley, 1989.

14. R. Mencía, M. R. Sierra, C. Mencía, and R. Varela, "A genetic algorithm for job-shop scheduling with operators enhanced by weak lamarckian evolution and search space narrowing," *Natural Computing*, vol. 13, no. 2, pp. 179–192, 2014.
15. E. Atashpaz-Gargari and C. Lucas, "Imperialist competitive algorithm: An algorithm for optimization inspired by imperialistic competition," in *Evolutionary Computation, 2007. CEC 2007. IEEE Congress on*, Sept 2007, pp. 4661–4667.
16. G. G. Minetti and C. Salto, "Imperialist competitive algorithm for the flowshop problem," in *Anales del XXI Congreso Argentino de Ciencias de la Computación*, 2015, pp. 1–10.
17. P. Brandimarte, "Routing and scheduling in a flexible job shop by tabu search," *Annals of Operations Research*, vol. 41, p. 157183, 1993.
18. M. Yazdani, M. Amiri, and M. Zandieh, "Flexible job-shop scheduling with parallel variable neighborhood search algorithm," *Expert Systems with Applications*, vol. 37, no. 1, p. 678687, 2010.
19. J.-Q. Li, Q.-K. Pan, P. N. Suganthan, and T. J. Chua, "A hybrid tabu search algorithm with an efficient neighborhood structure for the flexible job shop scheduling problem," *The International Journal of Advanced Manufacturing Technology*, vol. 52, no. 5, pp. 683–697, 2011.
20. J. Tang, G. Zhang, B. Lin, and B. Zhang, "A hybrid algorithm for flexible job-shop scheduling problem," *Procedia Engineering*, vol. 15, pp. 3678 – 3683, 2011.
21. L. Wang, S. Wang, Y. Xu, G. Zhou, and M. Liu, "A bi-population based estimation of distribution algorithm for the flexible job-shop scheduling problem," *Computers & Industrial Engineering*, vol. 62, no. 4, pp. 917 – 926, 2012.
22. L. Wang, J. Cai, M. Li, and Z. Liu, "Flexible job shop scheduling problem using an improved ant colony optimization," *Scientific Programming*, pp. 1–11, 2017.

A Study of Convolutional Architectures for Handshape Recognition applied to Sign Language

Facundo Quiroga¹, Ramiro Antonio¹, Franco Ronchetti¹, Laura Lanzarini¹,
and Alejandro Rosete²

¹ Instituto de Investigación en Informática LIDI, Facultad de Informática,
Universidad Nacional de La Plata
{fquiroga,fronchetti,laural}@lidi.info.unlp.edu.ar,
{ramiro.antonio}@outlook.com

² Universidad Tecnológica de La Habana José Antonio Echeverría
{rosete}@ceis.cujae.edu.cu

Abstract. Convolutional Neural Networks have been providing a performance boost in many areas in the last few years, but their performance for Handshape Recognition in the context of Sign Language Recognition has not been thoroughly studied. We evaluated several convolutional architectures in order to determine their applicability for this problem.

Using the LSA16 and RWTH-PHOENIX-Weather handshape datasets, we performed experiments with the LeNet, VGG16, ResNet-34 and All Convolutional architectures, as well as Inception with normal training and via transfer learning, and compared them to the state of the art in these datasets. We included experiments with a feedforward neural network as a baseline. We also explored various preprocessing schemes to analyze their impact on the recognition.

We determined that while all models perform reasonably well on both datasets (with performance similar to hand-engineered methods), VGG16 produced the best results, closely followed by the traditional LeNet architecture. Also, pre-segmenting the hands from the background provided a big boost to accuracy.

Keywords: convolutional neural networks, sign language recognition, handshape recognition.

1 Introduction

Automatic sign language recognition (SLR) is an important topic within the areas of human-computer interaction and machine learning that also requires the intervention of various knowledge areas, such as image and video processing, computer vision and linguistics. Robust recognition of sign language could assist in the translation process and the integration of hearing-impaired people, as well as the teaching of sign language for the hearing population.

The problem it tackles is how to convert a video where a signer is speaking in some variant of Sign Language to a text that conveys the same meaning

as the signs in the video. SLR implies a pipeline of tasks which starts with the recognition of the shape of the hands of the signer, their movement and position, the shape of lips and face expressions, plus the objects in the scene, in every frame of the video. Afterwards, the signs in the video must be identified, using both information about how signs are made and the syntax and semantics of the sign language. Afterwards, the sentences in sign language must be translated to some target written/spoken language, such as English or Spanish [13, 18]. Von Agris and Cooper present general reviews of the standard techniques of the field [2, 18]

Previous work has shown that the quality of the handshape recognition appears to be the most important factor when recognizing signs [11, 12]. Also, most experiments focused on finding adequate descriptors for the handshapes to be classified with conventional techniques such as SVM or Feedforward Neural Networks. Perhaps the most widely explored descriptors are the geometric descriptors computed on the contour of the hand [18], such as the perimeter, the orientation of the main diagonals, the area, the eccentricity, among others. Another widely used feature for handshape classification is the Histogram of Oriented Gradients (HOG) [3], where the image is divided in a square overlapped grid of small sub-images and it computes a histogram of directions of different gradient in each sub-image. In [13] the Radon transformation is used as a feature for the classification of handshapes.

There have been great advances in the last few years in text, sound, image and video processing by using Convolutional Neural Networks (CNNs), which are simply deep networks that employ convolutional layers [9]. These layers implement a convolution operation and have a sparse connection topology that allows a better processing of images and other signals.

These networks have been recently used for sign language and action recognition [1, 8]. In particular, [8] employs a semi-supervised training procedure with a pretrained Inception [16] network as base classifier to perform handshape recognition, with state-of-the-art results. However, to the best of our knowledge, there has been no systematic study of the applicability of CNNs for handshape recognition. This paper presents experiments on the applicability of various CNNs models (LeNet, Inception, All Convolutional, VGG, ResNet) to the problem of handshape recognition for sign language, using datasets LSA16 [13] and RWTH-PHOENIX-Weather [5].

2 Architectures

In this section we will introduce the convolutional architectures used to compare their performance on the task of handshape recognition. We have chosen a simple convolutional models like LeNet, one of the first convolutional networks, as baseline and also more complex and deep models that have proved their performance on image recognition problems in known datasets such as All Convolutional, VGG, Inception-v3 and ResNet.

2.1 LeNet

The LeNet model was proposed by LeCun et al.[10] in 1998. On their paper, the authors propose a convolutional architecture for OCR and character recognition in documents. It is one of the first convolutional neural networks and a common baseline in the field of Deep Learning.

A standard 2D convolutional layer applies a convolutional operation on an input feature map with two spatial dimensions and a filter/channel dimension. The convolutional mask has size Height*Width*Channels, and is learnable. A layer typically has many such convolutional masks, which determine the number of output channels/filters. Convolutions can be strided to perform downsampling in the same operation.

A max-pooling layer reduces the spatial size of the input to reduce the amount of parameters and computations in the network.

This architecture consists of two sets of convolutional, activation and max-pooling layers followed by two fully-connected layers and finishing with a softmax classifier.

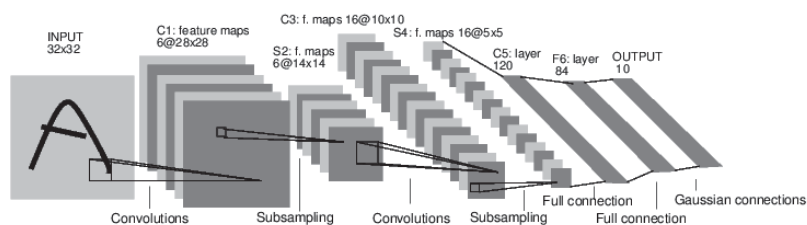


Fig. 1. Diagram of the LeNet5 architecture.

2.2 All Convolutional

This architecture was introduced by Springenberg et al.[15] in 2014. Their model is based on removing the pooling layers of a traditional convolutional network such as LeNet. The spatial downsampling is performed instead by using convolutional layers with a stride greater than 1. The model can achieve state of the art performance on several object recognition datasets such as CIFAR-10, CIFAR-100 and ImageNet, which provides some evidence for the hypothesis that the max pooling operation's contribution to a network comes mostly from the downsampling effect, and not from the maximum computation.

2.3 VGG

The VGG network was first described by Simonyan and Zisserman in their 2014 paper [14]. Its main contribution is providing evidence in favour of stacking

convolutional layers with small filters (3×3) instead of using a single layer with bigger filter sizes (5×5 , 7×7), since stacked 3×3 filters can approximate bigger ones. The network therefore consists of a many 3×3 convolutional filters stacked in groups of two or three, with max-pooling layers between them and finishing with two 4096-dimensional fully-connected layers

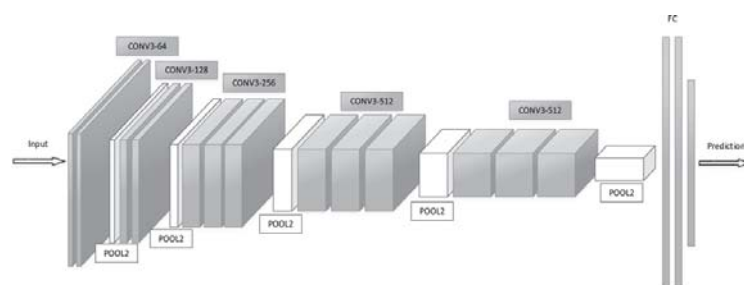


Fig. 2. Diagram of the VGG architecture with its stacked filters with max-pooling layers between them.

This gives a massive number of parameters and learning power. That is why this architecture is costly at runtime and many models that improve it try to reduce the dimensionality of the input to accelerate the computations. The training of VGG is very challenging, specially for the 16 and 19 depth versions. Symonyan and Zisserman trained versions of VGG with less weight layers first. When this smaller networks converged, they were used as initializations for the larger and deeper networks. Two disadvantages of this network are that its training is slow and that its architecture is quite large in terms of memory compared with other state of the art networks.

2.4 Inception

The Inception architecture was introduced by Szegedy et al.[16] in 2014. This original network was called GoogLeNet but later its newer versions were called Inception-vN, with N representing the version named by Google.

The main contribution of this model is the bottleneck layer applied before the parallel convolutions that reduces the computational cost of all the architecture. This bottleneck layer consists employing 1×1 convolutional filters to reduce the spatial input dimension giving a lower number of computations in the subsequent layers.

The goal of this model is to extract features at multiple levels in the same instance of the model. In that way, when a layer of this model receives its input, first, it reduces the input size in a bottleneck layer, and then computes 1×1 , 3×3 , and 5×5 convolutional filters in parallel. After that, the output of this filters are concatenated in a last filter maintaining the channel dimension. The main goal of this architecture was to reduce the computational cost of the network.

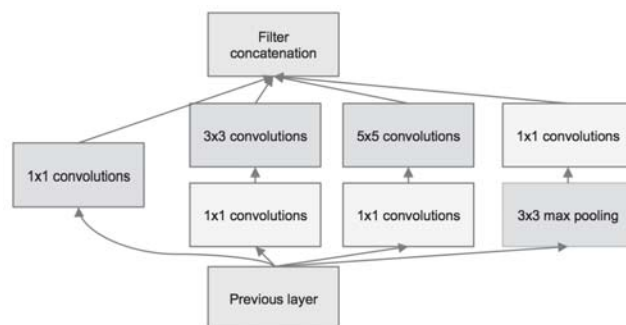


Fig. 3. Diagram of the inception module with its bottleneck layer applied before the 3x3 and 5x5 convolutional filters.

The Inception-v3 was described in 2015 by Szegedy et al.[17] on a paper which proposes updates to the architecture to boost its ImageNet classification accuracy. Even with these updates, the weights of this model are still smaller than those of the VGG model.

2.5 ResNets

This model was presented by He et al.[6] in the year 2015. It presented a modification of the convolutional architecture that allowed training of very deep networks. The problem it addresses is that while adding layers to a network should never lower its performance, since in the worst case the layer should just learn the identity function, in practice networks with many layers. Its main insight was the design of convolutional blocks called *residual modules*, which outputs the sum of the features obtained from the convolutional layers with the input. This design causes residual modules to be initialized to a slightly distorted identity function. Also, the fact that the input is bypassed to the output allows a shortcut path for the gradient to go backwards.

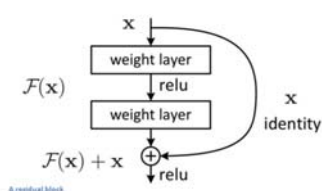


Fig. 4. Diagram of the ResNet architecture.

Hence, the most important contribution of this architecture is the evidence of a very deep network having hundreds of layers can actually be trained using standard stochastic gradient descent through the use of residual modules.

The network has 1x1 convolutional filters in the residual modules to form bottleneck layers as seen in the Inception architecture; this helps reducing the dimensionality inside the residual module, and then restoring it, since a hard constraint of this models is that the dimensionality of the output has to match that of the input. ResNet uses a global pooling layer plus softmax as final classifier.

3 Experiments and Results

We performed experiments with the models described in section 2 and two datasets that contain images of the hands of several people with different handshapes.

3.1 Datasets

LSA16 The LSA16 handshape dataset was originally presented in [13], for the purpose of being the first dataset for handshape recognition of Argentinian Sign Language (LSA), and freely available. It consists of 16 different handshapes of the most used in the LSA. There are 10 experimental subject that made 5 different poses of each handshape, with a total of 800 images in the dataset. The interpreters had colored gloves in their hands, to facilitate the task of segment the hands in the raw image, allowing so to focus on next classification tasks. Figure 5 shows the 16 handshapes configurations presented in the dataset.



Fig. 5. Samples of the 16 handshapes of the pre-segmented version of the LSA16 dataset.

RWTH-PHOENIX-Weather (RWTH) We are employing the labeled images collected for handshape experiments in [8], which contains 3359 samples and 45 classes. We will call this dataset RWTH-PHOENIX-Weather[5] since it contains images form that dataset. Originally presented in 2012. The dataset contains video clippings of a television channel where interpreters translated reports of the state of the weather.

It is important to highlight that this dataset is highly imbalanced: some classes have over 200 examples while others have 5. In order to make a fair comparison, we removed those classes with 7 or less examples, which reduced

the number of classes from 45 to 30. The resulting dataset was had 3289 examples (70 less than the original).

3.2 Comparison of models on LSA16 and RWTH

Methodology We measured the test set accuracy of the models on both datasets. Since there is no official test set for these datasets, and their size is relatively small, we averaged 100 runs of stratified randomized subsampling cross-validation to estimate the test set accuracy. We employed the Feedforward network as a baseline model, and also evaluated the LeNet, AllConvolutional, VGG, ResNet and Inception architectures.

In the case of Inception, we also employed an Inception network that was pre-trained on the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC2012) dataset, and performed transfer-learning experiments by replacing the last layer by either a linear support vector machine (SVM) or a two-layer feedforward network [4]. Only these last layers were retrained, and therefore we effectively employed the output of the penultimate layer of the network as a feature.

We trained all networks using a base learning rate of 0.0007 and the ADAM learning algorithm [7], which uses a per weight adaptive learning rate to remove learning rate effects from the experiment as much as possible. The Feedforward and LeNet architectures were trained for 20 epochs and the AllConvolutional, ResNet, VGG and Inception architectures for 50, 50, 60, and 120 epochs respectively, since they contain significantly more parameters. We determined these hyperparameters via monitoring the training loss and classification error on an initial run to make sure the networks converged. We did not perform hyperparameter optimizations or modify them in subsequent experiments.

In the transfer learning experiments with the Inception network, the SVM for the final layer employed a linear kernel and a regularization coefficient $C = 0.1$. The two-layer feedforward network had a hidden layer of size 512.

Since our datasets are smaller than the standard datasets on which these models are usually evaluated, we reduced the capacity of some models to both avoid overfitting and allow the network to train successfully. For the AllConvolutional architecture, we decreased the size of the filters to (96, 96, 96, 192, 192, 192), while keeping the topology untouched. For the VGG architecture, we used the VGG16 version and set the filter sizes to (32, 32, 64, 64, 64, 64, 64, 128, 128, 128, 128, 128, 128) and feedforward layer sizes to (512, 512). We used the 34-layer version of the ResNet architecture, without modification [6]. The Inception network's filter's sizes were not modified in neither the transfer learning experiments nor the normal end-to-end training. Meanwhile, the LeNet architecture employed four convolutional layers with sizes (32, 64, 128, 256) and 3x3 filters, and a 512-dimensional feedforward layer.

For all models, we found that adding Batch Normalization layers after each max pooling layer and replacing ReLU activation functions for ELUs reduced the training time while achieving the same accuracy. Finally, we used the pre-segmented RGB version of the LSA16 dataset. The code for the experiments is available at https://github.com/midusi/convolutional_handshake

Results In Table 1, we can observe that all models have a lower accuracy on the RWTH dataset, which is expected since it has more classes, unsegmented hand images and class imbalances. The feedforward model has the lowest accuracy in general, also expected. The convolutional architectures all have similar accuracies on each dataset. For the LSA16 dataset they achieve similar or better accuracy than ProbSom, a hand engineered method [13]. The accuracy of the DeepHand model [8] on the RWTH dataset is slightly bigger than for other models (85.50%). It must be noted that DeepHand used an Inception network pre-trained on ILSVRC2014 data, along with a weakly supervised training scheme that employed 1 million example images with noisy labels and then used the RWTH dataset as testing set. No such scheme was applied with the other models. Surprisingly, the simple LeNet architecture achieved relatively good recognition rates, which perhaps is an effect of the reduced size of the datasets, which is more important for bigger models such as VGG and Inception.

<i>Method</i>	<i>LSA16</i>	<i>RWTH</i>
DeepHand* [8]	-	85.50
Probsom* [13]	92.30	-
Feedforward	86.58	60.27
LeNet [10]	95.78	81.19
AllConvolutional [15]	94.56	80.29
VGG16 [14]	95.92	82.88
ResNet [6]	93.49	80.89
Inception [16]	91.98	75.33
Inception+SVM [16]	93.67	78.12
Inception+NN [16]	80.62	75.97

Table 1. Accuracy of various convolutional neural network models on two datasets: LSA16 [13] and handshapes from RWTH-PHOENIX-Weather [8]. Results from methods annotated with * were taken from other papers.

3.3 Comparison of preprocessing alternatives

Methodology We also ran experiments with LeNet on LSA16 to determine whether preprocessing the images provides a boost in accuracy. We chose the LeNet because it gave good results in the previous experiments and also because since it is the simplest convolutional architecture, the results should be more representative of the general family of models. We employed the same methodology for the experiments as before. We tested the model with the raw images, and also with several variants where the background has been segmented out of the image leaving only the hand color pixels. We tested the model with the segmented image (RGB), also converting it to a grayscale image, to a black and white image, and finally calculating the contour mask of the hand.

Results Table 2 shows the accuracies obtained with the different preprocessing schemes presented and the LeNet model. From the increase in accuracy when switching from the raw RGB image to the segmented RGB image, it would



Fig. 6. From left to right: raw image, segmented image, grayscale image, hand segmentation mask, and contour.

appear that the model has trouble separating the background and the hand, and that a segmentation step helps the network to distinguish the handshapes. The Grayscale and Black and White variants seem to have a lower accuracy. This may indicate that the texture information provided by the RGB pixels helps the model (possibly because fingers can be detected more easily).

<i>Preprocessing</i>	<i>Accuracy</i>
Raw (RGB)	83.54
<i>Segmented hand</i>	
RGB	96.18
Grayscale	87.08
Black and White	91.38
Black and White, Contour only	80.64

Table 2. Accuracy of the LeNet model on the LSA16 dataset with different preprocessing schemes.

4 Conclusion

We have performed experiments to evaluate the mean accuracy of various prominent convolutional architectures on two handshape recognition datasets.

In the case of the LSA16 dataset, the models showed a performance on par with the state of the art. For the dataset RWTH-PHOENIX-Weather, the best architecture, VGG16, performed slightly worse than the best published method, but it is difficult to perform a direct comparison since that method employed a pretrained Inception network with a weakly-supervised learning algorithm with 1 million noisily labeled examples.

Taking that into consideration, it would appear that convolutional networks in general perform reasonably well and on the order of hand-engineered methods for the problem of handshape recognition. We also found evidence supporting the hypothesis that a pre-segmentation step can help to improve the accuracy of the recognition for these methods.

In future work, we will focus on architectures that include localization and segmentation networks as previous steps and also allow end-to-end learning.

References

1. Camgoz, N.C., Hadfield, S., Koller, O., Bowden, R.: Using convolutional 3d neural networks for user-independent continuous gesture recognition. In: Proceedings

- IEEE International Conference of Pattern Recognition (ICPR), ChaLearn Workshop (2016)
2. Cooper, H., Holt, B., Bowden, R.: Sign Language Recognition, pp. 539–562. Springer London, London (2011)
 3. Cooper, H., Ong, E.J., Pugeault, N., Bowden, R.: Sign language recognition using sub-units. *J. Mach. Learn. Res.* 13(1), 2205–2231 (Jul 2012)
 4. Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., Darrell, T.: Decaf: A deep convolutional activation feature for generic visual recognition. *CoRR abs/1310.1531* (2013)
 5. Forster, J., Schmidt, C., Hoyoux, T., Koller, O., Zelle, U., Piater, J., Ney, H.: Rwth-phoenix-weather: A large vocabulary sign language recognition and translation corpus. In: *Language Resources and Evaluation*. pp. 3785–3789. Istanbul, Turkey (May 2012)
 6. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 770–778 (2016)
 7. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. *CoRR abs/1412.6980* (2014)
 8. Koller, O., Ney, H., Bowden, R.: Deep hand: How to train a cnn on 1 million hand images when your data is continuous and weakly labelled. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 3793–3802 (2016)
 9. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature* 521(7553), 436–444 (2015)
 10. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 86(11), 2278–2324 (1998)
 11. Ronchetti, F., Quiroga, F., Estrebou, C., Lanzarini, L., Rosete, A.: Sign Language Recognition Without Frame-Sequencing Constraints: A Proof of Concept on the Argentinian Sign Language, pp. 338–349. Springer International Publishing, Cham (2016)
 12. Ronchetti, F., Quiroga, F., Estrebou, C.A., Lanzarini, L.C., Rosete, A.: Lsa64: An argentinian sign language dataset. In: *XXII Congreso Argentino de Ciencias de la Computación (CACIC 2016)*. pp. 794–803. Red de Universidades con Carreras en Informática (RedUNCI) (2016)
 13. Ronchetti, F., Quiroga, F., Lanzarini, L., Estrebou, C.: Handshape recognition for argentinian sign language using probsom. *Journal of Computer Science and Technology* 16(1), 1–5 (2016)
 14. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. *CoRR abs/1409.1556* (2014)
 15. Springenberg, J.T., Dosovitskiy, A., Brox, T., Riedmiller, M.: Striving for simplicity: The all convolutional net. *arXiv preprint arXiv:1412.6806* (2014)
 16. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 1–9 (2015)
 17. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. *CoRR abs/1512.00567* (2015)
 18. Von Agris, U., Zieren, J., Canzler, U., Bauer, B., Kraiss, K.F.: Recent developments in visual sign language recognition. *Universal Access in the Information Society* 6(4), 323–362 (2008)

Active Learning to Reduce Cold Start in Recommender Systems

Luciano Silvi

Facultad de Matemática, Astronomía y Física - Universidad Nacional de Córdoba,
Córdoba, Argentina
{luciano.silvi@gmail.com}

Abstract. Every time a recommender system has a new user, it does not have enough information to generate recommendations with high precision, this is known as cold start. Adapting this problem to a classification problem allow us to apply Active Learning techniques that, as we will see, offer some methods to, given the less possible information about a new user, make right predictions with higher precision than the standard solutions applied in this situation.

Keywords: recommender systems, active learning, cold start

1 Introduction

When a new user starts to use a recommender system, the predictions of the system tend to be unsatisfactory because the system does not have enough information to make reliable predictions about the users tastes. This is known as the *cold start*, and it can span for a long period in the beginning of the user experience, until the system gets to know enough about the user to produce better predictions.

Cold start can be reduced if the system gets to know exactly which information about the user maximizes the accuracy of future predictions, instead of possibly redundant information.

The rest of the paper is organized as follows. In the next Section we discuss some related work on reducing cold start by smart methods. Then, in Section 3, we explain the classic architecture of recommender systems and propose the one we will use to run the experiments detailed in Section 5. In Section 4 we propose the clustering solution necessary to apply the Active Learning cycle. Finally we will present conclusions and different lines of future work that can be explored.

2 Related Work

Sometimes, recommender systems use a hybrid approach by combining collaborative and content-based methods, which helps to deal with certain limitations of pure systems, like the *cold start*. In these cases, different ways to combine both methods into a hybrid recommender system can be classified as follows:

- Implementing collaborative and content-based methods separately and combining their predictions,
- incorporating some content-based characteristics into a collaborative approach,
- incorporating some collaborative characteristics into a content-based approach,
- constructing a general unifying model that incorporates both content-based and collaborative characteristics.

However, although these implementations work well in some cases, usually the system needs to be extended [1]. An option for this extension is to apply different Active Learning techniques [3] or, depending on the state of the system, combining both requesting ratings from the user and using natural acquisition (when the user rates items on their own) [2].

Other approach [4] suggests considering existing users as new users and solve an Active Learning (AL) problem for each of them. In the end, aggregate all solved problems in order to learn how to solve the AL problem for a real new user.

3 Architecture

Classic Recommender Systems architecture is based in k-nearest neighbours, which may produce overfitting and is expensive (although optimizations are applied in actual systems). Other techniques consist of modified versions of Pearson correlation coefficient or the innovative SlopeOne algorithms [7]. All these methods are used to find similarities between users or items (depending on the implementation), usually by looking at the history of items consumed by a given user and, after analysing it, generating the best possible recommendations for the user.

However, the techniques mentioned above often suffer from the same problem: cold start. To deal with it, we present here a different approach by approximating the recommendation problem using classification and Active Learning. More concretely, we will use clustering (applying K-Means) over users profiles (known ratings for each user in the dataset) to build clusters of similar users. For each cluster we will also have a ranking of best rated items, this can be changed for other method if it is desired, although in our experiments the naive implementation worked well. On the other hand we have Active Learning, which is useful to optimize the amount of human labelling needed to achieve a given accuracy, this is why it is adequate for cold start, where accuracy is low.

Then, given a new user, we use AL to request him to rate items which provide more global information gain, using these ratings to find the more appropriate cluster for the user through a classifier. Once we have found the cluster, we use its ranking to make recommendations about items that have not been consumed yet by the user. The ranking for each cluster only considers the ratings for movies given by the users inside the cluster, but this can be extended including and pondering more information like directors, actors, etc.

The application of Information Gain was supported by Settles [5] and Forman [6] as a method to select features.

4 Unsupervised classes

Different configurations can be specified when applying clustering. Here we will generate relatively small groups using KMeans over profiles of 1000 users in the dataset. A small size of each group allow us to assure the similarity between its members.

After running the algorithm over users profiles (known ratings for each user) with random generation of the center in each group (because the dataset is not labelled and we can not previously determine indicative elements for each class) we obtained 10 clusters with sizes between 40 and 182 users in each one. Different numbers of clusters could be specified, but using 10 we found relatively small groups, as we wanted to generate significant recommendations.

5 Experiments

5.1 Dataset

The work proposed in this paper can be applied to any type of Recommender Systems, although in this case we will simulate movies recommendations to users by using the dataset from MovieLens, freely available on its website [10]. More specifically, we will use the dataset "MovieLens 100k", containing 100000 ratings (with possible values from 1 to 5) from 1000 users for 1700 movies. All the users in the dataset had rated at least 20 movies.

5.2 Evaluation Metrics

Different methods are known to evaluate recommender systems depending on what we desire to analyse: ratings predictions, usage of items on the system, coverage, etc. As proposed on Evaluating Recommender Systems [8] and on Recommender Systems (IJCAI 2013) [9], one of the most used metrics to analyse the precision of ratings predictions for a given user is the Root Mean Square Error (RMSE), whose expression is

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2}$$

where \hat{Y} is the vector with predictions made by the system and Y is the vector with the real values for that items. We will use this metric to analyse and compare results of our experiments.

5.3 Methodology

In the following experiments we will use ratings given by users to items in the system to generate new recommendations through different techniques, comparing the baselines methods with the proposed ones using Active Learning.

More concretely, what we are going to do is, for each experiment, build a matrix of preferences for all the users but the first one, who will be used for evaluation, in other words, we are simulating a system under production where a new user is added. Then, we will add preferences for that user under evaluation (from the known ratings) in a progressive way, generating recommendations based on those preferences and with the recommendations we will calculate the RMSE. Each user has less known ratings than the total quantity of items in the dataset, then, when a specific rating is not available, we put a 0 in the corresponding vector's position (remember that possible values for ratings are from 1 to 5, so we can use 0 as a synonym for 'no information').

This process will be applied to every user in the system, repeating the process but evaluating the second user, then the third one, and so on until the last one. Finally we will get an average of the square error for each number of preferences, results that will be shown through figures.

The only difference when applying Active Learning will be that, instead of calculating the error for all the users, it will be calculated only for a part of them, because the other part will be used to train the classifier.

5.4 Baselines

Experiment 1: Using Pearson correlation coefficient This is one of the most basic and used solutions because, in general, it provides acceptable results. Results can be observed on Figure 1, as we can see, the error increases when we add more preferences for the user under evaluation.

This situation usually happens when two users are detected as similar, even when they only share a low amount of similar ratings. Instead of using a 'pure' version of Pearson coefficient, some methods are applied to fix this problem like taking into account the total number of shared ratings, scaling the similarity coefficient when shared ratings are low, etc.

Experiment 2: Using SlopeOne SlopeOne [7] should avoid fake similarities and other common problems of Pearson correlation, because it was created specifically for recommendation and, when more information is added, results are better.

Results (Figure 1) show that, as expected, when more preferences are added, the error decreases. This offer an accented contrast with results from the previous experiment.

5.5 Active Learning

We will use here the Pool-Based Active Learning cycle presented on [11], where the model will be a classifier, and the oracle will be replaced with the known

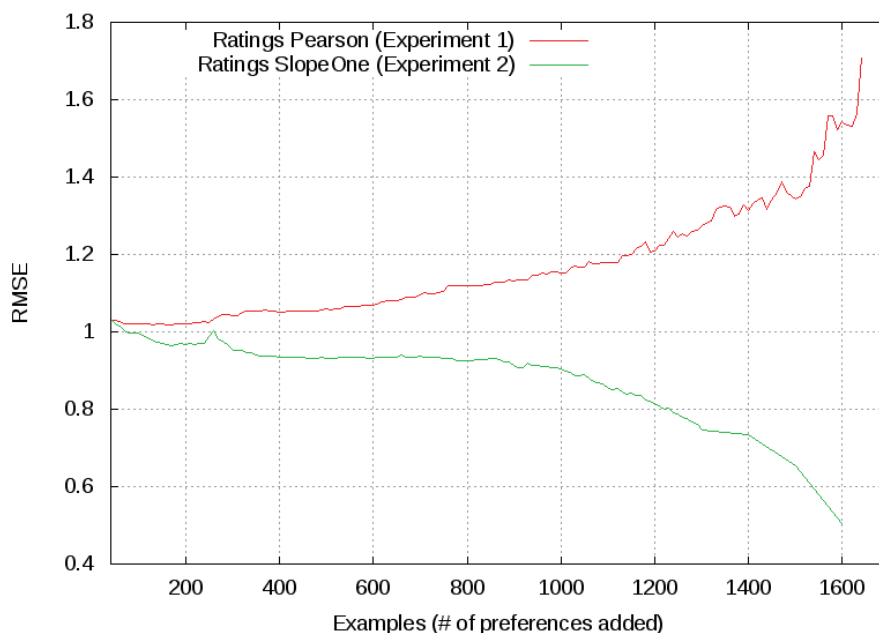


Fig. 1. Baselines

ratings for each user from the dataset. The classifier will be trained with resultant classes from clustering presented before in section 4.

Experiment 3: Using Passive Learning (using only classifier, adding ratings randomly) This experiment will be used to compare its results later with Active Learning. Here we only use a classifier to classify a given user and make recommendations from the ranking of his cluster, the preferences to analyse the improvement of the classification will be added randomly, without using any active learning method.

The classifiers used in this experiment were *Multinomial Naive Bayes* and *Support Vector Machines*. Both have their pros and cons, but the essential idea behind the application of them is to see if a simple classifier (MNB) is enough for our purposes or if other one (SVM) can help us to improve the results.

When analysing the results (Figure 2) we can see that using MNB the error decreases faster than using SlopeOne starting from about 400 known ratings from the user, while, in the beginning, the behaviour of both is similar. Using SVM, error is higher than using SlopeOne, so, if we would like to use only a classifier without AL techniques, MNB should be enough.

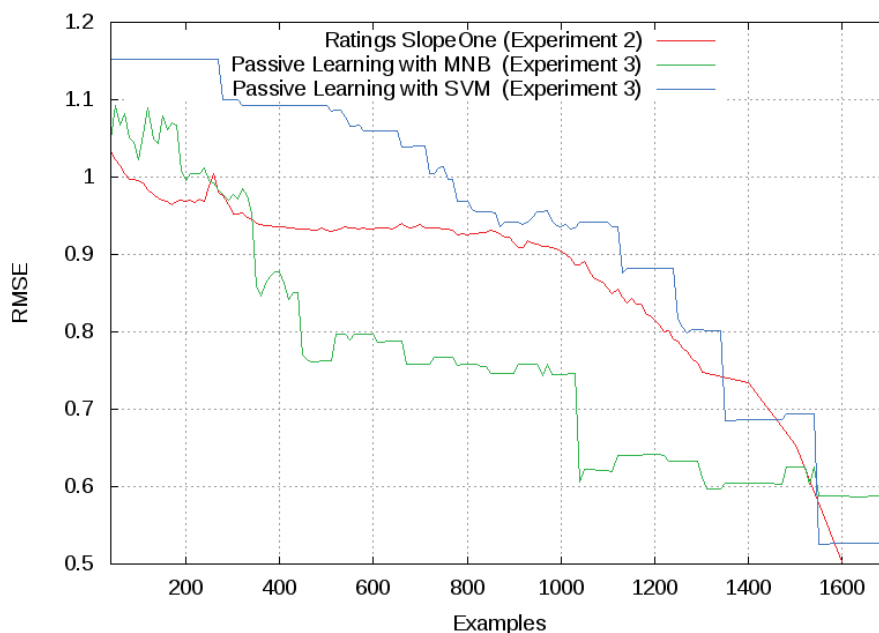


Fig. 2. Comparison of Passive Learning

Experiment 4: Active Learning with MNB and Information Gain As before, in this experiment we will add preferences on each iteration for the user, with the novelty that the preferences will be added depending on how many information they provide to improve the classification. To achieve this, we previously sorted items in the system by Information Gain, then, on each iteration, we add the preference that adds more information and that was not added before for the same user.

Same as before, once the user is classified, the recommendations will be generated using the ranking of his class.

The introduction of Active Learning produced improvements on the results, as we can see in the Figure 3, being the difference with the previous experiment more significant during the first 300 iterations. After that, the behaviour of both are almost similar. However, as we are focusing mainly on cold start, we can observe that the RMSE with this experiment is under SlopeOne, even during the first iterations.

Experiment 5: Trying others classifiers: SVM, DT and LR Looking for a better improvement, we have tried different classifiers in our Active Learning model: Decision Trees, Logistic Regression and Support Vector Machines.

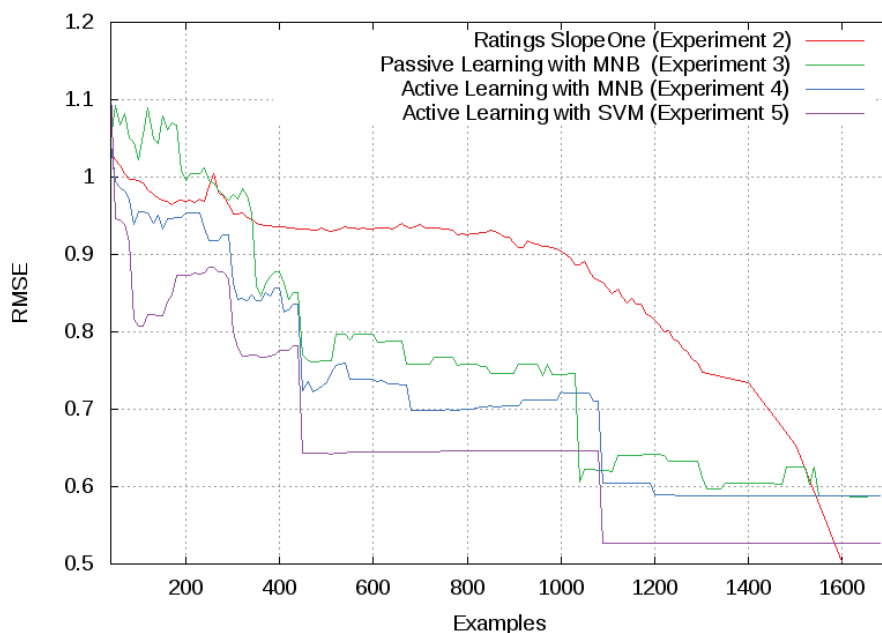


Fig. 3. Comparison of Active Learning

As we can observe on Figure 4, Support Vector Machines offer more precision mainly during the first iterations, with an RMSE of around 0.8.

5.6 Experiments adding users

In the previous experiments, we were focused on measuring RMSE for new users added to a stabilized system.

On the other hand, it is also possible to measure RMSE for the system, that could give us an idea of how fast the system reaches a stable stage when adding new users. To evaluate this, we did the following:

- Start the system with only 50 users, measure their RMSE and take the average as the RMSE of the system.
- Add 50 users on each iteration until reach the total number of users and repeat the calculus from the previous point.

This process was made over the baselines systems (Pearson and SlopeOne) and the one which uses Active Learning with SVM. Results are shown on Figure 5, where we can see that lowest RMSE is obtained when using AL. This allow us to conclude that recommender systems that implement Active Learning have a lower RMSE than other solutions.

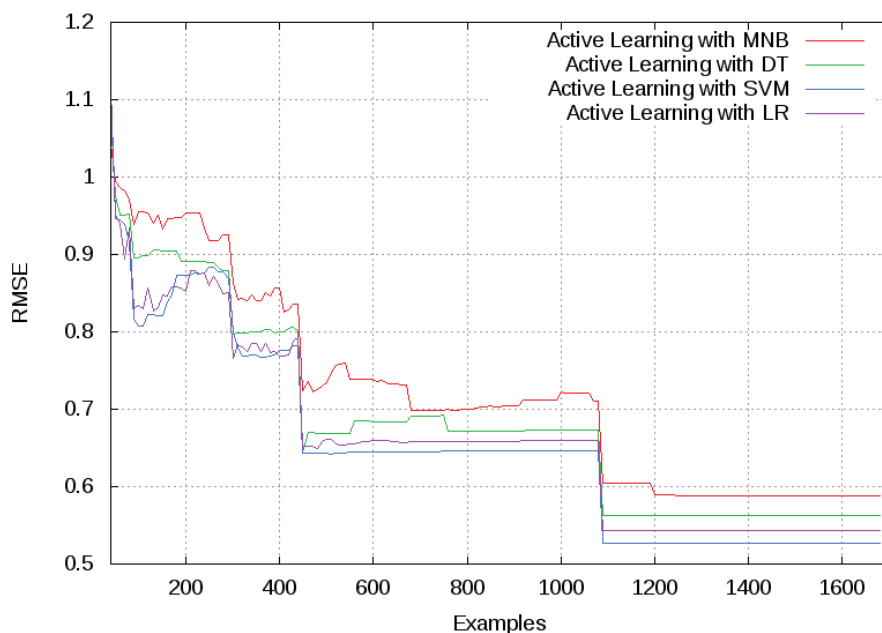


Fig. 4. Comparison of AL with different classifiers

6 Test of Significance

Statistical significance allows us to test that given results have not occurred randomly or due to sampling error. Here, we would like to know if advantages (reduction of RMSE during cold start) obtained using Active Learning and SVM are significant.

As proposed on [8], we will use Wilcoxon signed-rank test. Our null hypothesis will be: 'Results using AL and SVM are not significantly different from results using SlopeOne'. To apply the test, we will take as samples the results of RMSE (for each number of examples) obtained from experiments 2 and 5 (with SVM).

After applying the Wilcoxon signed-rank test, the p-value is $4.3268076925130965 \times 10^{-23}$, so we can see that, even with a low significance level ($\alpha = 0.01$), the p-value is considerably lower, that offers enough evidence to reject our null hypothesis, concluding that the results obtained with AL and SVM are significant with respect to results from SlopeOne.

7 Conclusions and Future Work

Our main goal in this paper was to reduce the system's learning time through maximization of the utility of the information provided by the user. In the ex-

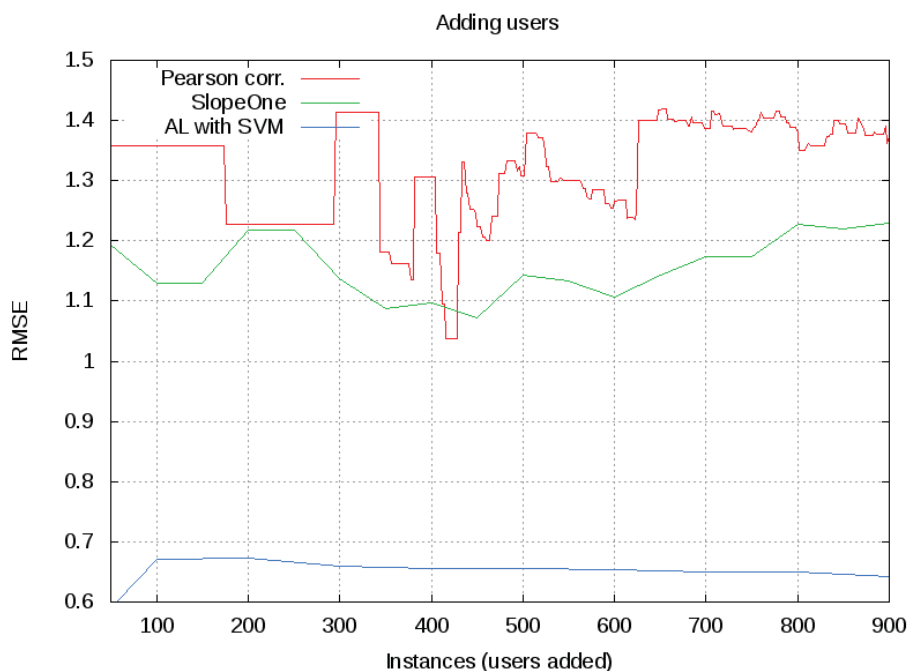


Fig. 5. RMSE adding users

periments we analysed that the best performance during the stage of cold start was achieved by applying Active Learning through information gain to a classification problem where the "new user" has to be located in the cluster with more users similar to him.

The experiments showed a reduction of RMSE and, at the same time, we have reduced the complexity regarding to the common approaches for recommendation (k nearest neighbours, Pearson, SlopeOne, etc.) avoiding the total exploration of search space and allowing the most intensive calculus to be done offline.

As future work, we propose the next lines to be considered:

- **Run the experiments with bigger datasets:** Because of computational limits, the experiments in this paper were done using a relatively small dataset regarding to a real system, where the amount of information could be really big. Then, using bigger datasets would allow us to do more detailed analysis.
- **Run the experiments using a system under production:** Build a system from scratch, or implementing the active learning cycle on an existent system would allow us to measure performance in a real environment, permitting also to evaluate the efficiency and speed because of the offline process.

- **Use an ensemble of classifiers:** As mentioned on "Active Learning for Recommender Systems" [12], one classifier can model in a better way certain users profiles while other one can model a different kind of profiles. In these cases could be convenient, depending on the profile of a given user, choose the classifier that offers a better representation of user's tastes.

References

1. Gediminas Adomavicius and Alexander Tuzhilin. 2005. Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. *IEEE Trans. on Knowl. and Data Eng.* 17, 6 (June 2005), 734-749.
2. Mehdi Elahi, Francesco Ricci, and Neil Rubens. 2012. Adapting to natural rating acquisition with combined active learning strategies. In *Proceedings of the 20th international conference on Foundations of Intelligent Systems (ISMIS'12)*, Li Chen, Alexander Felfernig, Jiming Liu, and Zbigniew W. Ra (Eds.). Springer-Verlag, Berlin, Heidelberg, 254-263.
3. Mehdi Elahi. 2011. Adaptive active learning in recommender systems. In *Proceedings of the 19th international conference on User modeling, adaption, and personalization (UMAP'11)*, Joseph A. Konstan, Ricardo Conejo, Jos L. Marzo, and Nuria Oliver (Eds.). Springer-Verlag, Berlin, Heidelberg, 414-417.
4. Rasoul Karimi, Alexandros Nanopoulos, and Lars Schmidt-Thieme. 2015. A supervised active learning framework for recommender systems based on decision trees. *User Modeling and User-Adapted Interaction* 25, 1 (March 2015), 39-64.
5. Burr Settles. 2011. Closing the loop: fast, interactive semi-supervised annotation with queries on features and instances. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP '11)*. Association for Computational Linguistics, Stroudsburg, PA, USA, 1467-1478.
6. George Forman. 2003. An extensive empirical study of feature selection metrics for text classification. *J. Mach. Learn. Res.* 3 (March 2003), 1289-1305.
7. Daniel Lemire and Anna Maclachlan. Slope One Predictors for Online Rating-Based Collaborative Filtering. *SDM*. Vol. 5. 2005.
8. Shani, Guy, and Asela Gunawardana. Evaluating recommendation systems. *Recommender systems handbook*. Springer US, 2011. 257-297.
9. TD3: Recommender Systems, <http://ijcai13.org/program/tutorial/TD3>
10. MovieLens, <http://grouplens.org/datasets/movielens/>
11. Settles, Burr. Active learning literature survey. *University of Wisconsin, Madison* 52.55-66 (2010): 11.
12. Rubens, Neil, Dain Kaplan, and Masashi Sugiyama. Active learning in recommender systems. *Recommender systems handbook*. Springer US, 2011. 735-767.

An agent-based simulation model using decoupled learning rules to (re)schedule multiple projects

Laura Tosselli¹, Verónica Bogado¹, Ernesto Martínez²

¹ Departamento Ingeniería en Sistemas de Información – UTN FRVM, Argentina
{ltosselli, vbogado}@frvm.utn.edu.ar

² INGAR (CONICET – UTN) – UTN FRSF, Argentina
ecmarti@santafe-conicet.gob.ar

Abstract. Competitive pressures and business globalization have led many organizations, mainly technology-based and innovation-oriented companies, to adopt project-based organizational structures. In a multi-project context within enterprise networks, reaching feasible solutions to the multi-project (re)scheduling problem represents a major challenge, where autonomy and decentralization of the environment favor agent-based simulation. This work presents and validates a simulation-based multi-agent model using the fractal company concept to solve the complex multi-project (re)scheduling problem in enterprise networks. The proposed agent-based model is tested through a set of project instances that vary in project structure, project parameters, number of resources shared, unplanned events that affect them, etc. Results obtained are assessed through different scheduling goals, such as project total duration, project total cost, leveling resource usage, among others to show that decoupled learning rules allow finding a solution which can be understood as a Nash equilibrium for the interacting agents and it is far better compared to the ones obtained with existing approaches.

Keywords: Multi-project (re)scheduling, project-oriented fractal organization, Multi-agent simulation.

1 Introduction

Multi-Project (re)scheduling is considered as a NP-hard problem, thus becoming a difficult task for project leaders when many tasks and resources are involved which prevents the application of optimization-based methods to find a repaired schedule [1]. Thus, in a multi-project context within enterprise networks, there are conflicting constraints and interrelationships among projects that cause an increase of the complexity, making project rescheduling a difficult problem to be addressed under real-time pressure and selfish behavior of concerned agents [2]. Particularly, the unplanned events impact on due dates and milestones achievement, budget consumption, and resources usage, which in turn affect timing and quality of projects goals delivery because of inefficient responses to such events. Then, the multi-project (re)scheduling problem within an enterprise network requires techniques and tools that allow decision-making in an autonomous and decentralized manner.

However, solving (re)scheduling problems are generally carried out in a centralized way, which makes it difficult to respond to unplanned events autonomously involving selfish agents. Thus, to achieve the project (re)scheduling problem dealing with unplanned events in a decentralized and efficient manner, organizations must adopt structures with a high degree of flexibility, allowing the reconfiguration of their constituent parts. In this context, the concept of a *Fractal Company* [3] has been proposed, which combined with the project-oriented approach gives rise to the needed flexibility, autonomy, decentralization, self-organization, among other features. Therefore, this proposal is based on the Project-oriented Fractal Company Model developed in [4] to implement a *Multi-agent System* (MAS) composed by autonomous and selfish entities that interact to solve the multi-project (re)scheduling problem. These entities do not have complete knowledge about the other strategies and payoffs, and there are also constraints related to the exchange of information. The establishment of *client-server* and *delegation* relationships among end-managers (projects) and mean-managers (resources) provides the flexibility for rescheduling tasks at different abstraction levels.

Several works related to the project scheduling problem have been presented using the agent-based approach [5, 6, 7] where an initial schedule is obtained from the interaction among relevant entities in the problem domain. However, many of these proposals are based on the assumption that the project schedule will be implemented as initially defined. This situation is not representative of the project management environment since 80% of project schedules are affected by unplanned events during their execution, affecting deadlines, costs and estimated resource usage [8].

In this context, arises the need to solve the multi-project (re)scheduling problem in a decentralized and autonomous way, providing metrics and indicators that support the decision-making process when unplanned events affect the initial schedule. Therefore, the objective of this work is to present and validate a Multi-agent Simulation model to deal with the complex multi-project (re)scheduling problem incorporating in each of the agents that compose it, a decoupled learning rule called *Learning by Trial and Error* [9], following different objectives and obtaining (re)scheduling solutions from the strategic interactions among them. Then, the Agent-based Simulation Model allows evaluating the emergent solution. A standard and representative set of multi-project problem instances [10] that vary in project structure, project parameters and number of shared resources is used to test the presented Multi-agent System against other multi-agent proposals. Results obtained are encouraging and demonstrate the applicability of agent-based simulation.

2 Problem Description

In this section, the multi-project (re)scheduling problem based on a project-oriented fractal model for enterprise networking is considered. This problem can be defined as follows:

- A set of I projects to be (re)scheduled within a project-oriented fractal organization, whose managers (agents) must respond to unplanned events and disturbances. These projects are interdependent and run to a certain extent in

parallel. Additionally, each project has properties such as deadline, budget, estimated start and finish time, resource requirements, dependence relationships, precedence relationships, among others. Some of these properties are considered as domain constraints. Each project i is considered as a *fractal unit* and it consists of a *recursive* structure, where a project can be composed in sub-projects, and these sub-projects can be composed of other sub-projects, and so on and so forth. For this reason, the recursive decomposition of a project continues until its minimum expression, i.e., a task that is also considered as a project. Each *fractal unit* or *project* is composed of a project manager and a managed object, where each managed object can be an *end* (e.g., a project, a sub-project, a phase) or a *mean* (a resource). Each project manager carry out functions such as: negotiation, (re)scheduling and has learning capabilities.

- A set R of links of *delegation* and *client-server* type. The delegation relationships are the result of recursive structure and client-server relationships are the result of negotiations, which are auction-based interactions among fractal units that manages projects or resources, where each manager of a project demands resources for fulfilling its scheduled work, whereas the managers of resources sell their specific capabilities and skilled workforce to different projects.
- A set of K available resources with properties such as: processing capacity, maximum availability, resource cost per unit time, list of scheduled tasks for processing. Many of these resources are shared among projects.
- A set of E unplanned events that may affect the execution of the initial schedule. In this work, the types of events are related to variations in the availability of resources.

The scheduling of different goals considered in this work are related to the multiple constraints that should be traded off in a multi-project context, such as time, cost and resources. These goals are:

- (1) Minimize the total duration of the project set I .
- (2) Minimize the total cost of the project set I .
- (3) Minimize the variation of resource utilization in each time period (solution stability) of project set I .

The first goal in addressing the rescheduling of multiple projects is related to the *Resource Constrained Project Scheduling Problem* (RCPSP), which seeks to obtain, for each project, a schedule that have the shortest possible total project duration (TPD), subject to limited amounts of the resources available and precedence/synchronization relationships among projects/tasks [10, 11]. This objective is defined as follows for any project i :

$$TPD = \min \{ \max \{ ft_i \} \} \quad (1)$$

The term ft_i is defined as the finalization time of latest task in the project i schedule.

The second scheduling goal considered in this work aims to obtain solutions that present the lowest possible total cost (TC), giving rise to solutions to rescheduling of multiple projects corresponding to the *Resource Investment problem* (RIP) [12]. In this context, the objective is defined as follows:

$$TC = \min \sum_{p=1, k=1}^{n, q} ral_{pk} \quad (2)$$

where ral_{pk} is the cost of each resource allocation that form part of the schedule for project i .

Finally, the third scheduling goal is related to the problem named *Resource leveling problem* (RLP) [12], and it aims to find solutions (schedules) that have the least variation in the resource usage between different time periods (stability). This goal SS tries to minimize the impact on the original schedule of the changes made:

$$SS = \min \sum_{k=1}^q (ur_{k,t} - ur_{k,t-1}) \quad (3)$$

where $(ur_{k,t})$ represents the amounts of resource usage in the time period t . This goal allows measuring the adaptability of the solution found to the changes in the project environment and leveling the resource usage within the availability constraints to solve resource over-allocations [2, 12].

3 Multiagent-based Methodologies

3.1 Agent-based modeling and simulation

Agent-based modeling and simulation (ABMS) is concerned with creating computational models such artificial societies where self-interested and autonomous agents interact among them following different goals [13]. In this proposal, the developed simulation model seeks to imitate the sequence of strategic interactions in the Project-oriented Fractal Organization when responding to events while addressing the multi-project (re)scheduling problem in a decentralized and distributed way. Thus, the multi-project (re)scheduling problem is modeled as a repetitive and interdependent *game* [9], compound of N players (agents) and M stages (repeated iterations), where each of the agents pursues its specific goals in a selfish way. Furthermore, the simulation model is a key tool to evaluate emergent scheduling behaviors from established *client-server* and *delegation* relationships among interacting agents.

The proposed agent-based model is composed of two kinds of agents, *Project* and *Resource* agents. Each agent class has different goals. A *Project* agent (PA) manages a project, a sub-project or a task, aims at minimizing its duration and total cost. On the other hand, a *Resource* agent (RA) manages a resource aiming to maximize its profits. These individual goals (micro-level) lead to global and emergent behaviors (macro-level) which are aligned to the scheduling objectives defined in *Section 2*. The defining properties for project and resource agents are described in Table 1 and Table 2 respectively.

Table 1. *Project agent properties*

Property	Description
pId	Project ID
$pLevel$	Project level
$pGoal$	Project goal
$estST$	Estimated project start time
$estFT$	Estimated project finish time
$deadline$	Project latest finish time
PR_i	Set of precedence relationships between other projects and project i
DR_i	Set of dependence relationships between other projects and project i
req_{ij}	Resource requirements for task j on project i
P_{u_i}	Payoff function for project agent i defined as: $\min TPD_i + TC_i$

Table 2. *Resource agent properties*

Property	Description
rId	Resource ID
$maxAvailability$	Resource Maximum available capacity per unit time
$rCapacity$	Resource Processing Capacity (%)
$rType$	Define local or global resource type
R_{u_i}	Payoff function for resource agent k defined as: $\max \sum_{p=1}^n ral_{pk}$

The above defined agent types for interacting managers used in both defining an initial scheduling and during (re)scheduling, where the latter is called for when an unplanned event occurs. The interactions of these agents in the proposed game, through decentralized mechanisms, provide an easier way to obtain a schedule flexibly adapted to the unplanned events mentioned above, where their decisions depend only on the negotiations that they carry out. To respond to abnormal disturbances, the agents must incorporate learning capabilities for (re)scheduling when choosing alternative courses of action and gaining experience from the situations that may arise. In the next section, *learning by trial-and-error* [9] using decoupled rules is incorporated in the proposed simulation model to advantage. The incorporation of decoupled learning rules to solve the multi-project rescheduling problem allows selecting at each strategic interaction among agents those actions that present a greater benefit to achieve agent's goals, according to their payoff functions. More importantly, the resulting repaired schedule corresponds to near Nash equilibrium for all concerned agents.

3.2 Learning by Trial and Error

One of the distributed learning techniques having as its main characteristic resorting to uncoupled learning rules for each individual agents is the proposal of Young et al. [9]. In this learning by trial-and-error approach, interacting agents respond to changes in their own rewards, which are affected, indirectly, by other agents' actions. Uncoupled learning rules can be implemented in environments where the agents cannot observe what the other agents might be doing. That is why this learning method has great potential for distributed optimization problems and complex adaptive systems involving many autonomous agents interacting strategically.

The trial and error learning technique is incorporated in each agent to obtain the best course of action to be followed through a simple implementation at each decision stage in the repeated game (presented in previous sub-section) to solve the multi-project (re)scheduling problem and to obtain a feasible schedule. Each interacting agent's state (z_i) is made up of a type of *mood* m_i (which can be *content*, *discontent*, *hopeful* or *watchful*), a reference action (\bar{a}_i) and a reference reward (\bar{u}_i). According to his mood and the obtained reward, each agent decides its next action to be selected. In Table 3, the properties for the implementation of decoupled learning rules in each agent are shown. The parameter ε defines the rate of exploration in the next game stage.

Table 3. Agent properties related to *Learning by Trial and Error* rules.

Property	Description
$z_i = (m_i, \bar{a}_i, \bar{u}_i)$	Reference state of agent.
$z_i = (m_i, a_i, u_i)$	Current state of agent.
ε	The exploration/exploitation rate of each agent

At the end of each game stage, once all the agents define their actions to follow, they simultaneously collect the stage payoffs according to the selected actions and agents' state transitions occur. The transitions between the different states that an agent can experiment during the simulation of the agent-based model are depicted in Fig. 1. The transitions "a" to "k" depends on agents' actions and payoffs only, while transitions "l" and "m" depends on a probability function called *response* function, which is monotonically increasing in u_i and monotonically decreasing in \bar{u}_i for project agents, and conversely for resource agents.

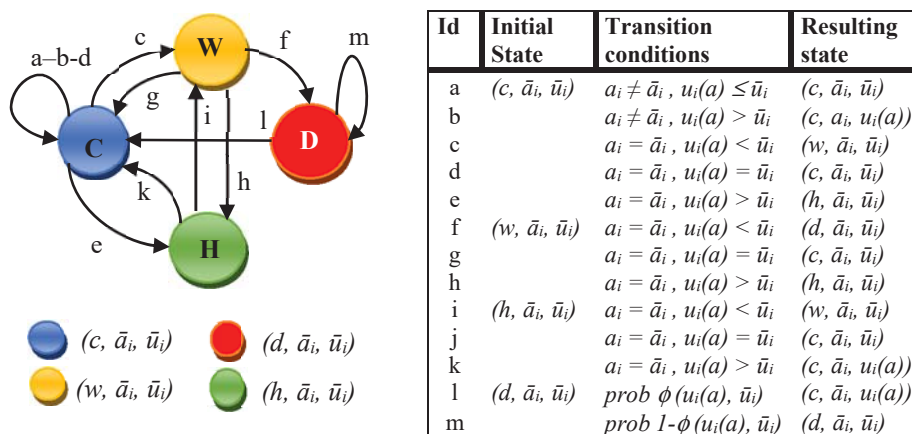


Fig. 1. State transitions experienced by agents using decoupled learning rules.

4 Computational results

This section presents and discusses results obtained through different simulations of the Multi-agent model for a Project-oriented Fractal Organization (MAS-MPR) developed

to solve the multi-project (re)scheduling problem in enterprise networking. These results allow assessing the response and performance of the MAS-MPR. The Multi-agent model was implemented in **Netlogo**, a multi-agent simulation environment for agent-based modeling that allows generating emergent behaviors resulting from ongoing interactions among autonomous, learning agents [13]. As a consequence of the initial concurrent projects scheduling received as input to the multi-agent model, the global (re)schedule of such projects is obtained in an autonomic and decentralized manner, so as to accomplish the scheduling goals defined in *Section 2*. The multi-agent model presented in this work is based on the prototype presented in [14]. This prototype has been extended to incorporate new goals based on additional project constraints such cost and resource leveling (since only the project duration is considered in [14]), and to eliminate assumptions defined during the initial implementation, such as that a resource can only process one task at a time.

Instances of multi-project problems with different features are used to test the proposed multi-agent system (Table 4). These problem instances are available in <http://www.mpsplib.com>. To adapt the problem instances to the simulation problem, different levels (L1, L2, L3) are defined for each problem, describing the number of project agents (managing either a project, sub-project or task) that interact at each level of fractal hierarchy. Then, four resource agents are available for each project, divided into shared (G) and local (L). In the scheduling phase of the multi-agent simulation model, 10 games per problem instance were considered, and 10 stages for each game are defined. Thus, the overall number of experiments performed in this work is 400.

Table 4. Multi-project problem instances

Set of Instances	L1	L2	L3	N. resources	
				G	L
MPJ30_a2	1	2	30	1	3
MPJ30_a5	1	5	30	1	3
MPJ30_a10	1	5	30	2	2
MPJ30_a20	1	5	30	2	2

In the virtual simulated world implemented using Netlogo, each estimated project schedule is graphically depicted by means of a Gantt diagram, on which changes will be made according to each agent action during interactions (Fig. 2). At each stage, agent decisions define the multi-project schedule, which is recorded as output of such stage. The MAS-MPR simulated response and performance can be evaluated through domain indicators (project total duration, project cost, leveling resources) based on obtained agent's payoffs after simulation execution, which can be utilized by project leaders to assess the multi-project (re)scheduling problem generated by unplanned events, thus providing a tool for find a solution in which all agents have no incentives to deviate, hence is a Nash equilibrium.

The scheduling efficiency for each project in the solution generated by the MAS-MPR is evaluated using a measure of performance defined as *Average Project Delay* (APD) [5], which in this work is calculated as follows:

$$APD = \frac{1}{N} \sum_{i=1}^N (ft_i - edd_i) \quad (4)$$

where f_i is the finalization time of the latest task in the generated schedule for project i , edd_i is the estimated due date for project i (considered as the critical path for the project), and N represents the number of game stages.

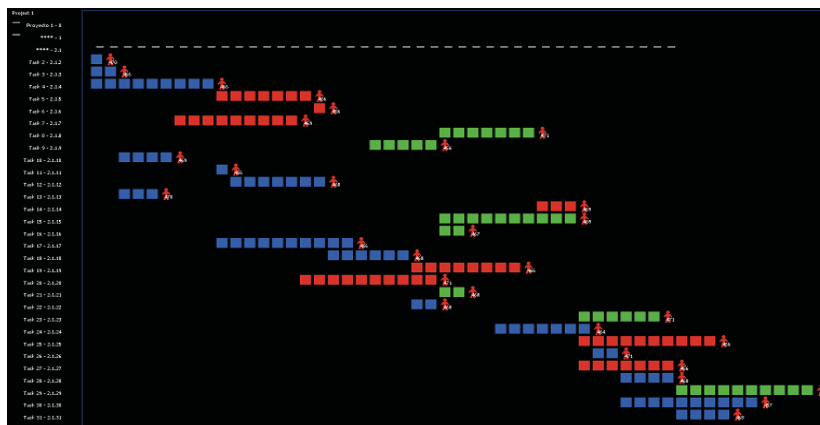


Fig. 2. Example of a project schedule obtained of MAS-MPR simulation as result of one stage.

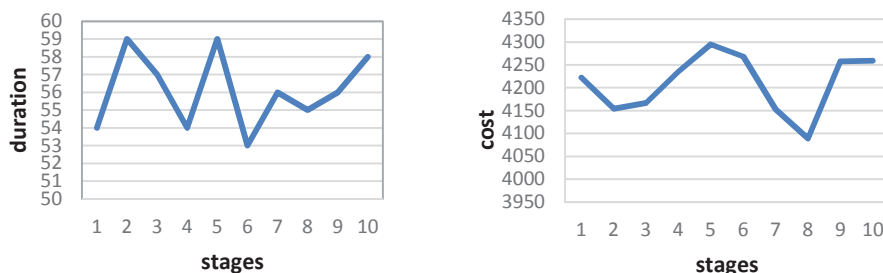
The APD values obtained through the simulation of the MAS-MPR model are shown in Table 5, together with the values for the different proposals presented in [5]. The project leader can analyze the solutions provides by the MAS-MPR, and assess each problem instance to obtain the best solution, i.e., the schedule that presents the minimum APD value.

Table 5. Results obtained by different proposals

Set of Instances	MAS-MPR	CMAS/ES	CMAS/SA	DMAS/ABN	MAS/CI
MPJ30_a2	10,59	13,8	12,8	15,9	19
MPJ30_a5	14,82	18,52	19,28	21,2	27
MPJ30_a10	69	84,4	99,3	87,5	84,14
MPJ30_a20	154	198,4	223,95	207,96	182,05

The obtained results in Table 5 vividly demonstrate the advantages of incorporating learning to the interacting agents in the resolution of the multi-project (re)scheduling problems, comparing against the results obtained in other proposals. From Table 5 it is seen that the MAS-MPR obtains better results for the four subsets of presented problems. These results correspond to the objective evaluation defined as minimization of the project total duration. Furthermore, the MAS-MPR provides other results to evaluate the different solutions obtained. In Fig. 3, the project total duration and costs after a game simulation is shown. To facilitate calculating the costs of the resulting project schedule, different processing costs were assigned to the shared resources. Thus, the project leader can choose the solution that is closest to the estimated duration and cost. For example, the global schedule in the *stage 6* presents the minimum duration, whereas in the *stage 8* presents the lowest cost. In addition, if the stage 2 is analyzed, it is observed that the cost of the project decreases, but the total duration increases. This

represents the situation in which project agents are selecting the most expensive resources. Thus, the project leader can analyze these two variables, namely time and cost, and then choose the generated schedule that fulfills the restrictions of its project.



(a) Multi-Project Total Duration in a game

(b) Multi-Project Total Cost in a game

Fig. 3. Project indicators in a simulated game of MAS-MPR.

It is worth noting that when using the MAS simulation model, computational times to find the solution to the multi-project (re)scheduling problem allows evaluating MAS-MPR response time in seconds. In this work, the average of such time is obtained from the following relation (where M is the number of stages played in a game):

$$AACT = \frac{1}{M} \sum actual_clock_time \quad (5)$$

To obtain a feasible schedule, the required real time (in seconds) changes from one problem instance to another. The average time simulation for each instance simulated in this work is shown in Table 6. The MAS-MPR also can carry out the multi-project (re)scheduling in similar simulation times to those used during the scheduling phase. The simulations were executed on a computer Intel Core I5 (2.5GHz, 8GB RAM).

Table 6. Average actual clock time for a simulated stage on MAS-MPR.

Instance	AACT (per stage)
MPJ30_a2	11 seconds
MPJ30_a5	70 seconds
MPJ30_a10	255 seconds
MPJ30_a20	720 seconds

5 Conclusions and Future Work

Multi-project (re)scheduling is considered a critical problem for organizations, mainly in those that share resources. Its importance is due to the inclusion of many resources, tasks, precedence constraints, unforeseen events, etc. In this context of high uncertainty and dynamism, project leaders claim for tools to evaluate alternative repaired schedules on a short notice, allowing the analysis of possible scenarios before making decisions. These scenarios can include unavailable resources, new tasks, technological complexity that causes delays, among other unplanned events that can generate project abnormal situations. In this work, the multi-project (re)scheduling problem is solved through a

multi-agent model of the project-oriented fractal organization that decreases the inherent environment complexity when unplanned events occur. This proposal through the agent-based simulation aims to provide different indicators that makes room for the project leader to evaluate project (re) scheduling alternatives, taking into account unplanned events that may arise during its project execution. Thus, multi-project (re)scheduling in real time using the proposed MAS, where feasible schedules emerge from simple learning mechanisms, permits an early analysis of different (re) scheduling solutions before the total project execution. Therefore, agents do not need to know the complete status of the project schedule, nor the actions taken by other agents. The results obtained highlight indicators that permit the analysis of the project total duration, cost, leveling resources, etc., and provide a quality evaluation of the MAS generated schedule. Current research work is about incorporating the multi-agent simulation models in project management tools with an user-friendly interface supporting the decision-making process and allowing training activities for project leaders that favor seeking a solution to solve the problem of project (re)scheduling in the framework of game theory and multi-agent learning.

References

1. Garey, M., Johnson, D. Computers and Intractability: A Guide to the Theory of NP Completeness. New York: W. H. Freeman (1979).
2. Project Management Institute, Guía de los Fundamentos para la Dirección de Proyectos (Guía del PMBOK), 4ª ed. (2008).
3. Warnecke, H.J. The Fractal Company: A Revolution in Corporate Culture, Springer-Verlag, Berlin (1993).
4. Canavesio, M., Martínez, E. Enterprise modeling of a project-oriented fractal company for SMEs networking, Computers in Industry, vol. 54, pp. 794-813 (2007).
5. Adhau, S., Mittal, M., & Mittal, A. A multi-agent system for distributed multi-project scheduling: An auction-based negotiation approach. Engineering Applications of Artificial Intelligence, vol. 25, pp. 1738-1751 (2012).
6. Bodea, C., Badea, I., Mogos, R. A Multi-Agent System with Application in Project Scheduling. Management & Marketing, vol. 6, No.4, pp. 573-590 (2011).
7. Araúzo, J., Pajares, J., López-Paredes, A. Gestión eficiente de carteras de proyectos. Dyna Ingeniería e Industria, vol. 84 (9), pp. 761-772 (2009).
8. Chaos Report. Standish Group. In <https://www.infoq.com/articles/standish-chaos-2015>, last access 07/07/2017.
9. Young, H.P., Learning by trial and error-Games and Economic Behavior, vol. 65, pp. 626-643(2009).
10. Homberger, J., A (μ, λ) -coordination mechanism for agent-based multi-project scheduling. OR Spectrum (2009).
11. Artigues, C., Demasse, S., Néron, E. Resource-constrained project scheduling- Wiley Online Library (2008).
12. Shahsavari, A., Najafi, A.A., Niaki, S.T.A. Three self-adaptive multi-objective evolutionary algorithms for a triple-objective project scheduling problem. Computers & Industrial Engineering, Elsevier, vol. 87, pp. 4-15 (2015).
13. Railsback, S., Grimm, V. Agent-Based and Individual-Based Modeling: A practical Introduction, Princeton University Press (2012).
14. Tosselli, L., Bogado, V., Martínez, E. Un Enfoque de Sistemas Multiagente para la Gestión Ágil de Riesgos en la Compañía Fractal Mediante la (Re) Planificación de Proyectos. Conaiisi 2015, Bs. As., Argentina (2015).

Deep Neural Networks for Shimmer Approximation in Synthesized Audio Signal

Mario Alejandro García and Eduardo Destefanis

Universidad Tecnológica Nacional Facultad Regional Córdoba, Argentina
mgarcia@frc.utn.edu.ar

Abstract. Shimmer is a classical acoustic measure of the amplitude perturbation of a signal. This kind of variation in the human voice allow to characterize some properties, not only of the voice itself, but of the person who speaks. During the last years deep learning techniques have become the state of the art for recognition tasks on the voice. In this work the relationship between shimmer and deep neural networks is analyzed. A deep learning model is created. It is able to approximate shimmer value of a simple synthesized audio signal (stationary and without formants) taking the spectrogram as input feature. It is concluded firstly, that for this kind of synthesized signal, a neural network like the one we proposed can approximate shimmer, and secondly, that the convolution layers can be designed in order to preserve the information of shimmer and transmit it to the following layers.

Keywords: shimmer, voice quality, deep learning, deep neural network, convolutional neural network

1 Introduction

Shimmer is a classical acoustic measure of the amplitude perturbation of a signal. This kind of variations in the human voice allows to characterize some properties, not only of the voice itself, but on the person who speaks [1].

Shimmer value is associated to voice quality [2–7], state of mind [8–13], age [14] and gender [15] of people. There are many research works that use shimmer (among other measures) with goals ranging from pathologies detection [6, 16, 17] to the improvement of human-machine interfaces through the estimation of the intensionality of a spoken phrase [19]. Regarding synthesized voices, Yamasaki et al. show in [18] that a certain shimmer level increases the degree of naturalness.

The application of deep learning techniques is the state of the art in automated audio analysis, with the detection of pronounced phonemes and the identification of the person that speaks as main objectives [20–26], but also used to detect emotions, age, gender, etc. [27–33].

Classifiers based on neural networks can be divided into two groups according to the type of input features, those using previously calculated acoustic measures [10, 14] and those using raw audio [22, 24, 25] or spectral data [21–23, 28–31, 34, 35]. In [26] a hybrid approach is applied by adding shimmer and other measures

to improve the recognition achieved with spectral data. It is important to clarify, for first group of classifiers, that shimmer calculation has a major complication, it depends on the previous detection of the fundamental frequency (f_0) of vocal cords vibration. It is difficult to estimate f_0 in pathological voices [36, 37]. The estimation of the actual f_0 value is still a research topic [36–40]. Regarding the second group of classifiers, it is not possible to know whether the outputs are influenced by the shimmer value of the signal.

1.1 Objectives

The objective of this work is to make an estimation of the value of shimmer in a synthesized audio signal through a neural model. The neural network must combine convolutional layers and feed forward layers. The inputs to the neural model will be the spectral values of the signal.

The main contributions of design a shimmer estimation neural network from the spectral features of an audio signal are, on the one hand, the procurement of a f_0 independent shimmer calculation method, and on the other, to answer the question about the extent to which amplitude disturbances of the original audio can influence the output of a deep learning model with raw audio or spectral data input. In other words, how much shimmer information is preserved to the last layers of the model.

1.2 Shimmer Calculation

There are different versions of shimmer. The most important difference between them is the window size (number of f_0 cycles) used in the calculation. Some versions can be seen in [41].

The chosen version of shimmer for this work is the proposed by Klingholz and Martin [42], also known as *Relative Shimmer*.

Relative Shimmer, hereinafter referred to as "shimmer", is a way of measuring the cycle-to-cycle amplitude perturbations of the fundamental frequency of a signal. It is shown as a *perturbation/total amplitude* relation.

$$shimmer = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |A_i - A_{i+1}|}{\frac{1}{N} \sum_{i=1}^N |A_i|} \quad (1)$$

where N is the number of periods of f_0 that the signal has and A_i is the maximum amplitude into the i period.

2 Methods and Materials

2.1 Neural Models

Deep learning models with ascending complexity were generated for problems of shimmer approximation. First, shimmer was approximated for f_0 variable, k

constant and f_{mod} constant. Then, shimmer was approximated for f_0 variable, k variable, and f_{mod} constant. Finally, a model was found to approximate shimmer with f_0 , k and f_{mod} variable.

In all cases, spectral audio data (instead of raw audio) were used as input features. There are two reasons, the improvement of training performance due the dimension reduction, and because in this manner models work in a similar way as the human auditory system, where spectral division is performed in the basilar membrane of the cochlea and not by the neurons of the auditory cortex [43].

2.2 Data

Audio Audio data without harmonics was generated. As in [1] the amplitude modulation of human voice was approximated by a sinusoidal wave. The expression of each audio signal $y(t)$ was:

$$y(t) = \frac{1}{1+k} \sin(\alpha + 2t\pi f_0)(1 + k \sin(\beta + 2t\pi f_{mod})) \quad (2)$$

where t is time [sec], f_0 is the frequency of vocal fold vibration [Hz], f_{mod} is the modulatory frequency [Hz], k is the constant of the amplitude modulator sensibility, α and β are constant to handle the phase of the signal to be modulated and the modulating signal respectively.

For the training, test and validation data generation, random values were taken with uniform distribution. f_0 got values in [200, 1000] Hz range, f_{mod} in [5, 10] Hz, k in [0, 0.4], α and β in $[0, 2\pi]$.

250 ms of audio generated with $f_0 = 200$ Hz, $f_{mod} = 8$ Hz and $k = 0.4$. is shown in Fig. 1.

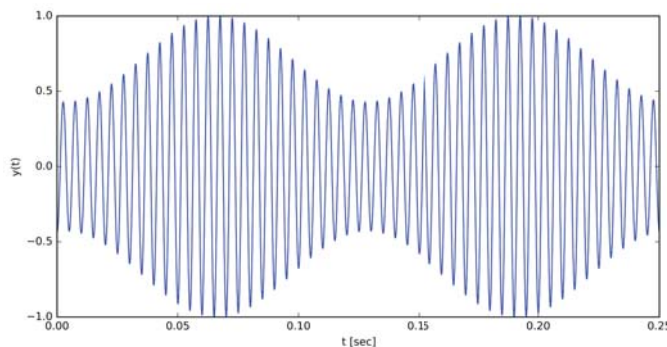


Fig. 1. Generated audio for $f_0 = 200$ Hz, $f_{mod} = 8$ Hz and $k = 0.4$

Training data To train each model, one 2500 elements training dataset, one 500 elements testing dataset and one 500 elements validation dataset were generated. Each element is composed by shimmer (Eq.(1)) value to be estimated and the spectrogram of generated audio.

Due to the fact that f_0 is known at the time of generating the audio, shimmer value can be calculated accurately.

The spectrogram is calculated on 2 sec of audio generated with 44100 *samples/sec*. For the calculation a *Tukey(0.25)* window of width = 256 was used, which determines a structure of shape 129 x 393 (frequency/time) containing the spectral density of the signal.

Fig. 2 shows the values of the second, third and fourth rows (index 1 to 3) of the spectrogram of signal in Fig. 1.

Spectrograms data and shimmer data was normalized between 0 and 1.

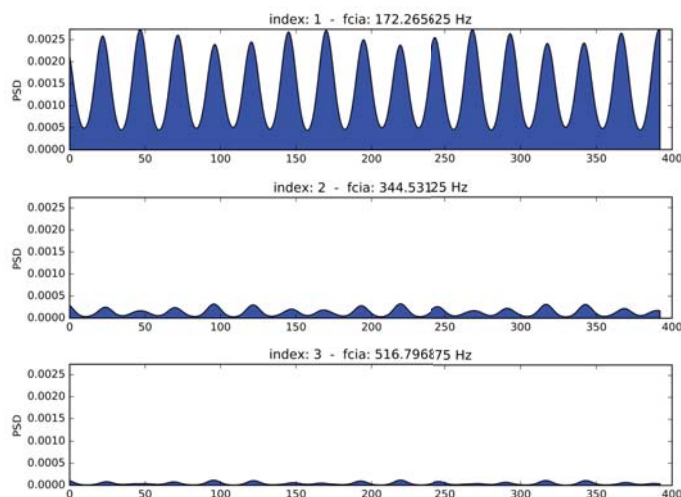


Fig. 2. Three rows with higher average value of Power Spectral Density (PSD) in spectrogram of audio generated with $f_0 = 200 \text{ Hz}$, $f_{mod} = 8 \text{ Hz}$ and $k = 0.4$

3 Results

An initial analysis was performed with f_0 , f_{mod} and k known data. It was found that a neural network with dense connections is able to calculate shimmer value with high precision if it gets f_0 , f_{mod} and k as input features. Optimal structure of this network was empirically found. This is a three layer network, two layers of 20 neurons with $\tanh()$ activation function and a linear neuron as output. In next models, convolution layers are used at the initial part of the network, and then, dense layers of 20, 20 and 1 neurons. The function of convolution layers

is to calculate the values of f_0 , f_{mod} and k in order to dense layers calculate shimmer.

It was noted that only the first 15 rows of the spectrogram (lower frequencies) would have significant information. Then, only for the scope proposed in this work, the rest of the frequencies were eliminated. Then, spectrogram shape change from 129×393 to 15×393 . This provides a important performance improvement.

3.1 Shimmer approximation with f_0 variable

Without harmonics, the calculation of f_0 from the spectrogram is simple, it is enough to obtain the energy average value weighted by the frequency that each spectrogram row represents. As expected, a network such as Fig. 3, where each complete row of the spectrogram connects to a neuron of an *Average pooling* layer, is able to perform the weighted average of frequencies and calculate the shimmer value in densely connected layers. Tests were performed with f_0 in $[200, 1000]$ Hz range, $k = 0.4$ and $f_{mod} = 8 Hz$. A satisfactory approximation was achieved, with a mean square error (MSE) $< 10^{-4}$.

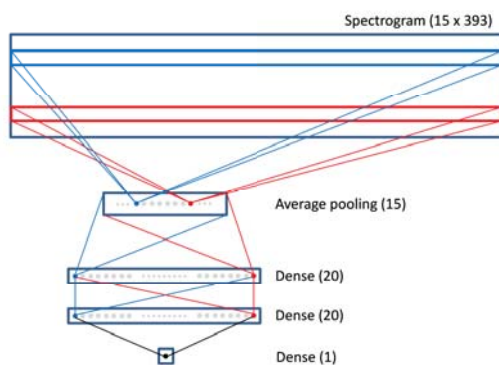


Fig. 3. Shimmer approximation model on signals with f_0 in $[200, 1000]$ Hz range, $f_{mod} = 8 Hz$ and $k = 0.4$. Each neuron in the *Average pooling* layer has a complete frequency of the spectrogram as its visual field. The activation function of hidden dense layers neurons is $\tanh()$ and the output neuron is linear.

3.2 Shimmer approximation with f_0 and k variable

The value of k affects inversely the area under the energy curve of the spectrogram. Therefore, information about the value of k can be obtained through the energy average of the spectrogram. The model presented in the previous section preserves the information necessary to estimate the energy average. Tests were

performed with audio data for f_0 in $[200, 1000]$ Hz range, k in $[0, 0.4]$ range and $f_{mod} = 8$ Hz . Results were satisfactory again. The model approximates shimmer with an $MSE < 10^{-4}$.

3.3 Shimmer approximation with f_0 , k and f_{mod} variable

For f_0 in the range $[200, 1000]$ Hz , k in the range $[0, 0.4]$ and f_{mod} in the range $[5, 10]$ Hz it was necessary to create a more complex model than the previous one. Shimmer depends on the modulation frequency, so a new transformation is necessary (the first one was the transformation from time domain to frequency domain in the spectrogram). The new (second) transformation is performed in a convolution layer at the initial part of the model (Fig. 4).

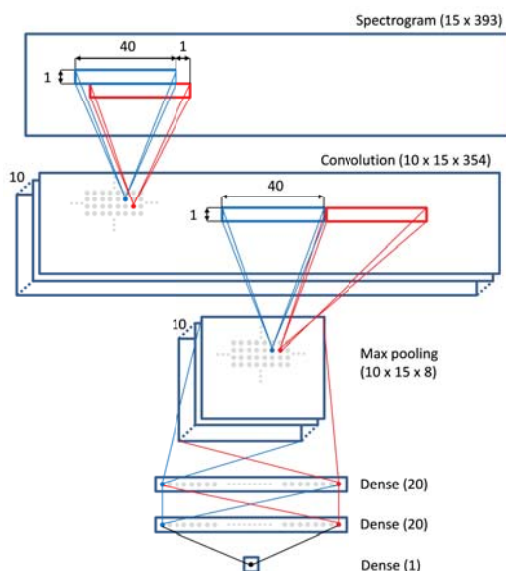


Fig. 4. Shimmer approximation model on signals with f_0 in the range $[200, 1000]$ Hz , k in the range $[0, 0.4]$ and f_{mod} in the range $[5, 10]$ Hz . The shape of convolutional layer windows is 1×40 , strides 1×1 . Convolutional layer has 10 sub-layers. The shape of *max pooling* layer windows is 1×40 , strides 1×40 . The network finish with three dense layers of 20, 20 and 1 neurons.

Convolutional layer Each convolution layer neuron is connected to spectrogram through a *height* = 1 and *width* = 40 window. Convolution is performed on a single frequency (*height* = 1) so that the f_0 detail level needed in the dense layers is not lost. The 40-element width is the minimum required to hold a cycle

of $\min(f_{mod})$. The number of elements of the spectrogram per modulation cycle (C) for a spectrogram of width W_s and an audio length L is:

$$C = \frac{W_s}{L \times \min(f_{mod})} = \frac{393 \text{ elements}}{2 \text{ sec} \times 5 \text{ Hz}} = 39.3 \text{ elements/cycle}$$

The window displacement in both directions is 1 step. It imply that on the frequency dimension there is no overlap, and in time dimension there are 39 overlapping elements between the windows of adjacent neurons. Finally, according to these definitions, the shape of each convolution filter or sub-layer is 15 x 354. The convolution layer is formed by 10 sub-layers. This amount is a compromise between performance and the detail level of f_{mod} on the information sent to dense layers. Neurons of this layer have linear activation function. Weights are initialized with orthogonal random values. An attempt was made to initialize them with wavelet families for sinusoidal waves between 5Hz y 10Hz, but no improvement was achieved on the prediction accuracy.

Max pooling layer The neurons in the *max pooling* layer have a 1 x 40 window size on the convolutional layer. Again, *height* = 1 allows allows f_0 information be able to be transmitted to dense layers with no losing details. The 40-element width extends the visual field of this layer neurons to 2 cycles of $\min(f_{mod})$ on the spectrogram. In this way, the output value is invariant to the modulation signal translations. There is no overlap between the windows, so the size of each of the 10 sub-layers is 15 x 8 neurons.

The outputs of *max pooling* layer are connected to three layers with dense connections equal to those of the previous model.

For this model, 20 training tests were performed. The size of training dataset was 2500 elements. In all cases, results were compared with a test dataset (500 elements) during training and a validation dataset (500 elements) at the end. The best result, with 150 training cycles, obtained a $MSE = 5.8 \times 10^{-5}$ on the test dataset. In Fig. 5 expected and calculated shimmer values are displayed in ascending order for the 500 elements of test dataset.

4 Conclusion

It was verified that, for simple audio signal modulated in amplitude by a sinusoidal wave, with variable parameters of fundamental frequency, modulating frequency and modulation sensitivity, it is possible to obtain a neural model able to approximate the value of shimmer.

Under the conditions presented in this paper, it is possible to calculate shimmer without knowing f_0 . Moreover, it can be affirmed that if the first layers of a deep neural network respects the structure of the second presented model, this neural network is able to use the value of shimmer, internally calculated, to perform other classifications.

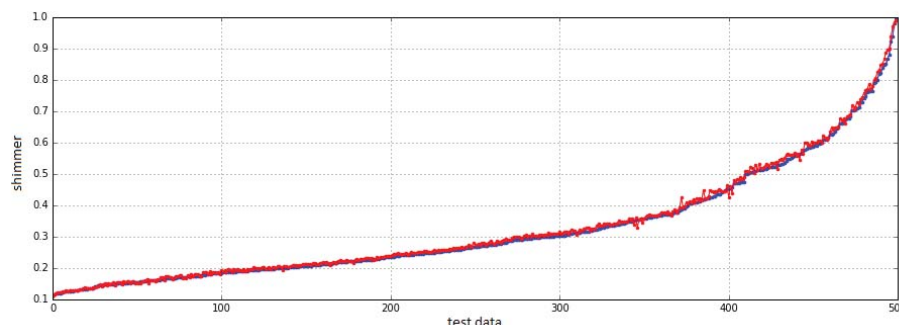


Fig. 5. Normalized shimmer. Expected (blue) vs. calculated (red) for elements in the testing dataset (in ascending order of shimmer value).

5 Future Works

It is planned to extend the analysis, first by expanding the ranges of f_0 y f_{mod} , then adding harmonics and noise to the synthesized signals. Finally, it is planned to analyze the behavior of deep learning models for shimmer calculation on natural voices.

References

1. Jafari, M., Till, J. A., Law-Till, C. B.: Interactive effects of local smoothing window size and fundamental frequency on shimmer calculation. *Journal of Voice*, vol. 7.3. pp. 235–241. (1993)
2. Nieto, R. G., Marín-Hurtado, J. I., Capacho-Valbuena, L. M., Suarez, A. A., Bolaños, E. A. B.: Pattern recognition of hypernasality in voice of patients with Cleft and Lip Palate. *Image, Signal Processing and Artificial Vision*. pp. 1-5. IEEE. (2014)
3. Holi, M. S.: A hybrid model for neurological disordered voice classification using time and frequency domain features. *Artificial Intelligence Research*, vol5.1, pp. 87. (2015).
4. Freitas, S. V., Pestana, P. M., Almeida, V., Ferreira, A.: Integrating voice evaluation: correlation between acoustic and audio-perceptual measures. *Journal of Voice*, vol 29.3, 390–e1. (2015)
5. Little, M. A., Costello, D. A., Harries, M. L.: Objective dysphonia quantification in vocal fold paralysis: comparing nonlinear with classical measures. *Journal of Voice*, vol 25.1. pp 21–31. (2011)
6. Lopes, L. W., Simões, L. B., da Silva, J. D., da Silva Evangelista, D., e Ugolino, A. C. D. N., Silva, P. O. C., Vieira, V. J. D.: Accuracy of Acoustic Analysis Measurements in the Evaluation of Patients With Different Laryngeal Diagnoses. *Journal of Voice*, vol. 31.3. pp. 382-e15. (2017)
7. Hillenbrand, J.: Perception of aperiodicities in synthetically generated voices. *The Journal of the Acoustical Society of America*, vol83.6. pp. 2361-2371. (1988)

8. Li, X., Tao, J., Johnson, M. T., Soltis, J., Savage, A., Leong, K. M., Newman, J. D.: Stress and emotion classification using jitter and shimmer features. ICASSP, vol. 4. pp. IV-1081. IEEE. (2007)
9. Kotti, M., Stylianou, Y.: Effective emotion recognition in movie audio tracks. ICASSP. pp. 5120-5124. IEEE. (2017)
10. Jacob, A.: Speech emotion recognition based on minimal voice quality features. ICCSP. pp. 0886-0890. IEEE. (2016)
11. Sondhi, S., Vijay, R., Khan, M., Salhan, A. K.: Voice analysis for detection of deception. KICSS. pp. 1-6. IEEE. (2016)
12. Palo, H. K., Mohanty, M. N., Chandra, M.: Sad state analysis of speech signals using different clustering algorithm. NGCT. pp. 714-718. IEEE. (2016)
13. Schuller, B., Steidl, S., Batliner, A.: The interspeech 2009 emotion challenge. Tenth Annual Conference of the International Speech Communication Association. (2009)
14. Kim, H. J., Bae, K., Yoon, H. S.: Age and gender classification for a home-robot service. In Robot and Human interactive Communication. RO-MAN. pp. 122-126. IEEE. (2007)
15. Teixeira, J. P., Fernandes, P. O.: Jitter, shimmer and hnr classification within gender, tones and vowels in healthy voices. Procedia Technology, vol 16, 1228-1237. (2014)
16. Tsanas, A., Little, M. A., Fox, C., Ramig, L. O.: Objective automatic assessment of rehabilitative speech treatment in Parkinson's disease. Neural Systems and Rehabilitation Engineering, IEEE Transactions, vol. 22.1. pp 181-190. (2014)
17. Gómez-Coello, A., Valadez-Jiménez, V. M., Cisneros, B., Carrillo-Mora, P., Parra-Cárdenas, M., Hernández-Hernández, O., Magaña, J. J.: Voice Alterations in Patients With Spinocerebellar Ataxia Type 7 (SCA7): Clinical-Genetic Correlations. Journal of Voice, vol. 31.1. pp. 123-e1. (2017)
18. Yamasaki, R., Montagnoli, A., Murano, E. Z., Gebrim, E., Hachiya, A., da Silva, J. V. L., Tsuji, D.: Perturbation Measurements on the Degree of Naturalness of Synthesized Vowels. Journal of Voice, vol 31.3, 389-e1. (2017)
19. Kotti, M., Patern, F.: Speaker-independent emotion recognition exploiting a psychologically-inspired binary cascade classification schema. International journal of speech technology, vol. 15.2. pp. 131-150. (2012)
20. Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. R., Jaitly, N., Kingsbury, B.: Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. Signal Processing Magazine, vol. 29.6, 82-97. IEEE. (2012)
21. Mitra, V., Sivaraman, G., Nam, H., Espy-Wilson, C., Saltzman, E., Tiede, M.: Hybrid convolutional neural networks for articulatory and acoustic information based speech recognition. Speech Communication, vol. 89. pp 103-112. (2017)
22. Collobert, R., Puhrsch, C., Synnaeve, G.: Wav2letter: an end-to-end convnet-based speech recognition system. arXiv preprint arXiv:1609.03193. (2016)
23. Amodei, D., Ananthanarayanan, S., Anubhai, R., Bai, J., Battenberg, E., Case, C., Chen, J.: Deep speech 2: End-to-end speech recognition in english and mandarin. International Conference on Machine Learning. pp. 173-182. (2016)
24. Palaz, D., Collobert, R.: Analysis of cnn-based speech recognition system using raw speech as input (No. EPFL-REPORT-210039). Idiap. (2015)
25. Sainath, T. N., Kingsbury, B., Mohamed, A. R., Ramabhadran, B.: Learning filter banks within a deep neural network framework. IEEE Workshop on ASRU. pp 297-302. IEEE. (2013)
26. Farrús, M.: Jitter and shimmer measurements for speaker recognition. 8th Annual Conference of ISCA. pp. 778-781. (2007)

27. Gu, Y., Li, X., Chen, S., Zhang, J., Marsic, I.: Speech Intention Classification with Multimodal Deep Learning. Canadian Conference on Artificial Intelligence. pp. 260–271). Springer. (2017)
28. Chang, J., Scherer, S.: Learning Representations of Emotional Speech with Deep Convolutional Generative Adversarial Networks.arXiv preprint arXiv:1705.02394. (2017)
29. Ghosh, S., Laksana, E., Morency, L. P., Scherer, S.: Representation Learning for Speech Emotion Recognition. INTERSPEECH. pp. 3603–3607. (2016)
30. Mao, Q., Dong, M., Huang, Z., Zhan, Y.: Learning salient features for speech emotion recognition using convolutional neural networks.IEEE Transactions on Multimedia,vol. 16.8. pp 2203–2213. (2014)
31. Ma, X., Yang, H., Chen, Q., Huang, D., Wang, Y.: DepAudioNet: An Efficient Deep Model for Audio based Depression Classification. Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge.pp. 35-42. ACM. (2016)
32. Abumallouh, A., Qawaqneh, Z., Barkana, B. D. : Deep neural network combined posteriors for speakers’ age and gender classification. CT-IETA. pp. 1–5. IEEE. (2016)
33. Qawaqneh, Z., Mallouh, A. A., Barkana, B. D.: Deep neural network framework and transformed MFCCs for speaker’s age and gender classification.Knowledge-Based Systems, vol.115. pp. 5–14. (2017)
34. Liu, Y., Wang, X., Hang, Y., He, L., Yin, H., Liu, C.: Hypemasality detection in cleft palate speech based on natural computation. ICNC-FSKD. pp. 523–528. IEEE. (2016)
35. Cummins, N., Epps, J., Ambikairajah, E.: Spectro-temporal analysis of speech affected by depression and psychomotor retardation.Acoustics, Speech and Signal Processing. pp. 7542-7546. IEEE. (2013)
36. Teixeira, J. P., Gonçalves, A.: Algorithm for jitter and shimmer measurement in pathologic voices.Procedia Computer Science, vol.100, pp. 271–279 (2016)
37. Shahnaz, C., Zhu, W. P., Ahmad, M. O.: A new technique for the estimation of jitter and shimmer of voiced speech signal. Electrical and Computer Engineering Canadian Conference. pp. 2112–2115. IEEE (2006)
38. Dong, B.: Characterizing resonant component in speech: A different view of tracking fundamental frequency.Mechanical Systems and Signal Processing, vol.88. pp. 318–333. (2017)
39. Liu, B., Tao, J., Zhang, D., Zheng, Y.: A novel pitch extraction based on jointly trained deep BLSTM Recurrent Neural Networks with bottleneck features. Acoustics, Speech and Signal Processing International Conference. pp. 336–340. IEEE. (2017)
40. Schlotthauer, G., Torres, M. E., Rufiner, H. L.: A new algorithm for instantaneous F0 speech extraction based on ensemble empirical mode decomposition. 17th European Signal Processing Conference. pp. 2347–2351). IEEE (2009)
41. Buder, E. H.: Acoustic analysis of voice quality: A tabulation of algorithms 19021990. Voice quality measurement. pp. 119–244. (2000)
42. Klingholz, F., Martin, F.: Quantitative spectral evaluation of shimmer and jitter. J Speech Hear Res, vol 28.2. pp 169–174. (1985)
43. Schnupp, J., Nelken, I., King, A.: Auditory neuroscience: Making sense of sound. MIT press (2011)

ESSIM-EA applied to Wildfire Prediction using Heterogeneous Configuration for Evolutionary Parameters

Miguel Méndez-Garabetti^{1,2}, Germán Bianchini¹, Paola Caymes-Scutari^{1,2},
María Laura Tardivo^{1,2,3}, and Verónica Gil-Costa^{2,4}

¹ Laboratorio de Investigación en Cómputo Paralelo/Distribuido, (LICPaD)
UTN-FRM, Mendoza, Argentina.

² Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET).

³ Departamento de Computación, Facultad de Ciencias Exactas, Físico-Químicas y
Naturales (UNRC), Córdoba, Argentina.

⁴ Laboratorio de Investigación y Desarrollo en Inteligencia Computacional (LIDIC)
Universidad Nacional de San Luis - Ej. de los Andes 950 - San Luis, Argentina.

`mmendez@mendoza-conicet.gob.ar, gbianchini@frm.utn.edu.ar,`
`pcaymesscutari@frm.utn.edu.ar, lauratardivo@dc.exa.unrc.edu.ar, gvcosta@`
`unsl.edu.ar`

Abstract. Wildfires devastate thousands forests acres every year around the world. Fire behavior prediction is a useful tool to cooperate in the coordination, mitigation and management of available resources to fight against this type of contingencies. However, the prediction of this phenomenon is usually a difficult task due to the uncertainty in the prediction process. Therefore, several methods of uncertainty reduction have been developed, such as the Evolutionary Statistical System with Island Models based on Evolutionary Algorithms (ESSIM-EA). ESSIM-EA focuses its operation on an Evolutionary Parallel Algorithm based on islands, in which the same configuration of evolutionary parameters is used. In this work we present an extension of the ESSIM-EA that allows each island to select an independent configuration of evolutionary parameters. The heterogeneous configuration proposed, at the island level, with the original methodology in three cases of controlled fires has been contrasted. The results show that the proposed ESSIM-EA extension allows to improve the quality of prediction and to reduce processing times.

Keywords: wildfire prediction, HPC, uncertainty reduction, metaheuristics.

1 Introduction

Wildfires are considered the natural phenomenon that causes most damages and losses worldwide. As a recent fact we can cite the great fire that occurred in the

provinces of La Pampa, Río Negro and Buenos Aires between December 2016 and January 2017, where fire consumed more than 2.5 million acres [1]. This kind of situations often brought about when there are long periods of drought, low humidity, high temperatures and considerable winds. In this context it can be very useful to have tools or methods that minimize the negative effects caused by wildfires. Such is the case of the wildfire behavior prediction, where we intend to determine with some time in advance the future behavior of the fire. It allows to identify the areas that are most at risk of being caught by fire, and therefore, to plan the efficient use of available resources. The behavior of a wildfire depends on several variables or factors, which are not usually known at the time of the fire.

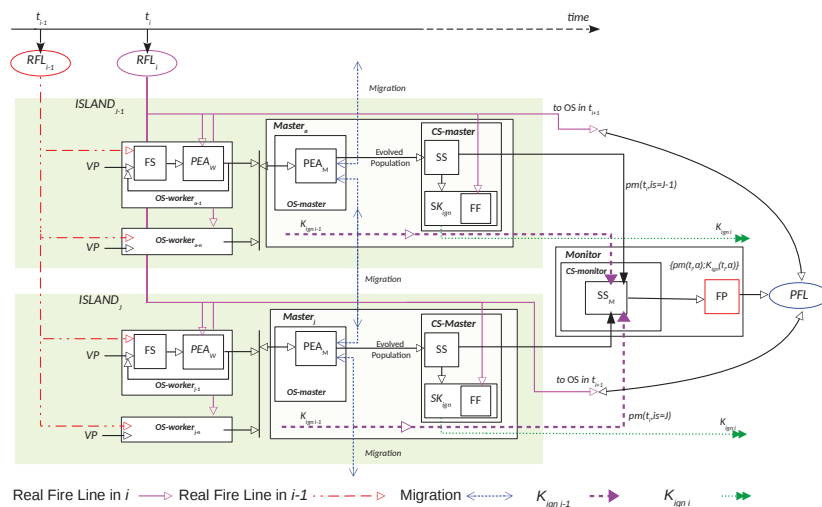


Fig. 1: ESSIM-EA: **FS**: Fire Simulator; **PEA**: Parallel Evolutionary Algorithm; **PEA_F**: PEA (Fitness Evaluation); **OS**: Optimization Stage; **SS**: Statistical Stage; **SK_{ign}**: search K_{ign} ; **K_{ign}**: Key Ignition Value; **FF**: Fitness Function; **CS**: Calibration Stage; **FP**: Fire Prediction; **PFL**: Predicted Fire Line; **RFL_x**: Real Fire Line on time x ; **PV**: input Parameters Vector; **SS_M**: Statistical Stage in monitor process; **pm**: probability map.

Due to this, the use of classical prediction tools does not usually allow to obtain good quality results, being necessary to resort to uncertainty reduction methods that allow to minimize the lack of precision in the system's input. By classical prediction we mean the use of a single instance of the model or simulator which feed each parameter with direct values, whether measured or estimated, but not including pre-processing or calibration data. The Evolutionary Statistical System with Island Model based on Evolutionary Algorithms (ESSIM-EA) [2], is a general uncertainty reduction method that has been successfully applied in

wildfire behavior prediction. ESSIM-EA uses Parallel Evolutionary Algorithms (PEAs) under a multi-population parallelization scheme based on islands [3]. In [2] ESSIM-EA obtained improvements in the quality of prediction in contrast to previously developed methodologies. Subsequently, in [4], ESSIM-EA underwent two complementary studies of static tuning [5] of the evolutionary parameters involved. In the tuning work [4], it was possible to increase the quality of prediction for some instants of the evaluated fires, without being able to determine a single optimum parameters configuration for all the fires. It is important to mention that the quality of prediction, for a particular fire, is calculated in discrete instants of time called prediction steps. Based on these results, this work presents an extension of the ESSIM-EA method where the islands use a heterogeneous configuration of the evolutionary parameters, allowing to obtain, for each prediction step, the result generated by the best parameter configuration. In this work this version is called ESSIM-EAh to differentiate it from the homogenous configuration version.

The Section 2 describes the ESSIM-EA method operation, together with the proposed heterogeneous configuration of evolutionary parameters. Then the details of the experiments are given, the working environment is described, and the results obtained are presented below in Section 3. Finally, in Section 4, the conclusions and the future work are presented.

2 ESSIM-EA: method description

ESSIM-EA has been successfully applied as an uncertainty reduction method in wildfire behavior prediction. As mentioned above, ESSIM-EA uses Parallel Evolutionary Algorithms under an island-based parallelization scheme as the optimization technique with dual master-worker hierarchy. Once the method is running, three different types of processes are executed: a) monitor process, b) master process and c) process worker. The arrangement and communication relationship between these processes can be seen in Fig. 1. The method operation begins in the monitor process, it is responsible for sending the initial information to each island. This initial communication includes: a) the Real Fire Line (RFL_x), b) the Value Ranges (VR) for each input parameter, and c) the PEA Parameter Setting values (PS_{PEA}). It is important to mention that the monitor process will have in memory an array with known parameter configurations, which will be used by the different islands (PS_{PEA}). Since each island can use a different evolutionary parameter configuration, these values (PS_{PEA}) are selected by the monitor and sent to the masters of each island (this is explained in more detail in section: 2.1).

Once each master has received the initial information, it can start the parallel evolutionary algorithm (PEA_M). Each master initiates a different population and applies the evolutionary operators using the parameters configuration previously received from the monitor process. The evaluation of individuals is carried out by workers (OS_{Worker}), which simulate their behavior in the Fire Simulation stage (FS).

Table 1: Parameter settings and ranking. (a) FT: fitness threshold, NIt: number of iterations, NI: number of individuals, NMI: number of migrated individuals, MF: migration frequency, CP: crossing probability and PM: probability of mutation. (b) Ranking of parameter configurations.

(a)							(b)				
Configuration	FT	NIt	NI	NMI	MF	CP	PM	Rank	Case A	Case B	Avg.
No. 0	0,7	200	200	5	5	60	0,5	1°	No. 0	No. 0	No. 0
No. 1	0,78	373	68	3	2	44	0,7	2°	No. 5	No. 9	No. 9
No. 2	0,75	419	194	17	4	37	1,7	3°	No. 18	No. 20	No. 18
No. 3	0,74	198	175	3	7	19	1,1	4°	No. 3	No. 18	No. 3
No. 4	0,71	302	41	11	10	42	0,3	5°	No. 14	No. 3	No. 14
No. 5	0,81	172	25	9	3	22	2	6°	No. 13	No. 11	No. 7
No. 6	0,56	359	125	4	9	27	2,1	7°	No. 7	No. 17	No. 4
No. 7	0,68	129	145	20	7	24	2,2	8°	No. 9	No. 7	No. 20
No. 8	0,69	117	91	5	7	58	1,7	9°	No. 4	No. 2	No. 12
No. 9	0,59	429	187	7	5	30	0,4	10°	No. 16	No. 14	No. 2
No. 10	0,57	144	51	6	4	28	0,6	11°	No. 12	No. 4	No. 5
No. 11	0,65	249	97	19	8	10	0,9				
No. 12	0,73	454	39	9	12	40	0,35				
No. 13	0,84	465	110	12	3	13	1,5				
No. 14	0,62	263	15	18	6	53	1,4				
No. 15	0,64	328	62	16	3	17	1,3				
No. 16	0,8	293	123	13	9	33	0,5				
No. 17	0,67	486	162	15	6	56	1,1				
No. 18	0,82	234	139	9	5	48	2,4				
No. 19	0,77	398	37	8	9	39	1,8				
No. 20	0,161	304	84	14	1	45	2,4				

In the simulation of each individual, the actual fire front of the fire at t_{i-1} (RFL_{i-1}), together with the values of each individual (or vector parameters, VP) in the simulator, is evaluated. At the end of the simulation, the difference between the simulated map and the actual map for that subsequent time instant t_i (RFL_i) is contrasted. The result of this comparison is the fitness value of each individual, which is used by the PEA. Subsequently, the results of the individuals evaluated are sent to their respective master, where the PEA stage is in charge of storing the partial results, evolve the population and migrate the individuals to the neighboring islands. This procedure is repeated until reaching the threshold of fitness or the maximum number of iterations, both previously established. At this time the calibration step begins to operate, i.e., the evolved population is sent to the Statistical Stage (SS) of the master process Calibration Stage (CS_{Master}). SS is responsible for generating a probability map, which is created by considering the simulated maps of each individual. The probability map generated by each island in t_i is used to calculate a pattern of fire behavior, called Key Ignition Value (K_{ign}). This value is calculated in the sub-step search for K_{ign} (SK_{ign}). It is important to note that each island generates in t_i a *key ignition value* and a *probability map*, which are sent to the monitor process ($CS_{Monitor}$). The monitor statistical stage ($SS_{Monitor}$) evaluates the pairs of data received by all islands by selecting those that offer the best fitness value for each prediction step. These data pairs are sent to the Fire Prediction stage (FP) to generate the Predicted Fire Line (PFL).

2.1 Heterogeneous Configuration of Evolutionary Parameters

From the point of view of configuration of evolutionary parameters, an island-based PEA can be implemented using a homogeneous or heterogeneous configuration. Homogeneous configuration is that all islands use the same configuration of evolutionary parameters, i.e. the same values of number of individuals, probability of crossing and mutation, etc., executing exactly the same algorithm with the exception of the populations initialized independently. In contrast, the heterogeneous configuration uses different parameter configurations on each island, whether obtained by known or calibrated configurations, even using random values within preset ranges. The work developed in [6], shows that combining multiple search threads with heterogeneous parameter configurations increases the robustness of the search engine, allowing to explore the space of solutions more efficiently and facilitating the obtaining of better results.

The PEA's capacity depends on certain parameters, such as population size, selection criteria, total number of generations, frequency of migration, among others. Determining the best values for each parameter is usually not an easy task, since there is no universal configuration of parameters and these are usually related to the problem's characteristics. Therefore, in order to obtain a suitable parameter configuration, a calibration or tuning study of the parameters [5] is necessary. Based on this, to determine the parameter configurations to be used, the results obtained in [4] were taken as a starting point, where a static calibration study was carried out on a group of evolutionary parameters of ESSIM-EA. This calibration consisted of evaluating 20 parameter configurations (from No. 1 to No. 20 in Table 1.a) generated through a Latin Hypercube Design [7] (LHD), along with an additional experiment obtained on the basis of a classical parameter configuration (configuration No. 0 in Table 1.a). Each parameter configuration was evaluated in two test cases corresponding to controlled fires. The metric used to perform the ranking was the quality of prediction for each simulation step, the average of each of them and the total execution time. The results allowed to determine those configurations that offered good quality of prediction values, without being able to determine a single configuration, since the results varied by each step of prediction. In order to use the best configurations (in terms of quality of prediction), a ranking was used to reduce the number of LHD configurations and to select those that offered better results.

The ranking was elaborated starting from the results obtained in [4] for each configuration of parameters of the Table 1.a. The classification of the parameter configurations was done for the two case studies (columns Case A and Case B in Table 1.b) and for the average of both (column Avg. in Table 1.b). The generated ranking can be seen in the Table 1.b. The values in the Case A, Case B, and Avg. columns indicate the parameter setting number of the Table 1.a. Therefore, of the 21 initial configurations, we were able to reduce, in a ranking of 11 positions, to a total of 15 final configurations, which have been highlighted in bold in Table 1.a). It is important to note that the ranking configurations of the Table 1.b are those that ESSIM-EA maintains at the moment of starting the execution of the method to be used by each of the islands. The selection of each

Table 2: Description of the case studies: dimension (m), slope (degrees), initial time, increment and final time (min).

Case	Width	Height	Slope	Initial Time	Increment	Final Time
I	89	109	21	2	2	14
II	89	91	21	2.5	2.5	12.5
III	60	90	6	2	2	10

configuration is done randomly in the master process of each island (previously all of the configurations are sent from the monitor process), this selection is carried out at runtime.

3 Experimentation and Results

The performance of the methodology has been evaluated by two metrics that allow us to evaluate important aspects of the behavior of the method: *(a)* the quality of prediction and *(b)* the total time consumed. The quality of prediction is a value calculated using a fitness function. This function compares the real fire map with the predicted map (simulated behavior), determining a numerical value that indicates the accuracy of the prediction. This value, called fitness value indicates the degree of accuracy of this prediction. These values are within the range $\{0,1\}$ where 1 represents a matching prediction with reality and 0 represents a totally wrong prediction. The second metric is to evaluate the total time elapsed for each experiment, this time will depend on the island with the most expensive parameter configuration. At the moment it has not been evaluated how to reduce the idle time in which each island waits for the completion of the rest of the islands, this is part of the future work.

3.1 Experiments

For the experiments, three cases of controlled burns were used in Serra de Louçã, (Gestosa, Portugal).

For each experiment, discrete time intervals have been defined which represent the progress of the fire front where the quality assessment is carried out. Since the terrains have different dimensions, the number of time steps is not the same in all experiments. The characteristics of each of them (i.e., dimension, slope, initial time and duration) are described in Table 2. Due to the non-deterministic behavior of the method, 20 executions of each experiment were performed using the same set of seeds. For each metric the average of them was considered. It is important to note that both ESSIM-EA and ESSIM-EAh cannot perform predictions for the first instant of time (e.g. Experiment I, minute 4), since the calibration step is performed at that time.

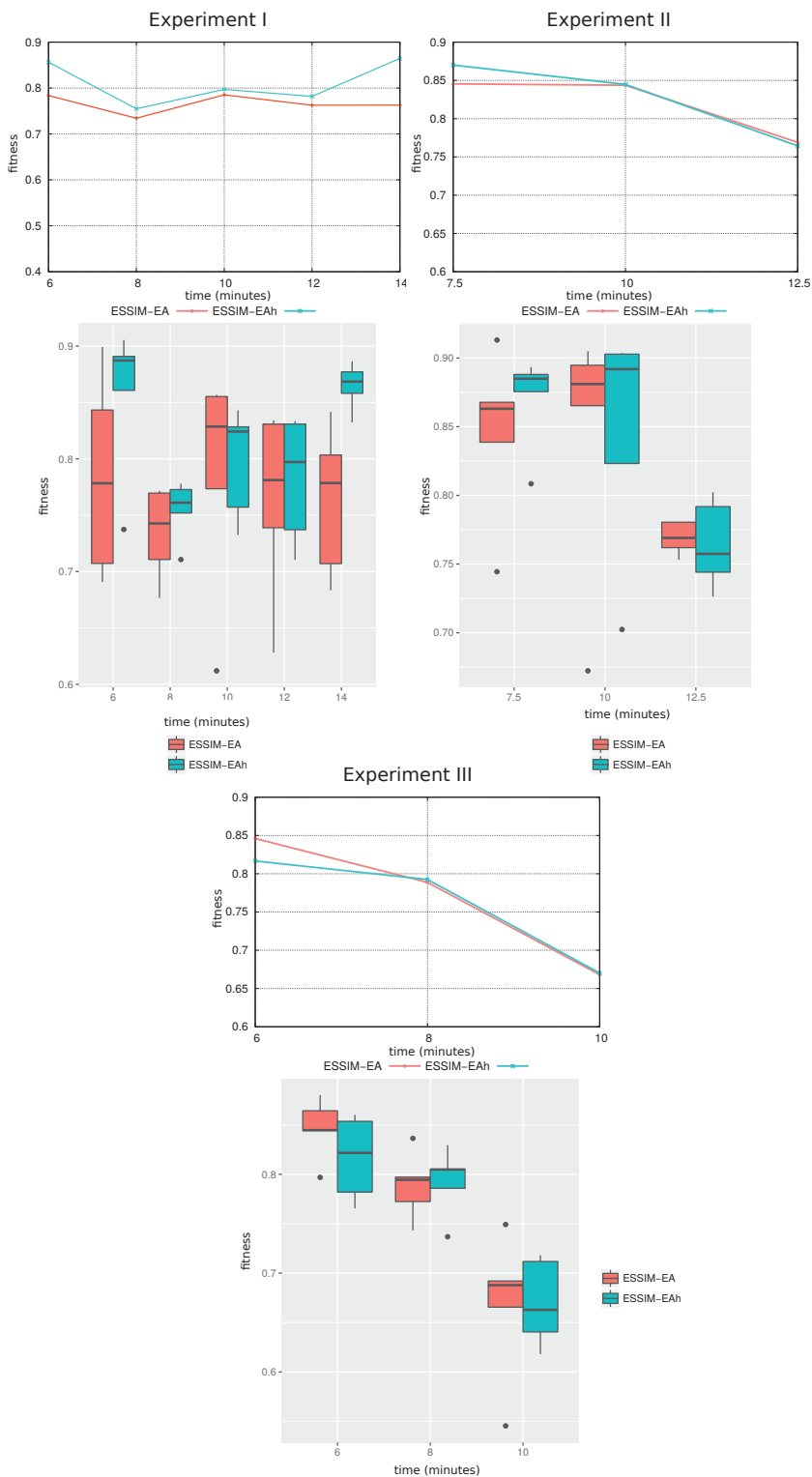


Fig. 2: Experiments I, II and III: Comparison between ESSIM-EA and ESSIM-EAh. Above: Comparison of quality of prediction. Below: Boxplots.

3.2 Results

The results obtained, both in terms of prediction quality and in processing time, have been compared with those obtained by ESSIM-EA using a homogeneous parameter configuration. The first one refers to the original implementation of ESSIM-EA based on a homogeneous configuration of evolutionary parameters, and ESSIM-EAh to the heterogeneous configuration version. It is important to note that the parameter configuration used by ESSIM-EA in these experiments corresponds to the configuration No. 0 of the Table 1.a, while in ESSIM-EAh each island operates using some of the 15 configurations generated by the ranking. In Fig. 2 the quality of prediction has been plotted for the N prediction steps of each experiment. This quality of prediction has been generated from the average of 20 executions. Also, in the lower part of the same figure (Fig. 2), we can see the distribution of these results, using boxplots diagrams. The experiments were carried out in a Linux cluster with 32 processing units (Intel-Q9550 Quad Core 2.83 GHz processors), 4GB RAM, Gigabit Ethernet network under an MPI environment.

Experiment I: this experiment (see Fig. 2), shows that the heterogeneous implementation exceeds the quality of prediction obtained by ESSIM-EA throughout the experiment. The greatest difference is observed in the first and last instants of time, with an average fitness value greater than 0.85. If we analyze the results considering the box diagrams (see Fig. 2), we can observe that in the time instants for which ESSIM-EAh exceeds ESSIM-EA (minutes 6 and 14), the distribution of results obtained by ESSIM-EAh has a smaller deviation. It can also be seen that in the 1 step (minute 6), about 99% of the results is above 0.85. And in the last step (minute 14) just over 75% is on the same value. In the step 2 (minute 8) a similar behavior can be observed, since although the difference in average is smaller, ESSIM-EAh has a minimum deviation concentrating near the 99% of the results in the range: 0.75 and 0.78. Finally, in Table 3 we can observe the amount of time consumed by each method for each experiment. In this table, the executions that offer the best performance have been highlighted in bold, where ESSIM-EA obtains a final difference of 24 less minutes compared to ESSIM-EAh. In this experiment, this difference is mainly related to the value assigned to the parameter that determines the maximum number of iterations (NI_t, in Table 1.a).

Experiment II: in this experiment the largest difference is obtained in favor of ESSIM-EAh in the first prediction step (7.5 minute) with an average value equal to 0.8700898 (Fig. 2). However in the subsequent steps, minute 10 and 12.5, the differences are negligible of the order of 10^{-3} being best ESSIM-EAh. Observing the boxplot, we can see that the performance of ESSIM-EAh exceeds ESSIM-EA. This is so because, at minute 7.5, the 99% of the results are above 0.87, except for an outlier with a value equal to 0.808473. In the rest of the prediction steps (minutes 7.5 and 10), although the quality of prediction average is similar in both implementations, it is observed that the homogeneous configuration offers a more limited distribution of results, indicating a better performance

Table 3: Average execution time.

Experiment	ESSIM-EA	ESSIM-EAh
I	00:57:20	00:81:30
II	01:37:54	00:38:17
III	01:04:13	00:31:40

of ESSIM-EA. The best results obtained correspond to those generated by the parameter configurations number 13, 4, 3 and 7 of Table 1.a.

If we evaluate the amount of total time consumed (Table 3) we can observe that in this experiment, ESSIM-EAh obtains a very significant gain regarding ESSIM-EA, with a difference of 59 minutes. The reduction of time obtained can be related to a greater speed of convergence of the evolutionary algorithm. This could be facilitated by the use of parameter configurations that allow the method to find good solutions with less effort. If this is the case, the maximum amount of iterations would not matter since the fitness threshold would be reached in a smaller number of iterations. To be able to determine this behavior with certainty, we should make a detailed analysis of the evolution of the different populations in each one of the islands.

Experiment III: this has a total area of 5400m². Although it has an equivalent amount of prediction steps to the previous experiment, it has a propagation velocity of the upper fire front, so the total simulation time is 8 minutes. When evaluating the results in terms of the quality of prediction average (Fig. 2) we can observe a similar behavior to the previous experiment, i.e., a significant difference in the first prediction step, in favor of ESSIM-EA, and minimal differences in the following steps. Looking at the box diagram we can see that the performance of ESSIM-EA is more stable than ESSIM-EAh as it concentrates about 99% of results in ranges of the order of 10⁻¹. Finally, the amount of time consumed (Table 3), shows a quite similar behavior to that of the Experiment II. Although here the gain is considerably lower, ESSIM-EAh reduces resource utilization by more than 50%. While the parameter configurations that offer the best results are 13, 9, 0 and 3 from the Table 1.a.

4 Conclusions

In this paper, a new implementation of the Evolutionary Statistical System with Islands Model based on Evolutionary Algorithms (ESSIM-EA) uncertainty reduction method has been presented. The original methodology of the method consists in using an evolutionary algorithm parallelized following an island-based scheme, operating with a unique configuration of evolutionary parameters. This work present the design and evaluation of a new implementation of this method, which consists in equipping ESSIM-EA with the ability to operate each island with different parameter configurations, i.e. with a heterogeneous configuration of evolutionary parameters. The implementation was based on results obtained in previous works, which consisted of studies of static tuning of the evolutionary

parameters. In these works it was possible to determine a series of configurations that allowed to improve the capacity of the method at certain instants of each fire. In this way, the new heterogeneous implementation uses these *known configurations*, and allows each island to select one of them in order to improve the performance of the method both in quality of prediction and in terms of resource utilization. The experimentation was carried out in three cases of controlled fires, evaluating the quality of prediction and the time consumed in each execution. The comparison of results was performed between the heterogeneous implementation (ESSIM-EAh) and the original methodology (ESSIM-EA). The results show that the heterogeneous configuration achieves improvements in both prediction quality and processing time. Although the increase in prediction quality is observed to a greater extent in one of the three experiments, and in the rest this increase is minimal, the new implementation manages to improve the prediction quality in 81% of the predicted steps. In addition to significantly reducing processing time in two of the three cases. As future work will evaluate alternatives to improve the use of resources (avoiding idle resources), will also implement this concept of “heterogeneous configuration of parameters” in a method that uses multiple metaheuristics.

References

1. Infobae, “Los incendios afectan más de un millón de hectáreas en La Pampa, Río Negro y Buenos Aires”, <http://www.infobae.com/sociedad/2017/01/05/sergio-bergman-los-incendios-afectan-mas-de-un-millon-de-hectareas/>. 2017.
2. M. Méndez-Garabetti, G. Bianchini, P. Caymes-Scutari, M. L. Tardivo, “Increase in the quality of the prediction of a computational wildfire behavior method through the improvement of the internal metaheuristic”, *Fire Safety Journal*, Volume 82, ISSN 0379-7112, pp. 49-62, May. 2016.
3. H. Mühlenbein, *Parallel genetic algorithms, population genetics and combinatorial optimization*, pages 398–406. Springer Berlin Heidelberg, 1991. doi:10.1007/3-540-55027-5_23.
4. M. Méndez Garabetti, G. Bianchini, P. Caymes Scutari, M. L. Tardivo, ESS-IM applied to Forest Fire Spread Prediction: Parameters Tuning for a Heterogeneous Configuration. 2016 35th International Conference of the Chilean Computer Science Society (SCCC), ISBN: 978-1-5090-3338-6 (USB), pp. 514-525
5. E. Talbi, *Metaheuristics: From Design to Implementation*, John Wiley and Sons, 2009.
6. E. Alba, F. Luna, A.J. Nebro, and J.M. Troya. Parallel heterogeneous genetic algorithms for continuous optimization. *Parallel Computing*, 30(5- 6):699–719, 2004.
7. R. L. Iman y W.J. Conover, “Small sample sensitivity analysis techniques for computer models, with an application to risk assessment,” *Communications in Statistics, Part A. Theory and Methods*, Vol. 17, 1980, pp. 1749-1842.
8. R. Real, & J.M. Vargas. (1996). The Probabilistic Basis of Jaccard’s Index of Similarity. *Systematic Biology*, 45(3), 380–385. <http://doi.org/10.1093/sysbio/45.3.380>

Fuzzy Harmonic Systems for Traffic Risk

Walter Bel¹, Daniela López de Luise¹,

¹ Universidad Autónoma de Entre Ríos (UADER),
Concepción del Uruguay, Entre Ríos, Argentina
bel.walter@uader.edu.ar, daniela_ldl@ieee.com

Abstract. This paper aims to present a model for traffic risk prediction. Its contribution is the adaptation of Fuzzy Logic applied to Harmonic Systems in order to make it more flexible and powerful in certain contexts. The possibility of having a good traffic risk prediction opens a practical possibility to successfully improve the security not only for drivers, but also for pedestrians and cyclists. The proposed model is able to process in real-time with simple data provided by the environment and the individual whose risk is being processed. The scope of this paper covers the technical description of the model, statistical analysis and comparison with alternates using Traditional Ruled-Expert Systems (RES), Harmonic Systems (HS) and Fuzzy Harmonic Systems (FHS). Also a short proposal for the prototype is described. Results indicate a remarkable improvement for the FHS predictor compared to RES and HS.

Keywords: Fuzzy Harmonic Systems, Harmonic Systems, Expert Systems, Traffic Risk Prediction

1 Introduction

With the development of the digital technology, new data sources are capable of generating great diversity and quantity of data. In accordance to the International Data Corporation (IDC) today data are increased by 50% per year [1]. These characteristics, when it comes to important numbers of records, they become unmanageable for traditional systems. It is thus that the techniques of big data and deep learning take place in these circumstances, generating the possibility of enhancing less procedural algorithms in pursuit of modern heuristics. All this is at the cost of a process of increasing the bias of the model, but with the convenience of becoming manageable problematic situations so far difficult until now difficult to solve.

The number of drivers on the roads and streets has increased dramatically over the years. In accordance to the Wards Auto [2] the number of vehicles in circulation in the world went from 980 million in 2009 to 1015 million vehicles in 2010, representing a 3,57% in just one year. The traffic generated as a consequence causes congestions and complications that increase over time, in the detriment of not only of

the drivers themselves but also of the rest of the cohabiting community. One of the most compelling problems in the world are traffic accidents and their consequences [3], [4] involving material losses, people injured and even multiple deaths [5], [6]. An effective risk assessment should include information of the environment, proper context information, vulnerability of the individual, sudden biomechanical forces [7], among other variables. There are many studies and statistics with models that partially represent the apparent risk, but they are not accurate predictors because they miss a representative number of hidden factors, which are indirect or partial cause. Examples of such uncovered tips are: the fact that pedestrian injuries increase with vehicle speed, sidewalk status, availability of suitable crossing facilities, lack of proper pedestrian crossings, number of lanes to cross, complexity of intersections, etc.

Among the items that are directly involved there are some that are typically considered, such as the use of seatbelt and airbag in vehicles. However there are other extra facts that relate to the impossibility of preventing risk situations (for example: sudden movements from other vehicles or pedestrians, animals crossing the sidewalk, etc.). The reaction time in such cases is often insufficient for appropriate risk nullification. On average, a person reacts in 1.5 seconds to a danger stimulus. This should be added to other facts that can slow down or alter this time making things worse. Some proposals try to generate some extra time, so the driver can limit the risk of accident, one example is the VANETS navigation platform [8].

There are proposals that are based on the use of certain technologies. By case, designing networks that can act in different environments (streets and routes) [9]. There are also works on communication between vehicles (cell phone, Bluetooth, RF, infrared, etc.) [10], [11], [12], [13], [14] as an element of additional support and prevention. In these proposals the vehicle communicates with two others (front and behind) while driving in its lane. But their success is limited and they are very sensitive to the weather conditions. Other proposals present the possibility of being coupled to the software of a GPS navigator or cell phone devices of personal use.

From Data Mining (DM) perspective, risk can be thought as a derivation of a set of variables heuristically selected as the most likely cause of some described accident. The HS are a type of mining centered on the rhythm, accelerations, static periods and other aspects related to the time characteristics of the selected patterns. HS also allows real-time processing. It is prepared for applications that require rapid reactions to data collected with certain criteria (for example a driver or pedestrian who compiles environmental data during its movement from one point to another) [15].

The FHS model extends HS with fuzzy predictions, enabling the semantic conception of the variables, thus allowing a better focusing of the predictor behavior. Instead of recognizing patterns configuration of certain parameters deterministically, it analysis just timings behavior at intervals. Taking that step in inferential capacity, it increases the bias of this model. This model changes the pattern of resonance parameters as processed by HS incorporating a fuzzy configuration. This way, according to the Occam Razzor theorem [16], a higher pattern detection power is expected in new situations. The rest of this paper presents the Fuzzy Harmonic System model (section 2), a statistical analysis and comparison of ES, HS and FHS

models (section 3), the proposal of prototype FHS implementation (section 4) and finally conclusions and future works (section 5).

2 Fuzzy Harmonic Systems Model

This section presents a technical description of the FHS [17].

FHS Goal. FHS systems can be applied to cases where the problem to be treated has two characteristics: response requirements in real time and the data are not discrete, also requiring a complex context of such data. Through the use of fuzzy patterns it is possible to detect implicit knowledge in the variables defining membership functions (for example low visibility, high speed, sunset and dawn hours, etc.), and the use of a word bag implementation for categorical attributes (for example `bag_rain = {Rain, Drizzle, Hail, Thunderstorm, Precipitation, Overcast, Squalls}`). Fuzzy patterns can be parameterized deterministically, by set of values or by intervals. This way it is possible to use FHS in problems whose data need to be interpreted with semantic sense, before being converted and evaluated as numbers.

FHS Problems. FHS can be used in the same types of problems as HS, provided that specific patterns of interest are properly defined (for example for problems like intrusion detection in a network, production failures, Decision Support Systems, process control, etc.). Additionally, it is possible to react upon changes of behavior in those patterns. It is important to note that the focus of the system emphasizes the time of occurrence of events, not the event itself.

FHS Functioning. FHS extend the HS model approach. A deeper description of HS is out of the scope of this paper but readers interested in details may find them in [18], and for more details about HS applied to traffic risk in [19], [20], [21], [22] y [23]. The main feature that is being added at the present model is the use of fuzzy patterns and certain basic expert rules as shown in fig. 1, providing the system with the ability to contextualize variables, and incorporating commonly implicit knowledge of the context in the form of membership functions. Through a set of specifically designed operators, it is possible to manipulate logical and subjective concepts of the problem through numbers. The result of the whole process is a weighted collection of non-numeric parameters, enriched with the contribution of subjective information not considered by other approaches.

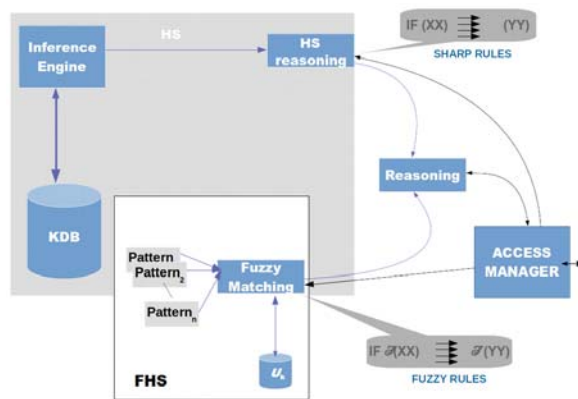


Fig. 1. FHS Model.

3 Statistical Analysis

Statistical analysis compares performance of a prototype implementing ES, HS (presented in [17], [22]) and a functional prototype that partially implements the FHS core. In order to be able to perform the comparison of the different models it is mandatory to test same cases.

-Dataset, is a collection of real world traffic at “Concepción del Uruguay” city (Entre Ríos province), these data are in the public domain (can be accessed from the link: https://drive.google.com/file/d/0B97BXscx2_9mSnpsZnB3Qm0wQnM/view?usp=sharing), fig. 2 shows a map of the city and its risk zones.

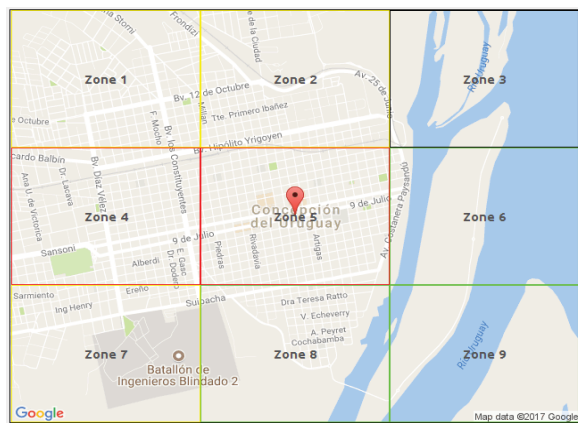


Fig. 2. Map of risk zones for Concepción del Uruguay.

As can be seen, the map has a mix of risks. Each cell represents a zone, described in the database as a tuple IDzone, two points of latitude and longitude, and risk. IDzone (identification code for the cell), its geolocation points and the level of risk associated with it (values can be HIGH, MEDIUM, or LOW). The map represents risk levels with green (low), yellow (medium), red (high).

-Patterns and rules, table 1 lists the 14 patterns used for testing HS and FHS systems. ES rules are the same as the ones described for HS patterns.

Table 1. Description of driver and pedestrian patterns for HS and FHS models.

ID	Patterns HS	Patterns FHS
Driver Patterns		
1	Type = Bike, Alcohol ≥ 0.2 Color = black, Belt/Helmet = false Time > Sunset	type = Bike, alcohol1 = high color = risk, belt_helmet = false time = nightfall
2	Type = Motorcycle, Alcohol ≥ 0.2 Weather = Rain	type = Motorcycle, alcohol1 = high weather = rain
3	Type = Motorcycle, Alcohol ≥ 0.2 Weather = Rain, Color = black	type = Motorcycle, alcohol1 = high weather = rain, color = risk
4	Type = Motorcycle, Belt/Helmet = false, Speed > 40, Time > Sunset	type = Motorcycle, belt_helmet = false speed1 = medium, time = nightfall
5	Alcohol ≥ 0.45 , Color = black Time > Sunset	alcohol = high, color = risk time = nightfall
6	Gender = Male, 22 \geq Age \geq 18 Day = Saturday, 12:00:00 > Time > 00:00:00	gender = Male, age = young day = Saturday time1 = dawn-md
7	Gender = Male, 22 \geq Age \geq 18 Day = Friday 23:59:59 > Time > 18:00:00	gender = Male, age = young day = Friday time = nightfall
8	Age ≥ 55 , Time > 18:00:00 Time < 23:59:59	age = old, time = nightfall
9	Visibility ≤ 4 , Color = black Time > Sunset	visibility = little, color = risk time = nightfall
10	Visibility ≤ 4 , Color = black Time < Sunrise	visibility = little, color = risk time = dawn
11	Type = Big Truck, Visibility ≤ 4 Speed > 80	type = big, visibility = little speed = high
Pedestrian Patterns		
12	Weather = Rain, HeadPhone = true Time > 18:00:00	weather = rain, headphone = true time = nightfall
13	Alcohol > 0.45, Time > Sunset	alcohol = high, time = nightfall
14	Alcohol > 0.3, Time < Sunrise Age ≥ 50 , Visibility ≤ 4 Temperature ≤ 10	alcohol1 = high, time = dawn age1 = adult-old, visibility = little temperature = low

For more details about fuzzy sets used for FHS patterns and fuzzification process see [17]. Table 2 shows the initial configuration values used for HS and FHS models.

Table 2. Initial configuration of HS and FSH patterns.

Parameters	Value
u - threshold	0.03
n+ - number of system resonance cases	0
nt - total cases treated by the pattern	0
nu - user threshold inertia coefficient	0.3
n - learning coefficient	0.05
nc - distribution threshold	80
li - average time of stamp i of the pattern	0
ti - time of last activation of the pattern	0

-Results, Table 3 presents a summary of the results of the testing process for the three prediction models. The number of cases studied, instances of risk, predicted risk level (risk values can be 0=NO RISK, 1=LOW RISK, 2=MEDIUM RISK, 3 = HIGH RISK), and the cases where the zone's risk level influences the risk level calculated by the model [23].

Table 3. Summary of systems testing.

Number of cases	Description
516488	Total testing cases
15440	Cases of risks detected
3036	Cases in which the risk of the zone influence in the risk defined

Table 4 presents a detailed description of the results obtained from testing, grouped according to the main characteristics of the behavior presented by ES, HS and FHS models. It can be observed similarities and differences of the risk obtained, the number of cases corresponding to each group and subgroup, and some remarks.

Table 4. Detailed description of test results.

ES	HS	FHS	Amount	Sub-Amount	Remarks
0	0	0	501048	501048	None of the three models detect a risk situation in the case.
1	1	1	309	2018	The three models return the same level of risk to the case.
2	2	2	890		
3	3	3	819		
0	0	1	1217	7847	FHS detects risk situations of different levels. HS and ES do not detect risk for the case. This difference is due to the flexibility of fuzzy patterns.
0	0	2	4035		
0	0	3	2595		
0	1	1	4	699	HS and FHS detect the same level of risk and ES does not detect a risk situation for the case. This is because both models focus on time and not on the event.
0	2	2	398		
0	3	3	297		

0	1	0	21	35	The HS detects risk situations of different levels. FHS and ES do not detect risk for the case. FHS does not detect them due to the demanding threshold of acceptance (only resonates in cases of marked risk).
0	2	0	11		
0	3	0	3		
1	2	2	703	1868	The difference in level is due to the fact that both HS and FHS add to the level of risk the risk inference of the zone.
2	3	3	1165		
1	0	0	65	710	The ES detects a risk situation of different levels. HS and FHS do not detect a risk situation. This is due to the acceptance threshold.
2	0	0	177		
3	0	0	468		
1	0	2	17	60	FHS and ES detect a risk situation but HS does not (this is due to the acceptance threshold), the difference in level is due to the inference of the risk of the zone.
2	0	3	43		
1	2	0	324	1108	HS and ES detect a risk situation but FHS does not (this is due to the acceptance threshold), the level difference is due to the inference of the risk of the zone.
2	3	0	784		
1	0	1	11	70	FHS and ES detect a risk situation but HS does not (this is due to the acceptance threshold).
2	0	2	51		
3	0	3	8		
1	1	0	113	1025	HS and ES detect risk situation but FHS does not (this is due to the acceptance threshold).
2	2	0	766		
3	3	0	146		

From the results, a random and significant sample is extracted to analyze the relation between the prediction results of the models. Contingency tables 5 and 6 show an interesting difference between the different predictions.

Table 5. Contingency table es_predict * hs_predict.

		HS_predict				Total
		0	1	2	3	
es_predict	0	6623	0	209	171	7003
	1	48	219	560	0	827
	2	147	0	777	931	1855
	3	246	0	0	70	316
Total		7064	219	1546	1172	10001

Table 6. Contingency table es_predict * FHS_predict.

		FHS_predict				Total
		0	1	2	3	
es_predict	0	2505	675	2405	1418	7003
	1	259	167	401	0	827
	2	925	0	390	540	1855
	3	316	0	0	0	316
Total		4005	842	3196	1958	10001

4 Proposal Prototype FHS

As mentioned in previous sections, the Kronos prototype partially implements the FHS model. Since the goal here is to validate its improvement of efficiency with respect to the HS model, it is not required a full implementation. The prototype is coded Java for its simplicity and flexibility to migrate to Android with minimal modifications. The global design of the system allows to interact via the Internet (with off-line support for cases where there is no Internet service). There is also the possibility to use it with several devices and alternate information systems (see fig. 3).

**Fig. 3.** Architecture of the prototype Kronos Mobile.

Kronos Admin (1). Management of patterns, maps and zones of risk, in accordance to the functionality previously described [17], [23].

Firestore (2). Real-time database for persistence of system data and data generated by it, with support for authentication and user management. It can be used also off-line.

Kronos Mobile (3). Mobile prototype that implements several functionalities.

- a. **user profile and vehicles**, allows the configuration and profiling of users and their vehicles.
- b. **blood alcohol model**, based on the Widmark formula [24] using UBE (Standard Drinking Unit). It helps to determine alcohol level in blood, and infers its decreasing behavior after time t .
- c. **weather model (5)**, based on the descriptions and needs determined in [17]. It is updated by means of a difference function after a time t , if the weather data present a significant difference from those obtained previously.

- d. **FHS model**, to calculate the level of risk of the user and issue an alert in case of being exposed to a possible risk situation.
- e. **speed calculation model (4)**, based on GPS geolocation, obtains the speed of the user's movement using the Haversine formula [25].

5 Conclusion and Future Work

This paper presents the FHS model, with a prototype that implements it. It is presented also a statistical comparison using ES, HS and a partial implementation of FHS. The characteristics of each model were presented: rules, patterns and basic semantic information about the zones used to infer the risk levels. The comparison between the different models shows interesting differences in the number of cases detected by FHS on HS and ES. This is due to the flexibility granted by fuzzy patterns and the ability to contextualize the variables and obtain subtle information.

As future work remains the full implementation of the FHS model and to perform real-time evaluations in different geographical locations of the country in order to measure the robustness of the model behavior in different situations.

References

1. International Data Corporation. <https://www.idc.com/> (2017)
2. Wards Auto. <http://wardsauto.com/> (2017)
3. D. López De Luise, MLW and bilingualism. Adv. Research and Trends in New Tech., Software, Human-Computer Interaction, and Communicability. IGI Global. USA. (2013)
4. O. Wash. Assessing pedestrian risk locations: a case study of WSDOT efforts. Department of Transportation. Washington State Library. Electronic State Publications. (1998)
5. Org. for Economic Co-operation and Development (OECD). <http://www.oecd.org>. (2014)
6. D. Alex Quistberg, J. Jaime Miranda, Beth Ebel. Reducing pedestrian deaths and injuries due to road traffic injuries in Peru. Rev. Peruana de Medicina Experimental y Salud Pública. Rev Peru Med Exp Salud Pública vol. 27 n. 2 Lima Apr/Jun. ISSN 1726-4634. (2010)
7. J. Oxley. Improving Pedestrian Safety. Curtin - Monash Accident Research Centre. Fact Sheet No. 6. (2004)
8. VANETs – A platform for the future Intelligent Transport System (ITS). Ms. Dahlia Sam. CSE (2014)
9. Vehicular ad hoc networks (VANETs): status, results, and challenges." Telecommunication Systems - Zeadally, Hunt, et al. (2010)
10. H. Fujii, K. Seki, and N. Nakagata, "Experimental research on protocol of inter-vehicle communication for vehicle control and driver support," 2nd World Congress on Intelligent Transport System. 1995, Yokohama, Japan, Nov. 9–11, pp. 1600–1605. (1995)
11. I. Sasaki, T. Hirayama, and T. Hatsuda, "Vehicle information network based on inter-vehicle communication by laser beam injection and retro reflection techniques," IEEE Vehicle Navigation and Information System Conference. Japan, pp. 165–169. (1994)

12. K. Mizui, M. Uchida, and M. Nakagawa, "Vehicle to vehicle communication and ranging system using spread spectrum technique" IEEE Vehicle Navigation and Information Syst. Conf. '94, Yokohama, Japan, Aug. 31–Sept. 2, pp. 153– 158. (1994)
13. D.W.Kremer, D. Hubner, S. Holf, T. Benz, and W. Schafer,"Computer aided design and evaluation of mobile radio local area networks in RTI/IVHS environments," IEEE J. Select. Areas Communication, vol 11, April. (1993)
14. J.-M. Valade, "Vehicle to vehicle communications: Experimental results and implementation perspectives,"2nd World Congress on Intelligent Transport Syst. '95, Yokohama, Japan, Nov. 9–11, pp. 1606–1613. (1995)
15. Daniela López De Luise, Walter Bel - "Cálculo de Riesgo en Tráfico y Peatón usando Sistemas Armónicos" - 2017, 978-3-639-53739-0. Editorial Académica Española. (2017)
16. Data Mining and Knowledge Discovery, Domingos, P. Data Mining and Knowledge Discovery 3: 409. doi:10.1023/A:1009868929893. (1999)
17. Parametric Prediction Model using Expert System and Fuzzy Harmonic System. W. Bel 1 , D. López De Luise 2 , A. Asmis 3 , D. Mansilla 4. SOFA. (2016)
18. D. López De Luise. Harmonics Systems for Time Mining. International Journal of Modern Engineering Research (IJMER). Vol. 3, Issue. 6. 3 pp-2719-2727. ISSN: 2249-6645. (2013)
19. A. Celayeta, C. Paredes, D. López De Luise, W. Bel. Traffic and Pedestrian risk evaluation with Harmonic Systems.ARGENCON. (2014)
20. D. Mansilla, D. López De Luise, W. Bel. Un Modelo con Conocimiento Experto y Sistemas Armónicos para Evaluación de Riesgos de Tráfico. EnIDI Argentina. (2015)
21. I. Acuña, E. García, D. López De Luise, C. Paredes, A. Celayeta, M.Sandillú, W. Bel. Traffic & Pedestrian risk inference using Harmonic Systems. SOFA. Romania. (2014)
22. Risk Prediction on Time with GPS Information . Daniela López de Luise, Walter Bel, Diego Mansilla, Rigoberto Malca la Rosa. "Information Technology and Computational Physics", P. Kulczycki, L.T. Kóczy, R. Mesiar, J. Kacprzyk (Eds.), Springer (2017)
23. Predicción de Riesgo basado en tiempo y patrones GPS. Daniela López de Luise, Walter Bel, Diego Mansilla, Rigoberto Malca la Rosa. ARGENCON. (2016)
24. Widmark - "Principles and Applications of Medicolegal Alcohol Determination" - Biomedical Publications - 978-0931890079. (1981)
25. Prof. Nitin R.Chopde, Mr. Mangesh K. Nichat - "Landmark Based Shortest Path Detection by Using A* and Haversine Formula" - IJIRCCCE Vol. 1, Issue 2, April - 2320–980. (2013)

Hacia un Formalismo Basado en Sistemas Multiagente para Evaluar Modelos de Organización Humanos. Caso de Estudio (I): *Holacracy*

Alexandre Ribas Hortal¹, Facundo Bromberg

Laboratorio DHARMA, Universidad Tecnológica Nacional,
Facultad Regional Mendoza. Rodriguez 273,
5500, Mendoza, Argentina
kozzak@gmail.com

Abstract. Se presenta una metodología en curso para evaluar modelos de organización humanos en base al concepto de *inteligencia*, entendida como la habilidad de una entidad para decidir las acciones que la llevan a mejorar su rendimiento dado cierto entorno. En los últimos años han surgido un gran número de modelos de liderazgo distribuido que emergen de manera local y con descripciones ad-hoc de su funcionamiento, razón por la cual no existe un lenguaje general que permita evaluarlos comparándolos entre sí. Este trabajo muestra como describir a las organizaciones humanas en términos de sistemas inteligentes permite compararlas con un lenguaje común y detectar características de sus mecanismos de organización que afectan a su rendimiento. A través de un caso de estudio, se ilustra como el modelo Holacrático permite aprovechar mejor los recursos disponibles para mejorar el rendimiento con respecto a un modelo de organización tradicional top-down.

Keywords: Liderazgo distribuido, Holacracy, Sistemas inteligentes, Planificación multiagente bajo incertidumbre, Dec-POMDPs

1 Introducción

Este trabajo surge del interés por utilizar conocimientos del área de la Inteligencia Artificial para profundizar en la comprensión de la cooperación humana, por lo que se parte de una perspectiva tecnológica para comprender a las organizaciones como una tecnología para resolver problemas que serían imposibles de resolver individualmente [6]. Así, el objetivo principal de este trabajo es proponer una metodología que brinde un marco contextual integrador basada en los formalismos de los sistemas multiagente para describir modelos de organización humanos en términos de su grado de inteligencia, entendiendo la inteligencia de una organización como su habilidad para maximizar una medida de rendimiento dado un entorno de trabajo [13]. La motivación principal para desarrollar esta metodología es que no existe un formalismo genérico en la literatura para comparar entre sí los diferentes modelos de liderazgo distribuido que aparecen con la emergencia de artefactos socio-tecnológicos inspirados por los valores de tradición cooperativista, los modelos *open source* y los principios de la auto-organización como las *Open Value Networks* [14], las *Hybrid Organizations* [1], o las *Decentralized Autonomous Organizations* [3]. Por lo que describir a las organizaciones humanas en términos de inteligencia permite analizarlas con un lenguaje común que admite comparativas respecto a cómo los mecanismos de organización utilizados afectan a su rendimiento.

Este primer paso hacia una metodología se desarrolla a través de un caso de estudio donde se compara bajo un mismo entorno de trabajo, el mecanismo Holacrático [12] con un mecanismo tradicional de gestión top-down, mostrando como el mecanismo Holacrático permite aprovechar mejor la información disponible en la toma de decisiones respecto a cómo actuar para maximizar su medida de rendimiento, por lo que en consecuencia muestra un grado de inteligencia superior a los mecanismos de gestión tradicional top-down. Sin embargo, también se muestra cómo la implementación de un mecanismo Holacrático tiene un costo computacional extra requerido para la toma de decisiones respecto a maneras óptimas de distribuir el liderazgo, lo cual puede ser demasiado costoso para organizaciones humanas con características particulares. Finalmente se discute como esta problemática podría superarse si los modelos de organización humanos contemplaran los costos de implementación tal y como hacen los modelos artificiales de planificación multiagente bajo incertidumbre [2, 5].

El resto del artículo se organiza de la manera siguiente: En la Sección (2) se presentan los formalismos que permiten describir a una organización como un agente inteligente, focalizando en la planificación multiagente bajo incertidumbre. En la Sección (3) se describen los principios básicos del modelo Holacrático y el funcionamiento de su proceso de toma de decisiones. En la Sección (4), se presenta la comparativa entre un modelo Holacrático y un modelo tradicional de gestión top-down en términos de sistemas inteligentes. En la Sección (5) se presenta una discusión y conclusiones finales y en la Sección (6) se recogen las referencias.

2 Planificación bajo Incertidumbre para Entornos Multiagente

En esta sección se describen los formalismos de sistemas multiagente (SMA) e inteligencia artificial distribuida (IAD) que brindan el marco teórico para evaluar el grado de inteligencia de una organización en base al modelo bajo el que están operando. En este sentido, las organizaciones como objeto de estudio, son de interés para los investigadores del área de los sistemas multiagente cooperativos, puesto que se ha demostrado que el diseño organizacional de un sistema de agentes cooperativo puede tener un impacto significativo en el rendimiento colectivo alcanzado [7]. Para profundizar en esta cuestión, se describen los formalismos para definir la inteligencia de un sistema de agentes racionales [13], y posteriormente se desarrolla el problema central del área de sistemas multiagente cooperativos: ¿cómo hacer que los agentes que conforman un sistema cooperen de manera efectiva? Adoptando uno de los enfoques más populares para la resolución del problema de planificación en sistemas multiagente cooperativos bajo condiciones de incertidumbre, los Dec-POMDPs (*Decentralized Partially Observable Markov Decision Processes*) [11].

2.1 Agentes Racionales como Modelos de Inteligencia

Desde la perspectiva de agentes racionales, la inteligencia se define como la habilidad de un agente situado en un entorno para elegir las acciones que mejoren su performance respecto a una medida de rendimiento dada. Russell y Norvig [13] describen el entorno de trabajo de un agente con la tupla $\langle R, E, A, S \rangle$, donde:

- R: es la medida de rendimiento.
- E: es el entorno externo.
- A: son los actuadores.
- S: son los sensores.

En relación al entorno externo, se contemplan diferentes características que permiten establecer la complejidad del problema dependiendo de las limitaciones perceptuales, actuadoras y habilidades cognitivas de los agentes. Las dimensiones características de los entornos recogidas por Russell y Norvig son: (1) Estático vs. dinámico, (2) Discreto vs. continuo, (3) Determinista vs. estocástico, (4) Totalmente observable vs. parcialmente observable, (5) Episódico vs. secuencial, (6) Simple agente vs. cooperativo vs. competitivo, con lo últimos de cada tupla como la alternativa más compleja. En el caso de las organizaciones humanas, éstas deben entenderse como agentes que operan en el entorno más complejo, puesto que operan en el mundo real, es decir en un entorno: dinámico (el mundo externo a la organización cambia a pesar de las acciones de la organización), continuo, estocástico, parcialmente observable (la organización no puede percibir todos los aspectos relevantes del entorno externo), secuencial (la mejor acción ahora depende de las acciones que se ejecutarán más adelante), y multiagente.

De modo que asumiendo esta formalización, tal y como recogen Horling & Lesser [7], se acepta que no hay un modelo de organización que sea el más inteligente para todas las situaciones, sino que dado un entorno de trabajo y una organización con unas características determinadas, existe un diseño ideal particular, que es el que permite aprovechar mejor los recursos disponibles para maximizar su medida de rendimiento.

2.2 Mecanismos de Organización Humanos como Dec-POMDP

Una organización humana se puede describir como un sistema de agentes racionales que coordinan sus acciones para alcanzar un propósito común explícito (i.e. conocido por todos), donde la medida de rendimiento de los agentes individuales está alineada con la medida de rendimiento de la organización a través de una recompensa única recibida por la organización.

Esta descripción coincide con la formalización de un SMA cooperativo, que es aquél donde los agentes actúan conjuntamente para alcanzar un objetivo compartido. El problema central de esta tipología de SMA es conseguir que los agentes cooperen de manera efectiva [10] y un enfoque para lograrlo es la planificación multiagente bajo incertidumbre, que se puede formalizar a través de Dec-POMDPs (Decentralized Partially Observable Markov Decision Processes). Este formalismo describe la problemática central de un sistema multiagente cooperativo actuando bajo condiciones de incertidumbre, donde los agentes individuales no pueden percibir el estado completo del entorno, aunque sí pueden hacerlo de manera conjunta. Una problemática similar a la que se enfrentan las organizaciones humanas, donde los individuos cooperan para alcanzar un objetivo común (i.e. el propósito de la organización), bajo condiciones de incertidumbre, y donde entre todos pueden observar el estado del entorno (procesos internos y estados externos de la organización). La complejidad de estos entornos de trabajo se concentra en la dificultad computacional de coordinar las acciones de los agentes para que actúen de la manera más inteligente posible, basándose en su percepción local y entendimiento individual del entorno, evitando conflictos con las acciones de otros agentes. Hasta donde se ha podido corroborar, los Dec-POMDPs son los modelos de SMA cooperativos que la comunidad considera más adecuados para modelar organizaciones realistas como redes de comunicaciones, redes de sensores, sistemas cooperativos de robots o sistemas de soporte a decisiones, puesto que permiten incorporar diferentes condiciones de incertidumbre en los procesos de razonamiento de manera cuantitativa [11]. A continuación, se describen sus características básicas.

2.2.1. Dec-POMDPS

Según Oliehoek & Amato [11], un Dec-POMDP es un formalismo que generaliza el modelo de planificación bajo condiciones de incertidumbre y parcialmente observable para agentes individuales llamado POMDP por su nombre en inglés (Partially Observable Markov Decision Processes), que a su vez es una generalización de un MDP (Markov Decision Process), un modelo de planificación para entornos estocásticos totalmente observables. La descripción formal de un MDP se ajusta mucho a la definición de un agente inteligente ofrecida por Russel & Norvig [13] puesto que presenta una interfaz de acciones y observaciones disponibles para interactuar con el entorno que se define con la tupla $\langle S, A, T, R, H \rangle$, dónde:

- S: el conjunto finito de estado del entorno
- A: el conjunto finito de acciones del agente
- T: la función de transición probabilística $T(s, a, s')$, donde $s, s' \in S$ y $a \in A$,
- R: la función de recompensa (representa la medida de rendimiento del agente)
- H: el horizonte del problema (durante el cual el agente interactúa con su entorno)

Así un MDP define el problema de un agente individual que debe aprender a actuar en un entorno estático, estocástico, secuencial, discreto y totalmente observable, en pos de maximizar su rendimiento o utilidad, que se formaliza como la maximización de las recompensas obtenidas sobre secuencias esperadas de estados alcanzables desde el estado actual. La función de recompensa inmediata $R: S \times A \rightarrow \mathbb{R}$ mapea estados con acciones conjuntas y se utiliza para especificar el objetivo de los agentes. R sólo especifica la recompensa inmediata de una acción conjunta, y puesto que el objetivo de estos problemas de toma de decisiones secuenciales es optimizar un comportamiento a largo plazo, es necesario seleccionar un criterio de optimalidad en orden de especificar el problema (i.e. definir exactamente qué se quiere optimizar). En los problemas de horizonte finito, en general una secuencia de acciones conjuntas deseable se corresponde con una recompensa a largo plazo formalizada como retorno (el modelo de recompensa más usado para horizontes finitos es el de *expected cumulative rewards*), en el caso de los problemas de horizonte infinito se puede definir una sumatoria finita mediante el uso de *discounted cumulative rewards* que es el modelo más avalado por la comunidad.

La solución al problema planteado por un MDP es una política, que determina la acción a tomar en cada estado del entorno, por lo que alcanzar la maximización del rendimiento se corresponde en un MDP a encontrar la política óptima, (i.e. saber para cada estado la acción que garantiza una secuencia de estados, y con ello de recompensas, que en el largo plazo maximice la utilidad recibida).

La extensión para entornos parcialmente observables (POMDP) incorpora una función de probabilidad que dado el historial de observaciones que el agente recolectó en sus visitas a una secuencia de estados, arroja la creencia de que el entorno se encuentra actualmente en cierto estado. Así, a diferencia del caso totalmente observable, el agente recurre a la distribución de creencias respecto a cada estado para decidir cuál es la acción más inteligente (i.e. la de mayor rendimiento esperado a largo plazo).

Por último, los Dec-POMDPS extienden el modelo de POMDP a un entorno multiagente en base a la tupla $MDecP = (D, S, A, T, O, R, H)$ formalizada de manera completa en Oliehoek & Amato [11], dónde:

- D= $\{1, \dots, n\}$ es el conjunto de n agentes.
- S: es el conjunto (finito) de estados.
- A: es el conjunto de acciones conjuntas.
- T: es la función de transición probabilística.

O: es el conjunto de observaciones conjuntas.
O: es la función de observación probabilística.
R: es la función de recompensa inmediata.
H: es el horizonte del problema.

Por lo que el modelo más general de Dec-POMDP asume un entorno estático, estocástico, secuencial, discreto, multiagente cooperativo, y parcialmente observable, donde dado un estado inicial, cada uno de los agentes ejecuta una acción, y cuya acción conjunta resultante influye en el entorno produciendo una transición de estado definida por la función de transición probabilística T , que especifica $Pr(s'|s, a)$. En el caso más general de un Dec-POMDP, el conjunto de acciones disponibles para cada agente puede ser diferente en cada paso de tiempo y además estos no pueden conocer las acciones que ejecutan los otros agentes, con lo que a pesar de participar de acciones conjuntas sólo pueden razonar respecto a sus acciones individuales. Lo mismo pasa con las observaciones, en donde cada agente sólo puede percibir una observación individual, pero entre todos tienen el potencial de producir una percepción conjunta más informativa que requiere de comunicación.

De modo que la problemática central de los Dec-POMDPs es la coordinación de acciones bajo incertidumbre, puesto que cada agente resuelve su POMDP de manera autónoma con información local, lo que puede provocar conflictos cuando los dominios de varios agentes se interfieren, y en consecuencia la coordinación puede no ser efectiva. Tratar de evitar estos problemas de coordinación tiene un costo muy alto, puesto que calcular las políticas óptimas conjuntas se convierte en un problema *NEXP-Complete* con tan solo dos agentes [11], con lo que se buscan maneras alternativas de resolverlo que sean más factibles en términos de capacidad de cómputo descomponiendo el problema para resolverlo de manera distribuida. Esta descomposición es una tarea compleja y constituye el problema central del área de *Distributed Problem Solving* [5]. En la literatura existen diferentes enfoques para abordar esta descomposición del problema en un Dec-POMDP, la mayoría en forma de extensiones particulares de un modelo más general como los Com-POMDPs, Factored-POMDP, MA-POMDP [11] puesto que cada problema particular funciona mejor con un tipo de algoritmo particular.

En este sentido, Witwicki & Durfee [18] describen tres características de la estructura de interacción de los agentes que combinadas permiten cuantificar los beneficios de utilizar un algoritmo de descomposición particular dado un caso concreto de Dec-POMDP: (1) el *agent scope size*, que hace referencia al número de agentes que afectan a las decisiones de otros agentes, (2) el *state factor scope domain size*, que hace referencia a la porción de valores característicos de un estado que deben ser considerados por los agentes para tomar decisiones de coordinación y (3) el *influence degree*, que hace referencia a la porción de caminos únicos en que los agentes pueden influir sobre las decisiones de los demás dado un esquema de partición. Estas características permiten entender por qué las clases existentes de algoritmos de descomposición funcionan bien en ciertos problemas así como predecir el rendimiento de un algoritmo particular en base a los atributos identificables de un problema, por lo que se considera estudiar si estas características también aplican para el caso de las organizaciones humanas y pueden llegar a ser de utilidad.

A continuación, se describe brevemente el enfoque de la auto-organización holónica [16] como solución para la descomposición de problemas en un Dec-POMDP, puesto que este enfoque particular de resolución se basa en el concepto filosófico de Holarquía [9], modelo que también inspira el mecanismo de organización elegido como caso de estudio: *Holacracy*. Por lo que se considera interesante ver cómo una misma idea para descomponer problemas se adapta en función de si va a implementarse en un sistema de agentes artificiales o en un sistema de agentes humanos. En este sentido, hasta donde se ha podido indagar, existen antecedentes de

algoritmos de descomposición de problemas multiagente que utilizan el enfoque de descomposición holónico como el caso del *DFG: Dissolution and Formation of Groups* de Weiß [17].

Este modelo Holárquico para descomponer problemas de manera auto-organizada se rige por el proceso de holonificación (i.e. la transformación estructural de dividir una unidad en sub-unidades o viceversa) y permite dos operaciones reversibles: (1) unión y (2) disolución. En los problemas de coordinación en SMA, la descomposición del problema consiste en agrupar los agentes en holones, lo que simplifica el problema al asumir que los agentes de un holón o grupo no pueden incurrir en conflictos por no poseer interdependencias entre sus acciones, y por lo tanto no requieren de coordinación; ahorrándose así importantes costos computacionales. Así, una solución al problema de un Dec-POMDP consiste en una descomposición en holones por un lado, y las políticas óptimas individuales por otro, aunque estos dos aspectos de la solución no son independientes entre sí, puesto que la política óptima de un agente es afectada por los agentes de los otros holones con los que debe coordinarse, lo que resulta en la práctica en una búsqueda intercalada de la descomposición y la política. Y en consecuencia, toda modificación de la descomposición holónica requiere de volver a ejecutar costosos algoritmos de aprendizaje de la política óptima.

3 Caso de Estudio: Holacracy

En esta sección se describe el mecanismo organizacional elegido para el caso de estudio: *Holacracy* [12]; presentando sus principios y funcionamiento; para luego en la Sección (4) contrastarlo con el formalismo de Dec-POMDPs y poder así evaluar cualitativamente su grado de inteligencia, resaltando sus fortalezas y debilidades respecto a los principios que rigen a los agentes inteligentes.

El mecanismo Holacrático es un mecanismo para organizaciones humanas que propone aprovechar las capacidades perceptuales y cognitivas de los individuos de la organización para encontrar de manera distribuida la política óptima de la misma. Para ello, según Robertson [12], *Holacracy* propone un mecanismo de auto-gobernanza o liderazgo distribuido inspirado por los sistemas vivos, que distribuye el poder, convirtiendo funciones que normalmente se reservan a un CEO o supervisor en procesos que deben resolverse en forma autónoma por roles que se encuentran más abajo en la jerarquía organizacional. El nombre del mecanismo de auto-gobernanza viene inspirado en la elección de la Holarquía [9] como topología organizacional, que propone una estructura de poder compuesta por unidades autónomas llamadas holones que pueden transferir parte de su poder a uno o más sub-holones para que operen de manera autónoma (para incluso definir su propia descomposición en más sub-niveles holónicos). En la jerga de *Holacracy* los holones pueden llamarse *roles* en el caso de no contener sub-holones, o *círculos* en caso contrario. Este mecanismo es además evolutivo, en tanto propone un algoritmo humano que permite ir encontrando la Holarquía ideal y con ello una manera ideal de distribuir el poder.

Esta gobernanza distribuida se ve plasmada en la Constitución Holacrática: compendio de normas y estructuración del poder que rigen los procesos ejecutivos de la organización. Esta introduce dos elementos fundamentales de las sociedades humanas en las organizaciones: un estado de derecho, definido por el proceso de gobernanza; y derechos de propiedad, definidos a través de dominios de autoridad asignados a los roles. A continuación, se describen los elementos principales de la Constitución Holacrática que es lo que define los procesos de gobernanza:

Roles: son las unidades elementales de la estructura organizacional. Los individuos que conforman la organización participan de ella energizando uno o más roles. Un rol está definido por los siguientes elementos:

Propósito: una meta que el rol perseguirá en nombre de la organización.

Dominio: (de autoridad): delimita los aspectos del entorno que el rol puede controlar, el espacio donde tiene autonomía para actuar. Alineado con el propósito de la organización tiene la autoridad para ejecutar cualquier acción que crea que sea útil para realizar el propósito y/o atender sus responsabilidades. Bajo esta definición entonces, ningún individuo o rol puede actuar en un dominio de otra entidad soberana (i.e. un rol o un círculo)

Responsabilidades: las responsabilidades o tareas continuadas de la organización que el rol va a realizar.

Políticas: son recursos que tienen los roles/círculos para controlar sus dominios de autoridad, pueden utilizarse para prevenir que nadie impacte en sus dominios o al contrario para invitar a hacerlo.

Círculos: son las unidades colectivas de la estructura, aparecen cuando un rol decide distribuir tanto su dominio de autoridad en sub-roles o sub-círculos, como así también distribuir sus responsabilidades

La fuerza motriz de la evolución propuesta por el método Holacrático son las *tensiones*, que expresan la capacidad humana para sentir disonancias en el momento presente y ver cambios potenciales con impacto positivo. Así, sentir una tensión puede verse como percibir la diferencia entre el estado de un aspecto concreto de la realidad y el estado (mejor) que podría alcanzar. Cada tensión sentada por un humano tiene el potencial entonces de llevar a la organización a expresar mejor su propósito, pero sólo puede aprovecharlo cuando es capaz de procesar las tensiones de cualquier miembro de manera rápida y efectiva. Para ello *Holacracy* propone un algoritmo humano basado en el consentimiento llamado *Integrative Decision Process* que estructura la comunicación entre los individuos. El consentimiento implica que todos los individuos están de acuerdo implícitamente salvo que haya razones sustentadas para creer que la propuesta atenta contra la expresión del propósito de la organización. El método Holacrático se desarrolla a través de dos tipos de reuniones presenciales: las reuniones de gobernanza, donde operando bajo el *Integrative Decision Process* se modifica colaborativamente la estructura de poder con propuestas que garanticen mejorar la expresión del propósito del rol y/o de la organización; y las reuniones tácticas, que resuelven el problema local de elegir el plan óptimo para expresar el propósito del rol y/o de la organización.

Así, una organización Holacrática nace con un único círculo fundacional que se da en llamar genéricamente *Anchor Circle* y en la medida que emergen tensiones la estructura va evolucionando con propuestas (consentidas) de cambios en la Constitución: distribuyendo roles en sub-roles con la creación de círculos, disolución de círculos o refinamiento de las responsabilidades, entre otros cambios estructurales.

4 Mapeo de *Holacracy* a los formalismos de los SMA

Una vez descritas las características básicas del método Holacrático y los principales formalismos de los SMA, se presenta una comparativa cualitativa que muestra cómo el modelo Holacrático permite a las organizaciones actuar de manera más inteligente que los modelos tradicionales de gestión top-down, en los términos descritos en la Sección (2), puesto que el modelo Holacrático aprovecha las capacidades perceptuales y cognitivas de todos los individuos para mejorar la estructura organizacional en términos de facilitar la coordinación, lo que reduce drásticamente el costo

computacional de encontrar una política óptima (al realizarse de manera distribuida), repercutiendo en una mejora de la performance.

Tal y como se ha descrito anteriormente, se asume que una organización (independientemente de su modelo) puede concebirse como un SMA, con lo cual el problema principal de ambos modelos es el mismo: encontrar una política óptima para expresar mejor el propósito de la organización (lo que equivale a maximizar su rendimiento). En este sentido, y asumiendo también que una organización puede modelarse como un Dec-POMDP, se mostrará como la principal diferencia entre un modelo Holacrático y un modelo tradicional top-down radica en la solución o algoritmo asociado a cada modelo para la resolución de este Dec-POMDP.

Así, en primer lugar se va a describir el entorno de trabajo, que es lo que define los problemas basados en agentes. En este sentido, los dos modelos a comparar (*Holacracy* vs. modelo tradicional) operan en el mismo entorno de trabajo que responde a una definición genérica de la tupla REAS para una organización humana descrita en los términos de este trabajo, dónde:

- R: es la medida de rendimiento. Se asume que existe una organización formada por un colectivo de individuos que tiene un propósito común.
- E: es el entorno. Se asume que el entorno donde operan las organizaciones humanas es el mundo físico por lo que presenta su configuración más compleja: dinámico, continuo, estocástico, secuencial, parcialmente observable y multiagente.
- (A) y (S): son Actuadores (A) y Sensores (S). Se asume que en el caso de las organizaciones humanas estos son los sensores y actuadores naturales de los individuos humanos sumados a otros artefactos tecnológicos.

Dado que la inteligencia se define como la habilidad para maximizar el rendimiento bajo un mismo REAS, la comparativa entre los dos modelos se reducirá a comparar el algoritmo utilizado para encontrar la política óptima. En este sentido, REAS es una formalización muy general del problema, por lo que tal y como se ha argumentado en la Sección (2), para el caso de las organizaciones se formalizará el problema más concretamente como un Dec-POMDP, lo que implica que los modelos comparados se evaluarán en términos de soluciones para resolver un Dec-POMDP.

En el caso del modelo Holacrático, el algoritmo humano utilizado para resolver un Dec-POMDP se basa en el enfoque holónico, es decir en descomponer el problema de manera auto-organizada rigiéndose por el proceso de holonificación (i.e. la transformación estructural de dividir una unidad en sub-unidades o viceversa), por lo que en el caso de una organización la descomposición del problema consiste en agrupar a los individuos en holones (roles o círculos), lo que simplifica el problema al asumir que los agentes de un holón o grupo no incurrir en conflictos por no poseer interdependencias entre sus acciones, y por lo tanto no requieren de coordinación; ahorrándose así importantes costos computacionales (i.e. comunicación requerida para la coordinación de acciones entre individuos). Con lo que la estructura Holocrática puede verse como una búsqueda de agrupamientos de roles y círculos que sea libre de conflicto. Sin embargo, como las organizaciones humanas son un ejemplo de Dec-POMDP de horizonte infinito, es habitual que ante cambios externos aparezcan conflictos donde antes no los había, por lo que ante estas situaciones el modelo Holacrático presenta herramientas como las reuniones tácticas y las reuniones de gobernanza que permiten evaluar la envergadura de dichos conflictos y resolverlos, de manera específica en cada caso. En general se contemplan dos niveles de conflictividad: (1) por un lado, puede darse el caso que dentro de un holón formado por más de un individuo (i.e. círculo) aparezcan conflictos de pequeña escala que pueden resolverse de manera ad-hoc a través de la comunicación directa entre los individuos afectados, mediado a lo sumo por las reuniones tácticas; (2) cuando la resolución de este tipo de conflictos es demasiado costosa en términos computacionales (i.e. requiere de mucha comunicación entre los individuos), el círculo

afectado reconoce que el conflicto es de tal envergadura que requiere de una modificación estructural (i.e. un nuevo reagrupamiento libre de conflicto). Este segundo caso, donde se requiere de una reorganización estructural, no sólo puede producirse por un grado elevado de conflictividad dentro del mismo holón (intra-holónico), sino que también puede generarse por conflictividad inter-holónica (i.e. dos holones presentan un grado de dependencia muy elevado, lo que hace difícil la coordinación entre ambos). En este caso la reorganización estructural se lleva a cabo a través de las reuniones de gobernanza.

Estos procesos de descomposición libre de conflicto que ofrece el modelo Holacrático se basan en dar autonomía y poder a todos los agentes de la organización para que puedan participar de ellos, es decir se busca aprovechar las capacidades perceptuales y cognitivas de todos los individuos para realizar esta búsqueda de agrupación en holones libre de conflicto, con lo que existe una inteligencia organizacional (relativa a los procesos de agrupación en holones libres de conflicto) de cada uno de los individuos que es aprovechada. Con lo cual este modelo propone una distribución de poder en la toma de decisiones que habilita a los individuos a actuar en pos de modificar el estado organizacional, es decir, los individuos disponen de acciones organizacionales y autonomía para ejecutarlas, dando así lugar a que cada individuo exprese su inteligencia organizacional, pudiendo aprovechar sus percepciones locales y su capacidad cognitiva para decidir qué acciones (organizacionales) son las que tienen el potencial de mejorar el rendimiento de la organización. Por lo tanto, se puede afirmar que es esta inteligencia organizacional la que está operando de manera distribuida en el caso Holacrático a diferencia de los modelos tradicionales de gestión top-down, donde esta inteligencia organizacional se reserva a las altas esferas de la jerarquía organizacional. Por lo que en términos de agentes, se puede concluir que mientras que en el caso Holacrático los agentes están habilitados para ejecutar acciones organizacionales, en el caso de la organización tradicional dichos agentes tienen los actuadores inhibidos o no están habilitados para ejecutar acciones de esta tipología, por lo que a pesar de que pueden detectar situaciones de mejora (i.e. tienen percepciones y poder cognitivo) no pueden expresar su inteligencia organizacional porque no tienen poder para realizar estas acciones.

5 Discusión y Conclusiones

Todo el razonamiento utilizado en la Sección (4) ha usado un marco conceptual específico de SMA, demostrando a través de un ejemplo concreto como este marco ha permitido comparar de manera clara y efectiva dos modelos de organización humanos en unos términos novedosos (i.e. en términos de otra disciplina), ofreciendo una comparativa entre modelos de organización humanos que aporta nuevos elementos relativos a sus características que resultan clarificadores, como los conceptos de inteligencia, SMA colaborativo, algoritmos de descomposición para problemas de planificación multiagente bajo incertidumbre, así como reflexividad respecto a los costos de implementación asociados a los modelos de liderazgo distribuido.

En este sentido, esta propuesta permite superar la problemática existente relativa a la evaluación y comparación de modelos de liderazgo distribuido, que básicamente radica en el hecho de que no existe un lenguaje general que permita compararlos entre sí más allá de sus terminologías propias y conceptos ad-hoc, con lo que la solución propuesta permite homogeneizar sus características (en términos de inteligencia) y comparar modelos con terminologías, léxicos específicos y lenguajes diferentes.

Por otro lado, se considera de interés destacar que el uso de formalismos del área de la IAD y los SMA permiten considerar y evaluar los costos de implementación asociados a los modelos de liderazgo distribuido humanos, tal y como se ha

demostrado con el mapeo entre organizaciones y Dec-POMDPs, por lo que se considera que profundizar en esta cuestión permitiría mejorar modelos existentes como el Holacrático, si por ejemplo, se considerara el costo de implementación asociado como una variable más que influye en el rendimiento de la organización tal y como ocurre en algunos modelos de planificación multiagente bajo incertidumbre, puesto que estos costos de implementación asociados son lo que hacen que modelos de liderazgo distribuido como el Holacrático puedan ser demasiado costosos de implementar en organizaciones humanas con características determinadas.

En este sentido, más allá de los algoritmos holónicos (que mapean con el modelo Holacrático) Oliehoek & Amato [11] recogen algoritmos que presuntamente son más efectivos para la resolución de Dec-POMDP, por lo que se considera como trabajo futuro indagar en estos algoritmos más recientes para analizar si es factible implementarlos en organizaciones humanas a través de nuevos modelos de liderazgo distribuido, y en caso que sea posible, explorar si estos nuevos modelos serían superadores al modelo Holacrático en términos de inteligencia.

Referencias

1. Billis, D.: Hybrid organizations and the third sector. Palgrave Macmillan, Basingstoke, Hampshire (2010).
2. Bulling, N.: A Survey of Multi-Agent Decision Making. *KI - Künstliche Intelligenz*. 28, 147-158 (2014).
3. Buterin, V.: A next-generation smart contract and decentralized application platform, (2014).
4. Drucker, P.: Post-capitalist society. Routledge, [S.l.] (2016).
5. Durfee, E. et al.: Trends in cooperative distributed problem solving. *IEEE Transactions on Knowledge and Data Engineering*. 1, 1, 63-83 (1989).
6. Gasser L.: Perspectives on Organizations in Multi-agent Systems. In: Luck M., Mařík V., Štěpánková O., Trappl R. (eds) *Multi-Agent Systems and Applications. ACAI 2001. Lecture Notes in Computer Science*, vol 2086. Springer, Berlin, Heidelberg (2001)
7. Horling, B., Lesser, V.: A survey of multi-agent organizational paradigms. *The Knowledge Engineering Review*. 19, 04, 281 (2005).
8. Irandoust, H., & Benaskeur, A. R. Multi-organizational structures. In: *Proc. of association for the advancement of artificial intelligence*, Chicago, 25-33 (2008).
9. Koestler, A.: *Janus*. Picador, Londres (Inglaterra) (1978).
10. Lesser, V.: Cooperative multiagent systems: a personal view of the state of the art. *IEEE Transactions on Knowledge and Data Engineering*. 11, 1, 133-142 (1999).
11. Oliehoek, F., Amato, C.: *A Concise Introduction to Decentralized POMDPs*. Springer International Publishing, Cham (2016).
12. Robertson, B.: *Holacracy*. Henry Holt & Company, New York (2015).
13. Russell, S., Norvig, P.: *Inteligencia artificial, un enfoque moderno*. Pearson Educación, Madrid (2004).
14. Siddiqui, Y., & Brastaviceanu, T.: Open value network: A framework for many-to-many innovation (2013).
15. Spillane, J. et al.: Towards a theory of leadership practice: a distributed perspective. *Journal of Curriculum Studies*. 36, 1, 3-34 (2004).
16. Tianfield, H.: A new framework of holonic self-organization for multi-agent systems. In *Systems, Man and Cybernetics, 2007. ISIC. IEEE International Conference on* (pp. 753-758). IEEE. (2007).
17. Weiß, G.: Distributed reinforcement learning. *Robotics and Autonomous Systems*. 15, 1-2, 135-142 (1995).
18. Witwicki, S. J., & Durfee, E. H.: Towards a unifying characterization for quantifying weak coupling in Dec-POMDPs. In *The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 1* (pp. 29-36). International Foundation for Autonomous Agents and Multiagent Systems (2011).

Image recognition for x-ray luggage scanners using free and open source software

Pablo Lázaro, Ariel Maiorano

Dirección de Gestión Tecnológica (DGT), Policía de Seguridad Aeroportuaria (PSA)
{plazaro, amaiorano}@psa.gob.ar

Abstract. This paper describes the status of a currently ongoing research and development project that proposes the implementation of a system to be connected to x-ray luggage scanning equipment used by the Airport Security Police (*Policía de Seguridad Aeroportuaria*, PSA), having as main objective to automatically recognize images obtained in real time from the scanners to alert the x-ray machine operators or screeners about those identified as threats (explosive devices, firearms, etc.) Technical instructions and source code of a web-service and auxiliary scripts developed for the evaluations were published on github in order to enable interested researchers, developers and other potential contributors to repeat or reproduce the training and testing summarized in this paper with freely available image data-sets.

Keywords: luggage control, x-ray images, airport security, explosive detection, firearms detection, image recognition, computer vision, deep neural networks, machine learning, TensorFlow library.

1 Introduction

Given the recent advances in computerized image recognition [1,6,7] and the availability of free and open-source software [12,13] that facilitates the practical application of these techniques [14,15,16], the project proposes the implementation of a system that, connected to the x-ray luggage scanning equipment currently in use by the PSA (Argentine Airport Security Police, or *Policía de Seguridad Aeroportuaria*), will have as main objective to automatically recognize images obtained from the scanners to alert the operator about images identified as threats (explosive devices, firearms, knives, other weapons or dangerous items, etc.)

Although a fair amount of literature has been published regarding this subject (for example [1,2,6,7,8,9,10]), including papers dealing with the recognition of x-ray images in search of threats such as those mentioned above, the data-sets or sets of images used in most of these publications are not in the public domain. Another factor that makes it difficult to reproduce published results is the availability of necessary software or the capacity to develop it.

On the basis of the above, in order to confirm the feasibility of the DGT project, and to be able to determine if the threshold or margin of precision of the estimates that the system would generate are acceptable, tests were performed on two data-sets; sample images obtained from PSA scanners, as well as on other sets of images in the

public domain [3,4]. Regarding software, the decision was made to implement the classification software using Google's free and open source software library for machine learning, TensorFlow [12,13].

This paper will include information about how the experimental models were trained and evaluated. Two of the data-sets used in this work are based in a freely downloadable one in the public domain; other, such as the one composed of sample images included by the manufacturer of x-ray luggage scanners from the PSA, are reached by redistribution constraints.

The technical instructions, source code of the web-service for classification of images that are passed to it as input, and general auxiliary scripts, will be published on github [18] in order to permit interested third parties to repeat the tests with the freely available sets of images.

2 Training and evaluation

As already stated, we used TensorFlow, an open-source, free-to-use library for the implementation of machine learning techniques in Python and C++ programming languages. Our goal was to test the performance of recognition of images obtained online from luggage scanners in order to alert or alarm, if applicable, about images identified as threats.

Regarding data-sets, with focus on explosive devices detection, images included by the manufacturer into x-ray luggage scanners used in the PSA were obtained and adapted to compose the first data-set. A second and third data-sets, aiming at the detection of firearms and knives, were composed with images obtained from a public x-ray images database that can be used free of charge for research and educational purposes.

2.1 TensorFlow library from Google

As stated in [12], TensorFlow is an open-source software library for Machine Intelligence. It is library for numerical computation using data flow graphs. Nodes in the graph represent mathematical operations, while the graph edges represent the multidimensional data arrays (tensors) communicated between them. The flexible architecture allows the deployment of computation to one or more CPUs or GPUs in a desktop, server, or mobile device with a single API. TensorFlow was originally developed by researchers and engineers working on the Google Brain Team within Google's Machine Intelligence research organization for the purposes of conducting machine learning and deep neural networks research, but the system is general enough to be applicable in a wide variety of other domains as well.

As stated in [15], modern object recognition models have millions of parameters and can take weeks to fully train. Transfer learning is a technique that shortcuts a lot of this work by taking a fully-trained model for a set of categories like ImageNet, and retrains from the existing weights for new classes. Inception-v3 model from Google was trained for the ImageNet Large Visual Recognition Challenge using the data from

2012. This is a standard task in computer vision, where models try to classify entire images into 1000 classes, like "Zebra", "Dalmatian", and "Dishwasher".

In this work the final layer of the Google's pre-trained Inception v3 model was retrained using this technique. To perform transfer learning the final classification layer of the network was removed and the next-to-last layer of the Convolutional Neural Network (CNN) extracted, in this case a 2048 dimensional vector. More information regarding this approach could be obtained, for example, from [19,20, 21,22].

2.2 PSA luggage scanners data-set

Example images included in equipment currently in use by the PSA were extracted to gather a data-set containing x-ray images classified in three groups: explosives, firearms, and general (negative) luggage.

After discarding duplicated and not suitable images, clipping of images with to increase the number of negative samples, rotation at different angles, changes in size, quality, etc.; the quantities per class obtained are shown in table 1.

Table 1. PSA luggage scanners data-set.

Explosive artifacts	4800
Firearms	2160
General luggage (negative)	1500

Example images of this data-set are shown in figures 1 and 2. Firearms class include images of pistols and revolvers; 357, 9mm, 38, 22; from different manufacturers (Magnum, Bersa, Colt, etc.) Explosive materials include C-4, Dynamite, Semtex, etc., with different switches or timers, batteries, and initiators.

This data-set objective was primarily used to train a model that would recognize explosive devices.

As summarized in [11], one of the primary weapons of choice for terrorists and lone wolf attackers has been the Improvised Explosive Device, or IED. These dangerous, homemade devices are typically constructed from common household items making them rather inexpensive to build.

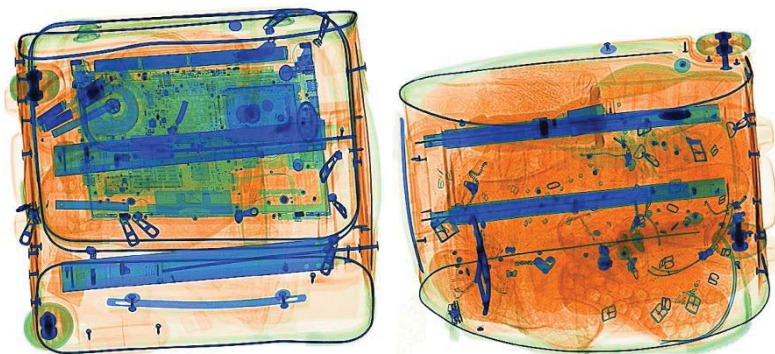


Fig. 1. Sample image from a PSA scanner showing luggage containing a laptop (left) and another one showing shoes, sneakers and slippers (right). These images are included in the “negative” class.



Fig. 2. Sample scanner image containing explosive artifacts composed of dynamite with a 9 volt battery (left) and TNT with AAA batteries (right). Both images belong to the “explosive” or “positive” class.

Images are colored, by many x-ray luggage scanners (including the ones used in this work), as follows:

- BLUE: "Hard" materials; Metals (also in BLACK), hard plastics, alloys; firearms, knives, cables, batteries, etc.
- ORANGE: Biological material mainly; Not exclusively natural; Leather, food, dynamite and other explosives (not plastic ones); Liquids, gels, organic powders, etc.
- GREEN: Alloys of lower density; Also ceramics; Not organic; Combined with BLUE in electronic devices.

2.3 GDXray - GRIMA database - firearms data-set

From [4], as a service to the X-ray testing and Computer Vision communities, Domingo Mery and his team collected more than 19.400 X-ray images for the development, testing and evaluation of image analysis and computer vision algorithms. The images are organized in this public database called GDXray: The GRIMA X-ray database (GRIMA is the name of a Machine Intelligence Group at the Department of Computer Science of the Pontificia Universidad Catolica de Chile). The X-ray images included in GDXray can be used free of charge, for research and educational purposes only.

The quantities per class, in two reduced data-sets made up for this particular project, to test the detection of firearms and of knives, separately, are shown in table 2 and table 3. Dataset was partitioned by the authors; all images from the original dataset were used and the final labeling of each class can be obtained from the github repository.

Table 2. Firearms GDXray data-set quantities per class.

Firearms	1117
General luggage (negative)	8445

Table 3. Knives GDXray data-set quantities per class.

Knives	2246
General luggage (negative)	7321



Fig. 3. X-ray sample images of firearms and general luggage from the GDXray GRIMA database.

3 Results of first experiments

According to common practice in training and evaluating machine learning models, and also following the default method in TensorFlow library, data-sets were divided into three subsets: a training subset comprising 80% of the complete, original data-set; one for validation, containing 10% of the samples -also used during the training-; and a last subset, for the final testing, containing the remaining 10% of the data-set images, which are left aside until this final, one time evaluation.

It is also worth noting that the results are represented in approximate ranges given that various training and data augmentation experiments have been performed, besides that in each case, there are variables that are initialized with random values.

3.1 Results obtained for the PSA images data-set

Using the PSA images data-set, we have evaluated the training, validation and testing, with three classes (explosives, firearms, and negative); With the main objective of recognizing explosive devices. Results are shown in table 4, considering explosive artifacts exclusively.

Table 4. Results for the PSA data-set.

Results for training subset	95 – 100 %
Results for validation subset	89 – 98 %
Results for testing subset	84 – 90 %

3.2 Results obtained for the GDXray GRIMA data-set

For the two GDXray GRIMA based sub data-sets (of firearms and knives), the results of training, validation and testing phases are summarized in tables 5 and 6.

Table 5. Results for the GDXray GRIMA firearms sub data-set.

Results for training subset	98 – 100 %
Results for validation subset	95 – 99 %
Results for testing subset	92 – 98 %

Table 6. Results for the GDXray GRIMA knives sub data-set.

Results for training subset	97 – 100 %
Results for validation subset	95 – 98 %
Results for testing subset	90 – 97 %

4 Open source project – RIS

The project was named RIS (for *Reconocimiento de Imágenes de Scanners* in spanish, or Scanners Image Recognition). General instructions and source code of the web-service and auxiliary scripts developed for the evaluations will be published on github at <https://github.com/dgt-psa/ris>. Information regarding the web-service for the classification of images of x-ray scanners can be found in the project README.md, including obtaining estimates from generated models, installing TensorFlow on Linux, installing the rest of the web service dependencies, the generation of new models, re-training the last layer of the Google inception model, etc.

The tests have been performed on Ubuntu server 14.04, with Python 2.7 and the installation of required software was done accordingly to what is described in https://www.tensorflow.org/install/install_linux#InstallingVirtualenv.

The project was first presented in ANDSEC security conference, when source code was made available in DGT's github repository.

This software can be redistributed and/or modified under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or any later version.

5 Future work

Planned future work includes continuing the compilation of relevant information from academic, government (other countries) and private experiences; and the evaluation of other free software libraries that enable the implementation of machine learning for image recognition.

Besides the implementation and operation of the online analysis of images for alerting in support or as a “second opinion” for the scanner operator, and regarding research specific tasks, the list includes:

- Acquisition of more x-ray images (positives and negatives) from PSA scanners.
- Evaluate other sets of images for training, validation and testing (from researchers, universities, other security or police institutions, etc.)
- Evaluate training and evaluation with different types of machine learning classifiers.
- Confirm technical feasibility and threshold or acceptable range of precision.
- Getting help from potential collaborators: contributions from the manufacturers and/or distributors of x-ray luggage scanners and/or other third parties.
- Train with dedicated hardware, considering the important performance difference in equipment with, for example, special video cards.
- Consider the use of the same technology for other applications, for example, the recognition of clandestine airstrips.

6 Conclusions

This research and development project, currently in progress, aims to confirm that available free and open source software could be used for image recognition to alert the operator (or screener) of x-ray luggage scanners about images identified as threats (explosive devices, firearms, knives, etc.)

Although we need to continue experimenting with different alternatives with regarding to software, pre-trained models, and image-sets, to confirm if the accuracy or threshold of acceptance of the classifications, we consider the results obtained in this first stage of the project encouraging.

Sharing on github the experiments that were carried out, the tools and the methodology used, we hope that interested third parties can reproduce the work described in this paper and eventually collaborate openly on the project.

7 Acknowledgements

The organic and functional structure of the PSA (Airport Police Security, or *Policía de Seguridad Aeroportuaria*) is integrated by a civilian management and administration structure, an operational police structure, and the higher institute of airport security. The DGT (Technological Management Department, or *Dirección de Gestión Tecnológica*) is responsible for managing, developing, providing and controlling the administrative, technological and communication support goods and services for data processing and access to information, as well as researching and implementing state-of-the-art technologies that ensure availability, integrity and reliability of information. (Ministry of Security Resolution # 1015/12).

The authors of this work would like to thank the PSA authorities for their support and encouragement in developing this project.

References

1. Vladimir Riffo, Domingo Mery, "Automated Detection of Threat Objects Using Adapted Implicit Shape Model", *Systems Man and Cybernetics: Systems IEEE Transactions on*, vol. 46, pp. 472-482, 2016, ISSN 2168-2216.
2. Domingo Mery, Erick Svec, Marco Arias, Vladimir Riffo, Jose M. Saavedra, Sandipan Banerjee, "Modern Computer Vision Techniques for X-Ray Testing in Baggage Inspection", *Systems Man and Cybernetics: Systems IEEE Transactions on*, vol. 47, pp. 682-692, 2017, ISSN 2168-2216.
3. Mery, D.; Riffo, V.; Zscherpel, U.; Mondragón, G.; Lillo, I.; Zuccar, I.; Lobel, H.; Carrasco, M. (2015): GDXray: The database of X-ray images for nondestructive testing. *Journal of Nondestructive Evaluation*, 34.4:1-12.
4. Mery, Domingo. GDXray - GRIMA Database website. [Online] Available: <http://dmery.ing.puc.cl/index.php/material/gdxray/> [Accessed 20 July 2017].
5. Wicher Visser, Adrian Schwaninger, Diana Hardmeier, Alexander Flisch, Marius Costin, Caroline Vienne, Frank Sukowski, Ulf Hassler, Irène Dorion, Abraham Marciano, Ger Koomen, Micha Slegt, Andrea Cesare Canonica, "Automated comparison of X-ray images

- for cargo scanning", Security Technology (ICCST) 2016 IEEE International Carnahan Conference on, pp. 1-8, 2016, ISSN 2153-0742.
6. Jaccard, N; Rogers, TW; Morton, EJ; Griffin, LD; (2016) Tackling the X-ray cargo inspection challenge using machine learning. In: Ashok, A and Neifeld, MA and Gehm, ME, (eds.) Anomaly Detection and Imaging with X-Rays (ADIX), 98470N (May 12, 2016); doi:10.1117/12.2222765. (pp. 98470N). Society of Photo-Optical Instrumentation Engineers (SPIE): Bellingham, Washington, USA.
 7. R. Gesick, C. Saritac, and C. Hung, "Automatic image analysis process for the detection of concealed weapons," in Proceedings of the 5th Annual Workshop on Cyber Security and Information Intelligence Research, 2009.
 8. K. Wells, D.A. Bradley, A review of X-ray explosives detection techniques for checked baggage, Applied Radiation and Isotopes, Volume 70, Issue 8, 2012, Pages 1729-1746, ISSN 0969-8043, <http://dx.doi.org/10.1016/j.apradiso.2012.01.011>.
 9. (<http://www.sciencedirect.com/science/article/pii/S0969804312000127>)
 10. G. Zentai, "X-ray imaging for homeland security", Proc. IEEE Int. Workshop Imag. Syst. Tech. (IST), pp. 1-6, Sep. 2008.
 11. Train, V. Safe Passage Security Training Blog. The Basics of an IED and its PIES Components. [Online] Available: <http://blog.safe-passage.com/the-basics-of-an-ied-and-its-pies-components> [Accessed 20 July 2017].
 12. Google Inc. TensorFlow library. [Online] Available: <https://www.tensorflow.org/> [Accessed 20 July 2017].
 13. Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Rafal Jozefowicz, Yangqing Jia, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Mike Schuster, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.
 14. Görner, Martin. Google Cloud Big Data and Machine Learning Blog. Learn TensorFlow and deep learning, without a Ph.D. [Online] Available: <https://cloud.google.com/blog/big-data/2017/01/learn-tensorflow-and-deep-learning-without-a-phd> [Accessed 20 July 2017].
 15. Google TensorFlow tutorials. How to Retrain Inception's Final Layer for New Categories. [Online] Available: https://www.tensorflow.org/tutorials/image_retraining [Accessed 20 July 2017].
 16. Hawkins, Chris. Recognising Beer with TensorFlow. [Online] Available: <https://medium.com/emergent-future/recognising-beer-with-tensorflow-9dedfee3c3c0> [Accessed 20 July 2017].
 17. Sémantique Données non-structurées, Machine Learning. Image classification with a pre-trained deep neural network. [Online] Available: https://www.kernix.com/blog/image-classification-with-a-pre-trained-deep-neural-network_p11 [Accessed 20 July 2017].
 18. DGT-PSA. RIS project github repository. [Online] Available: <https://github.com/dgt-psa/ris> [Accessed 20 July 2017].
 19. Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, Trevor Darrell. DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition. [Online] Available: <https://arxiv.org/pdf/1310.1531.pdf> [Accessed 18 September 2017].

20. Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich. Going Deeper with Convolutions. [Online] Available: <https://arxiv.org/pdf/1409.4842.pdf> [Accessed 18 September 2017].
21. Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, Zbigniew Wojna. Rethinking the Inception Architecture for Computer Vision. [Online] Available: <https://arxiv.org/pdf/1512.00567.pdf> [Accessed 18 September 2017].
22. Google's Tensorflow github repository. Inception in Tensorflow. [Online] Available: <https://github.com/tensorflow/models/tree/master/inception> [Accessed 18 September 2017].

Improving Artificial Bee Colony Algorithm with Evolutionary Operators

Gabriela Minetti¹ and Carolina Salto^{1,2}

¹ Facultad de Ingeniería, Universidad Nacional de La Pampa

² CONICET, Argentina

email: {minettig, saltoc}@ing.unlpam.edu.ar

Abstract. In this paper, we analyze the effect of replacing the method to create new solutions in artificial bee colony algorithm by recombination operators. Since the original method is similar to the recombination process used in evolutionary algorithms. For that purpose, we present a systematic investigation of the effect of using six different recombination operators for real-coded representations at the employed bee step. All analysis is carried out using well known test problems. The experimental results suggest that the method to generate a new candidate food position plays an important role in the performance of the algorithm.

Keywords: ABC algorithm - recombination - parameter tuning

1 Introduction

Swarm intelligence is the study of computational systems inspired by the co-operation of large numbers of homogeneous agents in the environment. An ant colony, a flock of birds, a honeybee or an immune system are typical examples of swarm systems. Tereshko and Loengarov [1] consider a bee colony as a dynamical system, where gathering information from an environment and adjusting its behavior in accordance to it produce an intelligent decision-making from enhancing the level of communication among the individuals. Taking the Tereshko and Loengarov's ideas into account, Karaboga et al. [2, 3] propose an optimization algorithm based on the intelligent behavior of honey bee swarm, called Artificial Bee Colony (ABC) algorithm. Considering that it works with a set of solutions, ABC is classified as a population-based metaheuristic.

In general, metaheuristics have a major drawback: they need some parameter tuning that is not easy to perform in a thorough manner. Those parameters are not only numerical values but may also involve the use of search components. Those parameters may have a great influence on the efficiency and effectiveness of the search. The ABC metaheuristic is not the exception [3].

Some works have deal with this problem, mainly by investigating value ranges of parameters more suitable to the solution of problems with certain features [4, 5]. Regarding the search components, several ways to create the initial solutions were presented in [6, 7] and local search methods to generate solutions at scout bee step have been researched in [6–8]. In [9], the authors changed the way to

calculate the selection probability of a solution by introducing the Euclidean distance between two solutions in the probability equation. Furthermore, different approaches introducing partial changes in the method to generate new food positions at each employed and onlooker bee step have been proposed [7, 9–11].

The cited proposals consider different solutions or add factors into the original method to create a candidate food position, but no new entirely distinct ways to generate it are presented. The mechanism used by ABC to produce a new candidate solution is very similar to the procedure carried out by the recombination operators for real-coded spaces in the evolutionary algorithm literature. In this sense, we consider the application of other mechanisms to generate new source positions using the recombination operators, whose effect on the ABC performance has not yet been studied and their impact could be more significant than the traditional approach. Consequently, the objective of this article is to analyze the effect of using six different recombination operators to create new food positions at the employed bee step of the ABC algorithm. With respect to the methodology followed to analyze the results, we study the performance of these ABC variants with respect to the traditional ABC version considering both the solution quality and computational effort. Therefore, the effectiveness and efficiency of three of the six proposed operators in comparison with the original ABC version is shown.

The rest of this article is organized as follows. In Section 2 and 3, we describe ABC and the suggested recombination operators, respectively. Section 4 explains the experimental analysis and the methodology used. Then, we study and analyze the results obtained by the different ABC variants in Section 5. Finally, we present our principal conclusions and future lines of research.

2 Artificial Bee Colony Algorithm

The artificial bee colony is one of the several algorithms have been developed depending on different intelligent behaviors of honey bee swarms. In the ABC algorithm, the position of a food source represents a possible solution to the optimization problem and the nectar amount in this source corresponds to the quality (fitness) of the associated solution. Three kinds of artificial bees can act on the food sources: *(i)* scout bees search new food sources in the environment surrounding the nest in a random way; *(ii)* onlooker bees wait in the nest waggle dances exerted by the other bees to establish food sources; and *(iii)* employed bees are associated with a particular food source. The last kind of bees finds and exploits new food sources, memorizes their locations, loads a portion of nectar to the beehive, and unloads it to the food area in the hive.

In the Algorithm 1 a pseudo-code of the ABC [2, 3] is shown. At the first step, an initial population of SN solutions is randomly generated. Each solution x_i ($i = 1, 2, \dots, SN$) is a D -dimensional vector. Here, D is the dimension of the function or problem to be optimized. Secondly, this population is iteratively modified by the employed, onlooker and scouts bees.

Algorithm 1 ABC Algorithm

```

1: Initialize the population of solutions  $x_i, i = 1, 2, \dots, SN$  ;
2: Evaluate the population;
3: repeat
4:   The employed bees generate the new solutions,  $v_i$ , from each  $x_i$  and evaluate them;
5:   The employed bees apply the greedy selection mechanism;
6:   The onlooker bees generate the new solutions,  $v_i$ , from the selected  $x_i$  and evaluate them;
7:   The onlookers bees apply the greedy selection mechanism;
8:   The scouts bees determine the abandoned solutions and replace them with new solutions,  $x_i$ 
9:   Memorize the best solution found so far;
10: until the stop criterion is meet
11: return The best solution;

```

The employed bees generate a candidate food position, v_i , from the old one in memory, modifying only the parameter j as is shown in the Equation 1:

$$v_{ij} = x_{ij} + \phi_{ij}(x_{ij} - x_{kj}), \quad (1)$$

where $k \in \{1, 2, \dots, SN\} \wedge k \neq i$ and $j \in \{1, 2, \dots, D\}$ are randomly chosen indexes. The value ϕ_{ij} is a random number between $[-1, 1]$, which is used to control the generation of food sources around x_i and compare the two food positions viewed by a bee. As the difference between the parameters x_{ij} and x_{kj} decreases, the perturbation on the position x_{ij} gets decreased, too. Thus, as the search approaches the optimum solution in the search space, the step length is adaptively reduced. If a parameter value produced by this operation exceeds its predetermined limit, the parameter can be set to an acceptable value, e.g. to its limit value.

An artificial onlooker bee chooses a food source, x_k , depending on the probability value associated with that food source and, then it applies the Equation 1 to obtain the candidate food position. In other words, a solution is chosen using the proportional selection method. But with this method a high pressure is applied to the selection. In order to avoid it in our experiments, the proportional selection method is replaced by the binary tournament selection [12].

When a candidate source position, v_i , is created and evaluated by an artificial bee, its fitness is compared with the old x_i one. If the new solution is equal or better than the old one, the new one takes the place in the memory. In other words a greedy selection mechanism is applied to select between v_i and x_i .

A food source is assumed to be abandoned when a position cannot be improved further through a predetermined number of cycles, known as *limit* for abandonment. When a food source is abandoned by the employed bees, it is replaced with a new food source found by the scouts. This is simulated by generating a random new position to replace the abandoned one.

3 Recombination Operators

To produce a candidate food position from the old one in memory, the employed bees uses a variation operator as the one presented in Equation 1. This technique, which generates one new candidate solution by combining the information

contained in two existing ones, is similar to the recombination operators from the evolutionary algorithms. In consequence, this section presents the recombination operators included in the experimental section. The most of them comes from the real-coded genetic algorithms literature and has never been used with this metaheuristic. The proposed ABC algorithms use the recombination operator to create a single trial vector. In what follows we name the ABC algorithm using the traditional variation operator as ABC_{TRAD} .

Let us assume that x_i and x_k ($i, k \in \{1, 2, \dots, SN\}$) are the two solutions considered to produce a candidate food position v_i . For a randomly chosen parameter j ($j \in \{1, 2, \dots, D\}$) in the solution, the operators sketch in the following description can be incorporated to the ABC instead of using Equation 1. The resting parameters l ($l \in \{1, 2, \dots, D\} \wedge l \neq j$) of v_i come from x_i .

Arithmetical Recombination (ABC_{AX}). This operator [13] chooses the parameter values of the candidate food position (v_{ij}) somewhere around and between the parameter values of x_{ij} and x_{kj} . Let $\lambda \in (0, 1)$ be an uniform random value, which are chosen for each new candidate solution. Then, the j -th parameter value of the candidate solution v_i is computed according to Equation 2.

$$v_{ij} = (1 - \lambda)x_{ij} + \lambda x_{kj} \quad (2)$$

Max-Min-Arithmetical Recombination (ABC_{MMAX}). In this case, the operator [13] generates four candidate solutions according to Equation 3. After evaluating them, a greedy selection mechanism is considered. Thereupon, four additional evaluations per new candidate solution are required. For the first two candidates, a $\lambda \in (0, 1)$ value is used that is an uniform random value.

$$\begin{aligned} v_{ij1} &= \lambda x_{ij} + (1 - \lambda)x_{kj} \\ v_{ij2} &= (1 - \lambda)x_{ij} + \lambda x_{kj} \\ v_{ij3} &= \min(x_{ij}, x_{kj}) \\ v_{ij4} &= \max(x_{ij}, x_{kj}) \end{aligned} \quad (3)$$

Linear Recombination (ABC_{LX}). This operator [14] is similar to the AX, but the λ remains fix and can take two possible values 0.5 and 1.5. Three candidate solutions v_i are generated according to Equation 4. After evaluating each new candidate position, the best one is considered. Consequently, this method requires three additional evaluations per new candidate solution.

$$\begin{aligned} v_{ij1} &= 0.5(x_{ij} + x_{kj}) \\ v_{ij2} &= 1.5x_{ij} - 0.5x_{kj} \\ v_{ij3} &= -0.5x_{ij} + 1.5x_{kj} \end{aligned} \quad (4)$$

Previous operators only change one parameter $j \in \{1, 2, \dots, D\}$ in v_i and the other ones are copied from x_i but, in what follows, more than one parameter j in the solution v_i is different than x_i ones. Another change is that the value of each parameter remains without any adjustment, i.e., it is only copied from x_i or x_k , depending of the operator, but no combinations of values for this parameter are made. The operators are the following.

Binomial Recombination (ABC_{BX}). The parameter values of the candidate food position are chosen from x_i or x_k (see Equation 5), depending on a random value u to be lower than the probability parameter $\rho \in (0, 1)$, which is defined by the user [15]. For this work, the ρ value was set to 0.5. Moreover, BX generates the candidate food position ensuring that it gets at least one variable from the k -th food position, as indicated in Equation 6.

$$v_{ij} = \begin{cases} x_{kj} & \text{if } u \leq \rho \\ x_{ij} & \text{otherwise} \end{cases} \quad (5)$$

$$v_{ij} = \begin{cases} x_{kj} & \text{if } v_{ij} = x_{ij} \\ v_{ij} & \text{otherwise} \end{cases} \quad (6)$$

One-Point Recombination (ABC_{1PX}). This operator randomly selects a cut point $p \in (1, D)$ and the tails of x_i and x_k are swapped to get the candidate food position, as is seen in Equation 7.

$$v_i = \{x_{i1}, \dots, x_{ip}, x_{k(p+1)}, \dots, x_{kD}, \} \quad (7)$$

Multi-Point Recombination (ABC_{mPX}). In this operator [16], m different cut points ($m_p \in (1, D - 1)$) are chosen at random with no duplicates and sorted into ascending order. Then, the variables between successive cut points are exchanged between x_i and x_k to produce a new candidate food position, as is shown in Equation 8.

$$v_i = \{x_{i1}, \dots, x_{im_1}, x_{k(m_1+1)}, \dots, x_{k(m_2)}, x_{i(m_2+1)} \dots x_{iD}, \} \quad (8)$$

4 Experimental Design

In this section we describe the experimental design used in this work to compare ABC_{TRAD} with the six different algorithmic variants: ABC_{AX} , ABC_{BX} , ABC_{LX} , ABC_{1PX} , ABC_{mPX} , and ABC_{MMAX} . Furthermore, the execution environment and the analysis methodology are detailed in this section.

A popular choice for evaluating the performance of algorithms in the literature is to use the IEEE CEC'2008 test suite [17]. This benchmark is specially designed with large scale real-parameter minimization problems (i.e. of dimensions $D=100, 500, \text{ and } 1000$). The mean and standard deviation of the error value are used to measure the performance of the algorithmic variants. The error is calculated as the difference between the current value of the global optimum and the best found value. Particularly, for function F7, the absolute value of the obtained optimum is recorded and compared, because for that function, the globally optimal function value is unknown.

In the experiments, the seven ABC variants use the same parameter settings. The colony size SN was set to 50. The control parameter *limit* is defined by $limit = SN \times D$ [2]. The stop criterion is to achieve the maximum number of function evaluations, computed as suggested in [17] ($5000 \times D$). Notice that we are not using highly specialized ABC algorithms, since our goal is not to outperform other algorithms, for the considered test suite, but to analyze the influence of the different mechanisms to generate candidate food sources in the behavior of the proposed algorithms.

Because of the stochastic nature of the algorithms, we performed 30 independent runs of each test to gather meaningful experimental data and apply statistical confidence metrics to validate our results and conclusions. Before performing the statistical tests, we first checked whether the data followed a normal distribution by applying the Shapiro-Wilks test. Since the results do not follow a normal distribution, the non-parametric Kruskal-Wallis (KW) test is applied. Then, the pair-wise Wilcoxon test is used, in order to assess individual differences in the performance of the algorithms. This pair-wise test must be adjusted in order to compensate for the family-wise-error derived from the performance of multiple comparisons, using the Holm's method. Furthermore, multiple comparisons using the Tukeys range test are used. All tests are carried out considering as significance value $\alpha = 0.01$.

5 Analysis of results

In the following, we present an analysis of the results obtained by the six ABC variants described in Section 3 and ABC_{TRAD} to solve the CEC'2008 functions. Our main goals are to offer different and efficient mechanisms that will be used by the employed bees to generate candidate food positions. For that purposes, we analyze the quality of results considering the error values obtained by ABC for the functions from $f1$ to $f6$ and the objective values for $f7$. Additionally, we study the hit rate and the distribution of the error values presented for all algorithms. Moreover the computational effort to find the best solution is evaluated.

In the Table 1, the mean error values found by the ABC variants are shown. The best values are bold faced. From this table, two well differentiable groups of algorithms can be observed when the quality of the solutions is considered, regardless of the dimension. The first group, which consists of ABC_{TRAD} , ABC_{AX} , ABC_{LX} , and ABC_{MMAX} , obtains the lowest error values, and some of them find the optimal values for the functions $f1$, $f5$, and $f6$ in all runs. In the second group, which is composed of the remaining algorithms, their best solutions are far from the optimum value (error values bigger than $1.28E+02$). These differences among the algorithms is statistically supported by post-hoc tests after KW test (p -values $< 4.9E-10$), as shown in the last column with the symbol $+$. The poor performance of algorithms in the second group is due to the method to generate the new candidate solution, which does not introduce any adjustment to the parameters at the employed bee step through the search process. Moreover, multi-point and binomial operators reduce the representational bias at the expense of increasing the disruption of parameters.

The ranking of the variants across the all dimensions is shown in Figure 1. To obtain these ranks, the mean errors of all variants on a same function and dimension were ranked from best (rank 1) to worst (rank 7). In the case of ties, average ranks are computed. Additionally, the $\#func$ row on top of each bar indicates the number of functions where each ABC variant finds the optimum value. This figure empirically confirms the differences between the two above

Table 1. Mean error values from $f1$ to $f6$, and mean objective values for $f7$.

Func.	Dim.	ABC _{TRAD}	ABC _{AX}	ABC _{BX}	ABC _{LX}	ABC _{1PX}	ABC _{mPX}	ABC _{MMAX}	KW
$f1$	100	0.00E+00	0.00E+00	4.88E+05	2.09E-01	1.42E+05	5.30E+04	0.00E+00	+
	500	0.00E+00	0.00E+00	3.03E+06	3.98E+00	1.80E+06	1.23E+06	0.00E+00	+
	1000	0.00E+00	0.00E+00	6.20E+06	3.81E+00	4.28E+06	3.29E+06	1.03E-03	+
$f2$	100	6.35E+01	7.65E+01	1.58E+02	8.71E+01	1.38E+02	1.28E+02	8.05E+01	+
	500	1.39E+02	1.01E+02	1.83E+02	1.14E+02	1.72E+02	1.64E+02	1.10E+02	+
	1000	1.60E+02	1.10E+02	1.87E+02	1.25E+02	1.80E+02	1.76E+02	1.24E+02	+
$f3$	100	7.87E+00	1.23E+01	4.39E+11	2.31E+01	5.92E+10	1.47E+10	1.70E+02	+
	500	1.30E+01	3.03E+01	3.44E+12	6.84E+01	1.50E+12	8.13E+11	8.35E+02	+
	1000	1.49E+01	5.51E+01	7.36E+12	1.81E+02	4.08E+12	2.68E+12	1.70E+03	+
$f4$	100	1.30E+00	4.46E-02	2.25E+03	3.35E+00	9.66E+02	5.42E+02	2.53E-01	+
	500	1.81E+01	5.17E+00	1.24E+04	2.60E+01	8.50E+03	6.78E+03	1.30E+00	+
	1000	4.10E+01	1.30E+01	2.52E+04	6.02E+01	1.97E+04	1.66E+04	2.54E+00	+
$f5$	100	0.00E+00	0.00E+00	4.15E+03	2.28E-02	1.11E+03	4.48E+02	0.00E+00	+
	500	0.00E+00	0.00E+00	2.58E+04	2.35E-01	1.50E+04	1.02E+04	0.00E+00	+
	1000	0.00E+00	0.00E+00	5.51E+04	2.04E-01	3.84E+04	2.97E+04	0.00E+00	+
$f6$	100	0.00E+00	0.00E+00	2.13E+01	3.79E-02	1.97E+01	1.80E+01	3.59E-02	+
	500	0.00E+00	0.00E+00	2.15E+01	4.63E-01	2.10E+01	2.06E+01	1.29E-01	+
	1000	0.00E+00	0.00E+00	2.15E+01	5.32E-01	2.12E+01	2.10E+01	1.56E-01	+
$f7$	100	-9.09E+02	-9.14E+02	-6.98E+02	-9.73E+02	-7.25E+02	-7.40E+02	-9.48E+02	+
	500	-5.50E+03	-5.52E+03	-3.12E+03	-5.76E+03	-3.33E+03	-3.33E+03	-5.61E+03	+
	1000	-1.16E+04	-1.15E+04	-6.03E+03	-1.19E+04	-6.37E+03	-6.34E+03	-1.16E+04	+

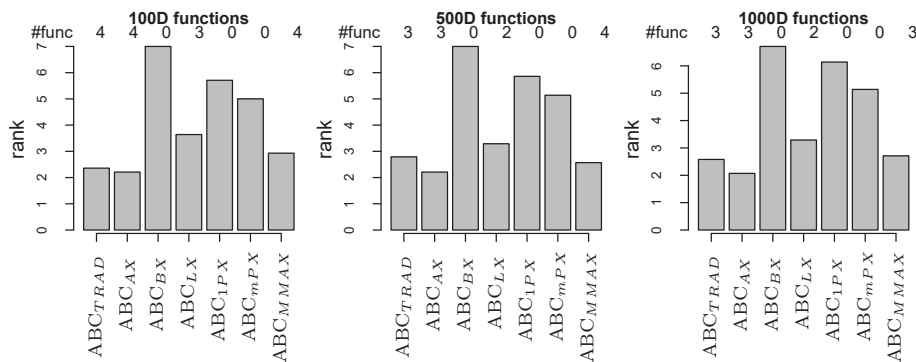


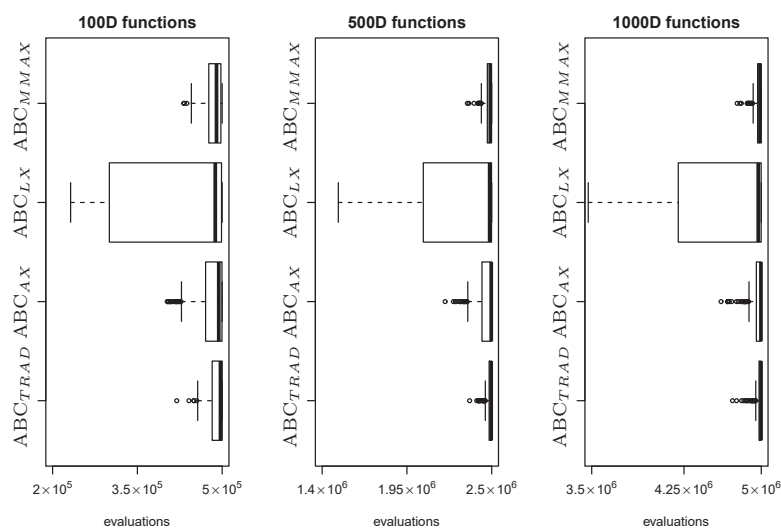
Fig. 1. Performance comparison based on the average rank over 7 functions. The ranking was computed using the average error value of each algorithm.

mentioned groups, since ABC_{AX}, ABC_{TRAD}, ABC_{MMAX}, and ABC_{LX} (in this order) are the best ranked algorithms. Furthermore, the three first algorithms solve optimally between three and four functions, while the all algorithms of the second group are not able to find the optimal value for any function and dimension ($\#func = 0$).

Regarding the previous analysis, we continue the result study considering the first group of algorithms. The Table 2 shows the minimum values obtained by each algorithmic variant considering all functions and dimensions. Additionally,

Table 2. Minimum error values from $f1$ to $f6$, and minimum objective values for $f7$.

Func.	Dim.	ABC _{TRAD}	ABC _{AX}	ABC _{LX}	ABC _{MMAX}
$f1$	100	0.00E+00	0.00E+00	0.00E+00	0.00E+00
	500	0.00E+00	0.00E+00	1.06E-01	0.00E+00
	1000	0.00E+00	0.00E+00	2.94E-01	0.00E+00
$f2$	100	5.74E+01	6.84E+01	7.67E+01	7.24E+01
	500	1.33E+02	9.62E+01	1.05E+02	1.04E+02
	1000	1.55E+02	1.02E+02	1.17E+02	1.19E+02
$f3$	100	9.23E-01	2.49E+00	1.63E+00	1.34E+02
	500	3.80E+00	2.02E+01	1.20E+00	7.55E+02
	1000	8.64E+00	4.62E+01	5.34E+01	1.57E+03
$f4$	100	0.00E+00	0.00E+00	1.90E+00	0.00E+00
	500	1.30E+01	2.00E+00	2.11E+01	0.00E+00
	1000	3.32E+01	1.02E+01	5.10E+01	7.00E-03
$f5$	100	0.00E+00	0.00E+00	0.00E+00	0.00E+00
	500	0.00E+00	0.00E+00	0.00E+00	0.00E+00
	1000	0.00E+00	0.00E+00	2.00E-03	0.00E+00
$f6$	100	0.00E+00	0.00E+00	0.00E+00	0.00E+00
	500	0.00E+00	0.00E+00	0.00E+00	0.00E+00
	1000	0.00E+00	0.00E+00	0.00E+00	3.00E-03
$f7$	100	-9.74E+02	-1.00E+03	-1.06E+03	-1.04E+03
	500	-5.71E+03	-5.70E+03	-5.88E+03	-5.83E+03
	1000	-1.19E+04	-1.22E+04	-1.20E+04	-1.23E+04

**Fig. 2.** Boxplot of number of evaluations to find the best solutions for each dimension.

the Figure 2 presents box-plots that show the distribution of the evaluations to find the best solution observed for each studied ABC variant on the 7 benchmark functions. On the basis of the results shown in the Figure 2, an important difference between ABC_{LX} and the rest of the algorithms is observed. ABC_{LX}

needs a small computational effort, i.e. the lowest number of evaluations to find its best solution. But when the error values from Table 2 are considered, ABC_{LX} presents a poor performance to solve the CEC' 2008 benchmark functions in comparison with the result quality obtained by ABC_{AX} , ABC_{TRAD} , and ABC_{MMAX} . Following with the analysis of the number of evaluations to find the best solutions, the results of the post-hoc tests remark statistically significant differences between ABC_{TRAD} and both ABC_{LX} and ABC_{MMAX} , and between ABC_{AX} and ABC_{LX} for 100D and 500D. In the case of 1000D, the difference between ABC_{LX} and ABC_{MMAX} is also presented.

Summarizing, the ABC_{AX} algorithm obtains the best tradeoff between the solution quality and the effort to find the best result. As a consequence, these results suggest that practitioners developing ABC-based solutions for applied optimization could adopt more efficient choices to generate new food positions, at the employed bee step. These choices can be based on standard recombination operators, such as the AX operator.

6 Conclusions

This article analysis the effect of considering other methods to generate a candidate food position in the artificial bee colony algorithm. For that purpose, an experimental evaluation of six recombination operators used to generate new solutions is carried out to solve the CEC' 2008 benchmark functions. These operators are taken from the real-coded genetic algorithm literature and they are used for the first time in the ABC framework.

The results obtained in these experiments suggest the use of the Arithmetical recombination as an interesting alternative for the traditional method to create new food positions. Since this operator presents a very good performance in terms of the average quality rank and the computational effort for all considered functions and dimensions. Furthermore, the performance of the use of the remaining operators that adjust variables in the solutions (Linear and and Max-Min-Arithmetical recombination operators) is similar to the behavior of the traditional algorithm. Instead the ABC algorithms, which apply Binomial, One-Point, and Multi-Point recombination operators and only copy parts of other solutions to create a new one, seems no good alternatives to be used in the ABC framework because of the low quality of their solutions for any function and dimension.

As future research line, we will experiment with combinatorial problems to extend the study presented in this work. These kind of problems has different solution representations, imposing new methods to generate food positions. Another future line is related with the parallelization of the ABC algorithm using the current parallel programming models.

Acknowledgments

The authors acknowledge the support of Universidad Nacional de La Pampa and the Incentive Program from MINCyT. The second author is also funded by CONICET.

References

1. V. Tereshko and A. Loengarov, "Collective decision-making in honey bee foraging dynamics," *Computing and Information Systems Journal*, ISSN 1352-9404, 2005.
2. D. Karaboga and B. Basturk, "A powerful and efficient algorithm for numerical function optimization: Artificial bee colony (abc) algorithm," *J. of Global Optimization*, vol. 39, no. 3, pp. 459–471, Nov. 2007.
3. D. Karaboga and B. Akay, "A comparative study of artificial bee colony algorithm," *Applied Mathematics and Computation*, vol. 214, no. 1, pp. 108 – 132, 2009.
4. S. Kockanat and N. Karaboga, "Parameter tuning of artificial bee colony algorithm for gaussian noise elimination on digital images," in *2013 IEEE INISTA*, 2013, pp. 1–4.
5. B. Akay and D. Karaboga, "Parameter tuning for the artificial bee colony algorithm," in *Proceedings of the 1st International Conference on Computational Collective Intelligence. Semantic Web, Social Networks and Multiagent Systems*, ser. ICCCI '09. Springer-Verlag, 2009, pp. 608–619.
6. B. Alatas, "Chaotic bee colony algorithms for global numerical optimization," *Expert Systems with Applications*, vol. 37, no. 8, pp. 5682 – 5687, 2010.
7. W. Gao, S. Liu, and L. Huang, "A global best artificial bee colony algorithm for global optimization," *Journal of Computational and Applied Mathematics*, vol. 236, no. 11, pp. 2741 – 2753, 2012.
8. D. Aydın, T. Liao, M. A. Montes de Oca, and T. Stützle, *Improving Performance via Population Growth and Local Search: The Case of the Artificial Bee Colony Algorithm*. Springer Berlin Heidelberg, 2012, pp. 85–96.
9. K. Diwold, A. Aderhold, A. Scheidler, and M. Middendorf, "Performance evaluation of artificial bee colony optimization and new selection schemes," *Memetic Computing*, vol. 3, no. 3, p. 149, 2011.
10. B. Akay and D. Karaboga, "A modified artificial bee colony algorithm for real-parameter optimization," *Information Science*, vol. 192, pp. 120–142, 2012.
11. A. Banharsakun, T. Achalakul, and B. Sirinaovakul, "The best-so-far selection in artificial bee colony algorithm," *Applied Soft Computing*, vol. 11, no. 2, pp. 2888 – 2901, 2011, the Impact of Soft Computing for the Progress of Artificial Intelligence.
12. M.-D. Zhang, Z.-H. Zhan, J.-J. Li, and J. Zhang, *Tournament Selection Based Artificial Bee Colony Algorithm with Elitist Strategy*. Cham: Springer International Publishing, 2014, pp. 387–396.
13. F. Herrera, M. Lozano, and A. Sánchez, "A taxonomy for the crossover operator for real-coded genetic algorithms: An experimental study," *International Journal of Intelligent Systems*, vol. 18, no. 3, pp. 309–338, 2003.
14. A. H. Wright, "Genetic algorithms for real parameter optimization," in *Foundations of Genetic Algorithms*. Morgan Kaufmann, 1991, pp. 205–218.
15. G. Syswerda, "Uniform crossover in genetic algorithms," in *Proceedings of the 3rd International Conference on Genetic Algorithms*. Morgan Kaufmann Publishers Inc., 1989, pp. 2–9.
16. L. J. Eshelman, R. A. Caruana, and J. D. Schaffer, "Biases in the crossover landscape," in *Proceedings of the Third International Conference on Genetic Algorithms*. Morgan Kaufmann Publishers Inc., 1989, pp. 10–19.
17. K. Tang, X. Yao, P. N. Suganthan, C. MacNish, Y. P. Chen, C. M. Chen, , and Z. Yang, "Benchmark functions for the CEC'2008 special session and competition on large scale global optimization," *Nature Inspired Computation and Applications Laboratory, USTC, China, Technical Report*, 2007.

Learning *When* to Classify for Early Text Classification

Juan Martín Loyola¹, Marcelo Luis Errecalde¹, Hugo Jair Escalante², and
Manuel Montes y Gomez²

¹ LIDIC - Universidad Nacional de San Luis, Argentina, ² INAOE - Puebla - Mexico

Abstract. The problem of classification in supervised learning is a widely studied one. Nonetheless, there are scenarios that received little attention despite its applicability. One of such scenarios is *early text classification*, where one needs to know the category of a document as soon as possible. The importance of this variant of the classification problem is evident in tasks like sexual predator detection, where one wants to identify an offender as early as possible. This paper presents a framework for early text classification which highlights the two main pieces involved in this problem: classification with *partial information* and deciding the *moment* of classification. In this context, a novel approach that learns the second component (*when* classify) and an adaptation of a temporal measurement for multi-class problems are introduced. Results with a classical text classification corpus in comparison against a model that reads the entire documents confirm the feasibility of our approach.

Keywords: Early text classification; classification with partial information; decision of the moment of classification

1 Introduction

Recent years have shown a tremendous growth in the machine learning field, solving very complex tasks with new algorithms, methods or architectures [1]. There are, however, settings of the classification problem that have received little attention despite its wide applicability. One of such scenarios is that of *early text classification* (ETC), which deals with the development of predictive models that can determine the class a document belongs to as soon as possible. Here a document is assumed to be processed sequentially, starting at the beginning and reading its containing parts one by one. In this context, it is desired to make predictions with as little information (as soon) as possible.

To date, only a few papers have approached this kind of scenarios [2, 3, 5]. Despite its low popularity, this topic has a major potential in practical applications. For instance, consider the problem of detecting sexual predators in chat conversations. Here, the goal is to sequentially read a conversation and to determine as fast as possible whenever a sexual predator is involved; clearly, a detection using the whole conversation can only be used for forensics rather than for prevention. Other potential applications include any kind of conversation analysis that

requires of a fast response (e.g., cyber-bullying prevention, detection of early traces of depression, suicidal speech identification) and classification processes with time constraint issues, where late classification implies a monetary cost (e.g., a real-time system where one might need to classify a document without processing it completely to give the user a fast response, otherwise the user might leave the site).

It is important to note that the early text classification problem consists of two related and complementary tasks. On the one hand, the task of *classification with partial information* (CPI), which consists of obtaining an efficient predictive model when only partial information is available that has been read sequentially up to a certain point in time. The emphasis in this case is to determine which classification methods are more likely to achieve performance comparable to that obtained when classified using the entire document. On the other hand, we have the task of *decision of the moment of classification* (DMC), that is, in which point in time one can stop reading and classify with some degree of confidence that the prediction is going to be correct. Both tasks need to be consistently integrated into any system for the ETC problem. However, as we will see in the related work section, little efforts have been dedicated to comprehensive approaches that simultaneously address them.

This paper addresses that previous research gap by presenting a simple framework for ETC which explicitly models the CPI and DMC components. In our proposal, the CPI component is learned with standard machine learning algorithms as in other works in ETC. The novelty of our approach consists in also learning the DMC component given an initial dataset. Evaluations of the ETC systems were carried out with standard classification measures and also with others that take into account the time dimension. In this context, another contribution of our work is the adaptation of a previous temporal evaluation metric for multi-class classification. Experimental results of our approach in a classical text classification corpus show the feasibility of the proposal for the ETC task.

The remainder of this paper is organized as follows. Next section reviews related work on early text classification. Then, Section 3 describes the framework and shows some evaluation metrics that consider the time of classification. Section 4 reports experimental results that indicate the effectiveness of the proposal. Section 5 presents conclusions and discusses future work directions.

2 Related Work

As far as we know, the whole ETC problem was initially approached in [2]; although the focus was not on making predictions earlier but on improving the classification performance with a sequential reading approach. Here, the authors process documents in a sentence-level basis. They proposed a Markov decision process (MDP), where two possible actions were allowed: read next sentence, or classify. A classifier is trained to learn good/bad state-action pairs on a high-dimensional space. In our opinion, the main drawback of this approach could be the well-known scalability problems that MDPs models are prone to suffer.

In [3] the authors propose an adaptation of Naïve Bayes to tackle the problem of classification with partial information. Although they achieve similar performance to state of the art models that read the entire document, they do not approach the DMC problem. Our work starts from this limitation and tries to solve this issue.

Recently, in [5] the CPI and DMC aspects are both addressed by learning the CPI component and using a simple heuristic rule for DMC that consists in classifying a text as positive when exceeding a specific confidence threshold in the prediction of the classifier. The problem with that DMC approach is that is very dependent on the problem and put all the burden of selecting the appropriate thresholds on the ETC system's implementer.

Nevertheless, if we consider the problem of early classification in general (not restricting it to text), we can find different groups that have tackled the problem. For example, in [4], the authors formalized the problem of early classification of time series as a sequential decision problem involving two costs: quality and delay of the prediction. Their method also provides the estimated time for classification, that is, how much of the remaining time series is needed to classify.

3 Early Text Classification Framework

Early text classification focus on the development of predictive models that determine the category of a document as soon as possible. It is assumed that the documents are read sequentially, starting at the beginning of the document and reading words in the order they appear. The objective is to predict the class of a document with as little information (as soon) as possible.

In an abstract way, it is like the classic text categorization problem, that is, for a given document it is classified under the class that best fit what had been seen in the training phase. However, differences appear when we want to measure performance. While for classic text categorization problems we use measures like accuracy, precision, recall and the F_1 measure, for early text classification those are not enough. We need measurements that consider the *time* of (*delay* in) the prediction.

This need for temporal evaluation remarks two problems that are related and complementary to each other: classification with partial information and decision of the moment of classification. Our framework defines the way the initial corpus should be divided to train and test this task. The initial corpus is divided into a training and test sets, the first one to train the early text classifier and the second one to test it. We will denote the training set as Tr and the test set as Te . What follows describes the construction of the corpus and classifiers for the different parts of the problem.

3.1 Classification with Partial Information

The task of classification with partial information consists in obtaining an effective predictive model that predicts the class of a document when only partial

information sequentially read up to a certain point of time is available. To achieve this, it is necessary to evaluate the performance of the model on partial documents. It should be noted, however, that when it comes to training the model, the entire document is used.

Given the training set Tr , we partition it in a training set for CPI, denoted TrP , and a test set for CPI, denoted TeP . Since we want to evaluate the performance on partial documents, we must modify the test set for CPI. To achieve this, it is necessary to define a window value $w \in \mathbb{N}$, which indicates the number of terms that are read in each time step. In this way, if $d_j = (t_1, \dots, t_r)$ is the j -th document in TeP with r the number of terms in the document d_j and t_1, \dots, t_r the terms in the order they appear in it, then the documents $d_{j,1} = (t_1, \dots, t_w)$, $d_{j,2} = (t_1, \dots, t_{2w})$, \dots and $d_{j,\frac{r}{w}} = (t_1, \dots, t_{\frac{r}{w}w})$ are part of the pumped test set TeP' . This process is repeated for all the document in TeP . For example, given the document “Do not look a gift horse in the mouth” from the test set TeP and $w = 3$, then “Do not look”, “Do not look a gift horse” and “Do not look a gift horse in the mouth” will belong to TeP' .

Once the training and tests sets were constructed, we apply machine learning procedures to get a good representation and to train the model. Nonetheless, we need to adapt the evaluation method to consider the performance of the model as the reading of the partial documents proceed. For this purpose, we build subsets of TeP' with the same number of documents as the initial TeP test set. The subset TeP'_t is defined as all partial documents of TeP where the length (measure in number of terms) is less than or equal to $w \cdot t$. Also, if there exists $d_{j,l}$ and $d_{j,l+v}$ partial documents of d_j with l multiple of w and $l < l + w \leq w \cdot t$, only the largest partial document will belong to TeP'_t , in this case $d_{j,l+v}$. Thus, the model performance is calculated for the different TeP'_t subsets by evaluating it as the terms of the windows are read.

3.2 Context Information

Trying to decide when to stop reading a document only using the class the CPI model returns is difficult. For this reason, we augment the data the DMC model gets with context information, that is, data from the body of the document that could be helpful for deciding the moment of classification. We propose to obtain these from three different sources:

- *Current document*: characteristics relative to the content of the current document. For example, number of: terms, different terms, most relevant terms for each class, stop words, etcetera.
- *Output from CPI*: features produced by the CPI model. This can be the class predicted, current window number and additional information from the model (this depends on the type of model picked for CPI). Regarding the latter, in the case of probabilistic classifiers we can have the probability assigned to each class.
- *Historic data*: related to the context information obtained in previous windows. That is, we apply an aggregation function δ (average, max, min, count,

or other function) to some context information from previous window. An example of historic information is the average probability given to all classes by the CPI model.

The features that in the end are provided to the DMC model depend on the problem under consideration and which are estimated to be more informative to this decision problem. The context information is calculated only for the test set TeP' since the training and test sets for DMC are constructed from it.

3.3 Decision of the Moment of Classification

The task of decision of the moment of classification is to determine the point at which the reading of the document can be stopped with some certainty that the prediction of the classification made is correct. Given a feature vector, this model predicts if we should stop or keep reading. It is expected that when the model for DMC decides to stop reading, the class predicted by the model for CPI is the right one.

The initial corpus for DMC is constructed based in the context information, that is, the training and test sets are formed by the feature vectors generated based in the partial documents of TeP' . The problem here is that we do not have the labels indicating when to stop reading. They must be manually obtained or an automatic process should be devised to do it. Here we propose an automatic way to label the feature vector: if the category chosen by CPI is the correct one then the reading can be stopped; if, however, this is not correct we should keep reading. Finally, the corpus is divided into a training set TrM and a test set TeM . The construction and evaluation of the DMC model does not present any particularity, reason why any machine learning technique could be applied.

3.4 Architecture

Once the CPI and DMC models and the context information procedure have been defined, we can formalize the final architecture for our ETC framework. Figure 1 shows the role every model fulfills: CPI is responsible for predicting the category of the partial document, the context information procedure builds the feature vector and, finally, DMC is the one in charge of making the decision of the moment in which the reading of the document must be stopped.

When classifying using this architecture, a document follows the next steps:

1. Read w contiguous terms;
2. Build the vector representation of the partial document for CPI;
3. Classify the partial document with CPI;
4. Build the feature vector for DMC;
5. Classify using DMC;
6. If DMC suggests keeping reading terms, return to point 1;
7. If DMC suggests stopping reading then return the category chosen by CPI for the partial document.

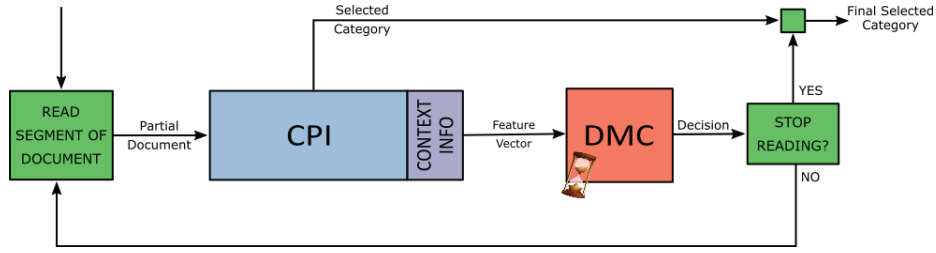


Fig. 1. Early Text Classification Architecture.

3.5 Evaluation Metric

Since the ETC problem has not been addressed for many of the possible configurations, many aspects of it have not been yet defined. Among these is the evaluation of the model, since there is no measure to evaluate the temporary performance of it in a multi-class context. There exists, nonetheless, an evaluation metric for binary early classification [5] that considers the accuracy of the prediction and the delay taken by the system to make the decision. Here the delay is measured by counting the number of terms seen before giving the answer. Given a decision d made by the model at time k , the early risk detection error (ERDE) is defined as:

$$\text{ERDE}_o(d, k) = \begin{cases} c_{fp} & \text{when the decision } d \text{ is positive and it is incorrect} \\ c_{fn} & \text{when the decision } d \text{ is negative and it is incorrect} \\ lc_o(k) \cdot c_{tp} & \text{when the decision } d \text{ is positive and it is correct} \\ 0 & \text{when the decision } d \text{ is negative and it is correct} \end{cases}$$

The values given to c_{fp} and c_{fn} depends on the application domain and the implications of false positives and false negatives decisions. The factor $lc_o(k) \in [0, 1]$ encodes a cost associated to the delay in detecting true positives. In domains where late detection has severe consequences we should set c_{tp} to c_{fn} , that is, late detection is equivalent to not detecting the case at all. The function $lc_o(k)$ should be a monotonically increasing function of k . Losada and Crestani suggest:

$$lc_o(k) = 1 - \frac{1}{1 + e^{k-o}}$$

This function is parametrized by o , which controls the point where the cost grows more quickly. The overall error will be the mean of the ERDE values for all the documents.

Based on this metric, we propose a generalization for ETC when there are more than two classes. ERDE is redefined for each class as:

$$\text{ERDE}_o(d, k, i) = \begin{cases} c_{fp}^i & \text{if it is a } FP_i \\ c_{fn}^i & \text{if it is a } FN_i \\ lc_o(k) \cdot c_{tp}^i & \text{if it is a } TP_i \\ 0 & \text{if it is a } TN_i \end{cases}$$

where i indicates the category on which the error is being calculated, d represents the decision made for this category and k the time when the decision is made. The constants c_{fp}^i , c_{fn}^i and c_{tp}^i indicate the cost associated with the decision on the category being false positive, false negative or true positive, respectively. The values given to these constants depend on the particular addressed problem. Function $lc_o(k)$ is defined as before.

A problem with this function is that it does not consider the length of the document. It does not make sense to have one fixed point of penalization when the documents have very dissimilar lengths. For example, if the corpus contains papers and books, a fixed penalization point will harm one of them depending if it is small or big. We proposed an alternative function to tackle this problem:

$$lc_o(k) = \frac{k}{o}$$

where o represents the length of the document measured in number of terms and k represents the number of terms read at the time of stopping the reading.

Then the early detection error (EDE) for a document is given by the sum of the ERDE for all categories. That is:

$$EDE_o(d, k) = \sum_{i=1}^{|C|} ERDE_o(d, k, i)$$

Since only one category can be chosen by the model (single label problem) and the cost associated with true negatives is zero then the early detection error can be reduced to:

$$EDE_o(d, k) = \begin{cases} lc_o(k) \cdot c_{tp}^i & \text{when the category } i \text{ chosen by CPI is correct} \\ c_{fn}^j + c_{fp}^l & \text{when the category } j \text{ chosen by CPI is incorrect} \\ & \text{and the category } l \text{ was correct.} \end{cases}$$

Overall early detection error is obtained by averaging on all documents.

4 Experiments and Results

To test the feasibility of our approach, we used in the experiments the well-known dataset R8 [6]. Based on the Reuters-21578 collection, R8 contains documents belonging to the eight classes with the highest number of training documents in that collection. These documents belong to only one class, thus allowing the corpus to be used for single-label text categorization. More detailed information of R8 is shown in Table 1.

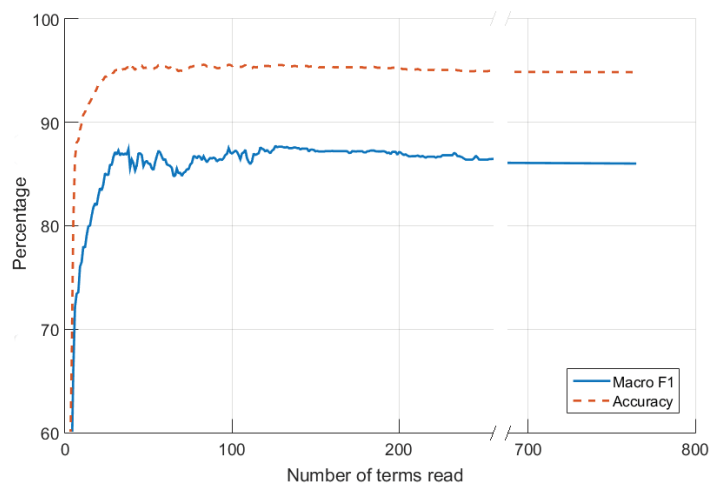
For the experiments, the corpus was processed as follows: a bag-of-words representation of the documents was obtained using the TMG toolbox with a term-frequency (tf) weighting scheme [7]. Then, we split the corpus in training and test set for the early text classification model in general and CPI. The

Table 1. Composition of the corpus R8.

Class	# train doc	# test doc	total # doc
acq	1596	696	2292
crude	253	121	374
earn	2840	1083	3923
grain	41	10	51
interest	190	81	271
money-fx	206	87	293
ship	108	36	144
trade	251	75	326
Total	5485	2189	7674

window size chosen was $w = 3$, that is, three terms were read between each run of the early text classification framework.

Based on Escalante's work [3] we chose a naïve Bayes classifier for the CPI model. The performance for the partial documents can be seen in Fig 2. Clearly, we can classify documents without reading all terms.

**Fig. 2.** Evaluation of the sets TeP_t .

Next, we needed to find out what features we could extract to help decide when to classify. In the present work, the selected features were:

- Number of terms of the partial document.
- Number of distinct terms of the partial document.
- Number of relevant terms for each class¹.

¹ In this work, we used the most frequent terms in each class.

- Class score given by the CPI model.
- Current window number.
- Historic data from previous windows. For each class we calculated the mean class score given by the CPI model in previous windows.

Finally, we used these features from the documents to build the training and test sets for the DMC model. We tested models trained with three different approaches provided by MatLab: *Naïve Bayes*, *k-ridge*, and an ensemble method (*GentleBoost*) which used neural networks as weak classifiers. The performance of each classifier is shown in Table 2. From these results, we chose the GentleBoost classifier for the DMC model.

Table 2. Final estimation results for the precision, recall and F_1 measure.

Predictive DMC Model	Precision Estimation	Recall Estimation	F_1 Measure
Kridge	54.46%	57.52%	55.94
Naïve Bayes	56.19%	56.79%	56.48
Gentleboost	60.12%	78.87%	68.23

Having trained the CPI and DMC models, we had all the pieces to test our framework. The final results are shown in Table 3, where EDE_{10}^1 is the early detection error using the definition of $lc_o(k)$ that does not consider the length of the document with $o = 10$, and EDE^2 is the one that considers the length of the document. It can be observed that, for the F_1 measure (measure not considering time), a standard (full reading) model obtains a better result than the temporal one, although the result of our approach is still acceptable. However, when evaluating the temporal aspects, the advantages of the proposal presented in this work are evident: with respect to EDE_{10}^1 , a reduction of 0.73 to 0.57 is achieved while for the EDE^2 error the reduction is of 1.05 to 0.73. It can also be observed that there is an average saving of 41.21 terms from the temporal versus the standard approach. Considering the average size of the documents in the R8 collection is 150 terms, this is a significant number of terms that are saved (28% of terms in average).

Table 3. Final results of the linear against the temporal model.

Type of Model	F_1 Measure	Average Unread Terms	EDE_{10}^1	EDE^2
Standard	85.97	0	0.73	1.05
Temporal (ETC architecture)	78.99	41.21	0.57	0.73

5 Conclusions

Early text classification has not been yet studied in depth, but numerous new applications in on-line detection and real-time systems give a new impetus to research works in this area. This trend has been evident this year where the first conference directly related to ETC will be proximately held in Dublin (<http://early.irlab.org/>).

In this paper, we have formalized a simple framework that makes explicit two critical parts in ETC systems: (i) the classification with partial information, and (ii) the decision of the moment of classification. In this context, a novel approach that learns the second component is proposed and a new measure for evaluating multi-class ETC problems is defined.

While promising results were obtained with the R8 corpus reducing considerably the number of terms needed for classification, several directions for improvement are easily identified for future work. First of all, we can boost up the basic implementation used in the present work by augmenting the contextual information of the DMC model with other more informative features (for instance, as relevant words, using the words with the highest information gain). Furthermore, different document representations and predictive models should be tested for CPI and DMC. Finally, we should test this framework in a more recent and competitive early classification corpus like the one presented by Losada and Crestiani [5] and also on other data sets where ETC approaches can be critical like the detection of sexual predators in chats or detection of suicidal discourse.

References

1. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Advances In Neural Information Processing Systems*, pp. 1–9, 2012.
2. G. Dulac-Arnold, L. Denoyer, and P. Gallinari, "Text classification: A sequential reading approach," in *Advances in Information Retrieval, Proc. of 33rd European Conf. on IR Research, (ECIR'11)*, vol. 6611 of *LNCS*, pp. 411–423, Springer, 2011.
3. H. J. Escalante, M. Montes-y-Gómez, L. V. Pineda, and M. L. Errecalde, "Early text classification: a naïve solution," in *Proc. of the WASSA@NAACL-HLT 2016, June 16, 2016, San Diego, California, USA*, pp. 91–99, 2016.
4. A. Dachraoui, A. Bondu, and A. Cornuéjols, "Early classification of time series as a non myopic sequential decision making problem," *LNCS*, vol. 9284, pp. 433–447, 2015.
5. D. E. Losada and F. Crestani, *A Test Collection for Research on Depression and Language Use*, pp. 28–39. Cham: Springer International Publishing, 2016.
6. A. Cardoso-Cachopo, "Improving Methods for Single-label Text Categorization." PdD Thesis, Instituto Superior Tecnico, Universidade Tecnica de Lisboa, 2007.
7. D. Zeimpekis and E. Gallopoulos, *TMG: A MATLAB Toolbox for Generating Term-Document Matrices from Text Collections*, pp. 187–210. Springer, 2006.

Optimización de un Método de Reducción de Incertidumbre Aplicado a Incendios Forestales

María Laura Tardivo^{1,2,3}, Paola Caymes-Scutari^{1,2}, Miguel Méndez-Garabetti^{1,2} y Germán Bianchini¹

¹ Laboratorio de Investigación en Cómputo Paralelo/Distribuido (LICPaD), Dpto. de Ingeniería en Sistemas de Información - UTN-FRM - Mendoza - Argentina.

² Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)

³ Departamento de Computación - UNRC- Córdoba - Argentina.

Resumen La predicción del comportamiento de los incendios forestales representa un gran desafío desde el punto de vista matemático y computacional. La dificultad radica en la imposibilidad de medir en tiempo real todos los parámetros que determinan el comportamiento del incendio. ESSIM (*Evolutionary Statistical System with Island Model*), es un método de reducción de incertidumbre que utiliza estadística, computación de alto rendimiento y estrategias evolutivas para orientar la búsqueda hacia mejores soluciones. ESSIM ha sido implementado con dos estrategias de búsqueda: el método ESSIM-EA utiliza como método de optimización a los Algoritmos Evolutivos, por su parte ESSIM-DE utiliza el algoritmo Evolución Diferencial. ESSIM-EA ha demostrado obtener buena calidad de aproximaciones, mientras que ESSIM-DE obtiene mejores tiempos de respuesta. En este trabajo se presenta una alternativa para mejorar la calidad de las soluciones encontradas por ESSIM-DE, a partir del análisis de la velocidad de convergencia de la estrategia evolutiva y su relación con la distribución de la población al inicio de cada paso de predicción.

Keywords: Incendios Forestales, Predicción, Modelo de Islas, Algoritmos Evolutivos, Evolución Diferencial, Paralelismo.

1. Introducción

La predicción de fenómenos naturales, como los incendios forestales, se considera una tarea muy importante que implica un alto grado de complejidad y precisión. Generalmente, las herramientas de simulación implementan modelos que intentan explicar y predecir la propagación del fuego sobre el terreno. Estos modelos deben ser alimentados por todos los parámetros de entrada necesarios para representar los diversos factores dinámicos que determinan el comportamiento del fuego. Sin embargo, es poco probable disponer con anticipación de valores precisos para todos estos factores, lo cual afecta negativamente la calidad de la predicción. Por lo tanto, el desafío radica en el desarrollo de métodos computacionales que puedan mejorar el conocimiento

sobre los valores de los parámetros de entrada y reducir la incertidumbre sobre dichos valores, con el fin de realizar predicciones más realistas.

El Sistema Estadístico Evolutivo con Modelo de Islas (*Evolutionary Statistical System with Island Model, ESSIM*) es un método de reducción de incertidumbre que ha sido aplicado a la predicción de incendios forestales. ESSIM utiliza estrategias evolutivas para guiar el proceso de búsqueda hacia soluciones de calidad, un componente estadístico para analizar la tendencia del fuego, y evaluación en paralelo para reducir el tiempo de cómputo. Su primer diseño computacional incorporó como estrategia de búsqueda a los Algoritmos Evolutivos, y el método pasó a denominarse ESSIM-EA. Posteriormente, se utilizó la metaheurística Evolución Diferencial, dando surgimiento al método ESSIM-DE. ESSIM-EA ha demostrado obtener buena calidad de predicción, por su parte ESSIM-DE reduce significativamente los tiempos de respuesta. Ambos métodos han sido estudiados en los últimos años con el objetivo de mejorar su desempeño. En [1] y [2] se realizaron dos estudios de sintonización estática de los parámetros evolutivos de ESSIM-EA. En [3] y [4] se realizaron dos estudios de calibración estática para ESSIM-DE, que involucraron los parámetros relativos al modelo de islas y los parámetros evolutivos, respectivamente. En estos últimos dos estudios se encontraron mejoras en algunos pasos de predicción, sin embargo, no fueron demasiado significativas. Los estudios antes mencionados están relacionados a la calibración de parámetros de control. En este trabajo se presenta una propuesta para mejorar la calidad obtenida por ESSIM-DE a partir del análisis de la distribución de la población al inicio de cada paso de predicción. El objetivo es determinar la relación existente entre la velocidad de convergencia de la metaheurística que guía el proceso de búsqueda, respecto de la calidad de soluciones encontradas en los diferentes instantes de tiempo del desarrollo del incendio. Este estudio representa el punto de partida para la obtención de un patrón de comportamiento general, que permita determinar la distribución que debe tener la población en cada paso de predicción.

El trabajo se organiza de la siguiente forma. La Sección 2 describe el método ESSIM y los métodos de búsqueda utilizados por ESSIM-EA y ESSIM-DE. La Sección 3 presenta la propuesta para mejorar la predicción de ESSIM-DE, denominada ESSIM-DE(r). La Sección 4 presenta la experimentación realizada para validar la propuesta. Finalmente, en la Sección 5 se presentan las conclusiones y los trabajos futuros.

2. Descripción general de ESSIM

ESSIM es un método de reducción de incertidumbre que posee evaluación en paralelo a través de un esquema de comunicación de doble jerarquía *Master/Worker*. En una jerarquía superior se dispone de un proceso denominado *Monitor*, que es el encargado de enviar información de inicialización, recolectar los datos procesados en la etapa final de la simulación y determinar los valores de salida. En una jerarquía inferior, los procesos de ESSIM están organizados en islas. En cada isla, un proceso *Master* es el encargado de inicializar una población

de individuos, los cuales representan diferentes combinaciones de valores para las condiciones en las que se desarrolla el incendio. La Fig. 1 describe el diagrama general de ESSIM. Una vez que el *Master* inicializa la población, comienza el ciclo evolutivo a través de la aplicación de los operadores evolutivos para generar nuevos individuos (etapa \mathbf{EO}_M). El *Master* distribuye cada individuo entre ciertos procesos *workers*, quienes realizan la Etapa de Optimización (\mathbf{EO}_W), evaluando el desempeño de cada uno a través de la aplicación de la función de aptitud, para luego retornar al *Master* la evaluación obtenida. La etapa \mathbf{EO}_W posee dos subetapas internas denominadas Simulación de Incendios (\mathbf{SI}) y la aplicación de la función de aptitud propia de la metaheurística asociada a ESSIM (\mathbf{Met}_a). \mathbf{SI} debe ser alimentada con la línea de fuego real del incendio en el instante de tiempo t_{i-1} ($LF R_{i-1}$) junto con el individuo que representa el vector de parámetros de entrada (\mathbf{VP}). Cuando \mathbf{SI} concluye con la simulación de los individuos, el resultado de cada simulación es introducido en la etapa \mathbf{Met}_a para comparar el mapa simulado con el mapa real del instante de tiempo t_i ($LF R_i$), lo que determina el valor de aptitud para cada individuo. En ESSIM, la función de aptitud está dada por la expresión:

$$aptitud = \frac{|A \cap B|}{|A \cup B|} \quad (1)$$

en donde A representa el conjunto de celdas en el mapa real sin el subconjunto de celdas quemadas antes de iniciar el proceso predictivo, y B representa el conjunto de celdas en el mapa predicho sin el subconjunto de celdas quemadas antes de iniciar la predicción. Por lo tanto, el valor de aptitud estará comprendido en el rango $[0,1]$, un valor igual a 1 representa una predicción perfecta, y un valor igual a cero indica el máximo error. La eliminación del conjunto de celdas quemadas antes de iniciar la predicción se realiza para evitar resultados sesgados. El *Master* también se encarga de efectuar la migración de individuos, seleccionado un conjunto de individuos de la población actual para ser enviado a otra isla. Los individuos que arriban reemplazan a los peores de la población destino.

Una vez que la población completa las diferentes generaciones evolutivas o alcanza cierta aptitud, es introducida en la Etapa de Calibración (\mathbf{EC}_M). En esta etapa la población que ha evolucionado alimenta a una sub-etapa denominada Etapa Estadística (\mathbf{EE}). La salida de EE es un mapa de probabilidades y es utilizado en la sub-etapa BK_{ign} (búsqueda del valor clave de ignición K_{ign}) para calcular el patrón de comportamiento del incendio. El valor hallado, K_{ign} , es utilizado para realizar la predicción para el próximo instante de tiempo (t_{i+1}). La evaluación de la aptitud del mapa de probabilidades se lleva a cabo en la etapa que implementa la Función de Aptitud (\mathbf{FA}). Además, la salida BK_{ign} del instante de tiempo anterior se combina con el mapa de probabilidades proporcionado por EE para generar la predicción en la Etapa de Predicción (\mathbf{EP}). En la etapa \mathbf{EC}_M , la salida de EE se envía a \mathbf{EC}_{MON} junto con los j valores de K_{ign} calculados por las j islas. En la Etapa \mathbf{EC}_{MON} el proceso *Monitor* selecciona el mejor valor de K_{ign} de todas las islas y lo utiliza para realizar la predicción para el próximo instante de tiempo (t_{i+1}). Para obtener

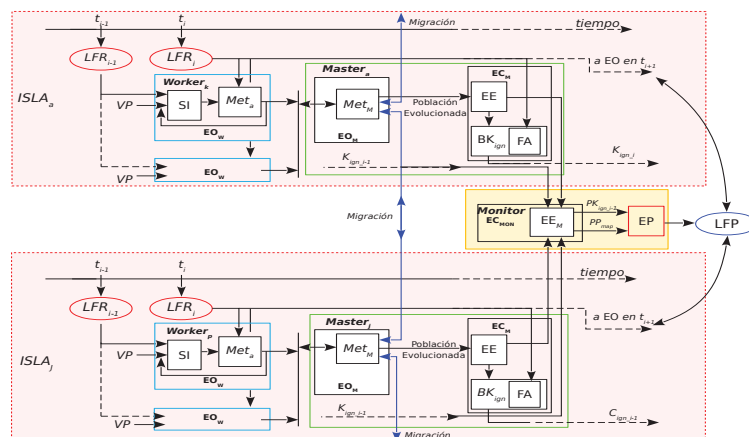


Figura 1. Diagrama General de ESSIM. **SI**: Simulador de Incendios, **Met_M**: Metaheurística (instancia *Master*); **Met_a**: Metaheurística (evaluación de aptitud); **EO**: Etapa de Optimización; **EE**: Etapa Estadística; **BK_{ign}**: Búsqueda de K_{ign} ; **K_{ign}**: valor clave de ignición; **FA**: Función de Aptitud; **EC**: Etapa de Calibración; **EP**: Etapa de Predicción; **LFP**: Línea de Fuego Propuesta; **LFR_x**: Línea de Fuego Real en el instante x ; **VP**: Vector de Parámetros de entrada; **EE_M**: Etapa Estadística Monitor; **PKign**: Valor clave de ignición pre-calculado; **PP**: Mapa de Probabilidad pre-calculado.

mayor detalle de los métodos de predicción y del funcionamiento de ESSIM puede consultarse [4] o [5].

Se puede observar que es posible instanciar el método ESSIM con diferentes estrategias de búsqueda en las etapas evolutivas. A continuación se describen brevemente dos optimizadores que han sido utilizados en ESSIM.

ESSIM-EA: Algoritmos Evolutivos: Los Algoritmos Evolutivos (EAs) se consideran métodos de búsqueda eficientes para resolver problemas de optimización, inspirados en las teorías de la creación de nuevas especies y su evolución [6]. El espacio de búsqueda se organiza como un conjunto de posibles soluciones, llamadas individuos, que constituyen una población. La población evoluciona iterativamente, a través de generaciones, imitando los principios de la evolución biológica y la supervivencia del más apto. Para ello se selecciona un subconjunto de individuos de la población que será utilizado como progenitores de la población de la siguiente generación. Estos progenitores son sometidos a diferentes operadores para generar el conjunto de descendientes. Posteriormente, los descendientes se introducen a la población reemplazando a los individuos con las peores características, es decir, los menos aptos para sobrevivir.

ESSIM-DE: Evolución Diferencial: El algoritmo de Evolución Diferencial (DE) [8] es un optimizador estocástico basado en poblaciones que utiliza diferencia de vectores para modificar cada vector de la población. En contraste con EA, los operadores de mutación, cruzamiento y selección son aplicados sobre cada uno de los individuos de la población. El operador de mutación modifica la población aplicando diferencias de vectores entre cada "individuo actual" y ciertos individuos seleccionados al azar. Posteriormente, se somete cada individuo

mutado junto con el individuo actual al operador de cruzamiento, generando un nuevo vector, llamado “vector de prueba”. Finalmente, en la etapa de selección se determina el mejor candidato entre el individuo actual y el vector de prueba. Aquel que tenga mejor valor respecto de la función a optimizar será el que sobrevivirá a la generación siguiente.

3. ESSIM-DE(r): optimización de ESSIM-DE

ESSIM-DE ha demostrado obtener una tendencia decreciente en cuanto a los promedios de aptitud alcanzados. En [5] se pudo observar que ESSIM-DE inicia con un desempeño bueno, pero a medida que el incendio transcurre, los promedios de aptitud decaen. En dichos experimentos la población es inicializada con valores aleatorios siguiendo una distribución uniforme. Este proceso se realiza una única vez, al comienzo de la simulación. Luego, la población evoluciona a través de la aplicación de los operadores evolutivos de Evolución Diferencial. La cantidad de individuos que participan en el proceso de mutación y cruzamiento depende del modo de operación de cada metaheurística. En EA, este valor está determinado por un parámetro definido por el usuario y generalmente es inferior al 50% de los individuos totales. En cambio, en Evolución Diferencial todos los individuos participan en el proceso de mutación y cruzamiento: en cada generación se recorre la población y cada individuo es mutado y cruzado con otros individuos de la población.

Teniendo presente la manera en la que cada metaheurística aplica los operadores evolutivos, se propuso analizar cuál es la distribución de la población en los diferentes pasos de predicción. La hipótesis asociada es que ESSIM-DE puede tener una convergencia prematura: la población completa evoluciona de manera acelerada en los primeros pasos de predicción, por lo cual converge a un espacio de soluciones acotado que impide obtener mejores individuos en los pasos sucesivos de la evolución del incendio. Por lo tanto, sería posible alterar dicha distribución reiniciando la población al inicio de cada paso de predicción. De esta forma, Evolución Diferencial podría explorar, en cada paso de predicción, un espacio de soluciones nuevas e independientes de las anteriores. En consecuencia, ESSIM-DE(r) fue diseñado con un operador de diversificación poblacional aplicado al inicio de cada paso de predicción, el cual **reinicia la población** con valores aleatorios dentro del rango de cada variable, siguiendo una distribución uniforme.

Para verificar la hipótesis, se propuso comparar la distribución poblacional alcanzada tanto por ESSIM-DE como por ESSIM-EA. Para ello, se sometió a ambos métodos a una experimentación utilizando diferentes casos de quemas controladas realizadas en Serra de Louçã, (Gestosa, Portugal) [7]. El tamaño de cada mapa fue definido de acuerdo a la superficie disponible, a las características del terreno y a los requerimientos del proyecto. Para cada caso se han definido intervalos de tiempo discretos que representan el avance del frente de fuego.

La Tabla 1 describe las características de los tres casos utilizados en este trabajo. Es importante destacar que en el primer instante de tiempo se encuentra

Tabla 1. Descripción de los casos de estudio: dimensión, pendiente y tiempo de inicio, fin e incremento.

Caso	Ancho (m)	Alto (m)	Pend. (grados)	T. ini. (min)	Incr. (min)	T. fin (min)
A	95	123	21	2	2	12
B	60	90	6	2	2	10
C	75	126	19	3	1	9

operando la etapa de calibración, por lo cual no se realizan predicciones. La distribución de la población obtenida con cada método se analiza mediante gráficos de caja (boxplots) con los valores de aptitud de cada uno de los 200 individuos.

Con el objetivo de facilitar la comprensión de los resultados, se seleccionaron las salidas obtenidas en una ejecución del caso de estudio C, que consta de seis pasos de simulación, y se estableció como condición de finalización alcanzar 15 generaciones evolutivas en cada paso. La Fig. 2 muestra tres gráficos correspondientes al segundo, cuarto y sexto pasos de simulación obtenidos con ESSIM-EA. Para simplificar el análisis, en cada gráfico se incluyen tres boxplots que representan la distribución de la población cada cinco iteraciones. Como se puede observar, en el segundo paso la población inicia con valores de aptitud cercanos a cero (la mediana está próxima a cero). A medida que la población evoluciona por las diferentes generaciones, la distribución se modifica levemente hacia valores superiores. Se puede observar que en la iteración 15 del sexto paso de simulación la mediana alcanza un valor cercano a 0.2 y el boxplot de distribución llega a un valor máximo de 0.6 (*whisker superior*).

La Fig. 3 muestra los gráficos correspondientes al segundo, cuarto y sexto pasos de simulación para ESSIM-DE. A diferencia del caso anterior, se puede ver claramente que al finalizar el segundo paso de simulación la población alcanza valores de aptitud cercanos a 0.8. También se observa que al finalizar las primeras cinco iteraciones en el segundo paso de simulación la población posee baja distribución, esto provoca que el método alcance su máxima capacidad de exploración, por lo tanto en los pasos sucesivos Evolución Diferencial no podrá aportar mejoras. En este ejemplo, así como en el resto de los casos considerados, se observa que ESSIM-DE converge más rápidamente que ESSIM-EA. Este comportamiento provoca una tendencia acelerada de las soluciones hacia óptimos locales en los instantes de tiempo iniciales del desarrollo del incendio. Por lo tanto, en los últimos pasos de predicción la metaheurística no puede aportar mejoras, lo que conduce a un decaimiento de la aptitud.

Como meta ideal, se pretende que los métodos desarrollados obtengan soluciones a corto plazo y que estas sean confiables, pero la velocidad de respuesta y la alta probabilidad de convergencia suelen ser objetivos contradictorios [8]. Este es el caso de ESSIM-DE, puesto que obtiene soluciones a corto plazo, con buena calidad en los primeros pasos de predicción, pero con tendencia decreciente. Con el objetivo de mejorar la calidad de la predicción, el operador de diversificación poblacional de ESSIM-DE(r) permite incorporar un nuevo espacio de soluciones cada paso de predicción. De esta manera, el esquema de procesamiento original de Evolución Diferencial no es alterado y la

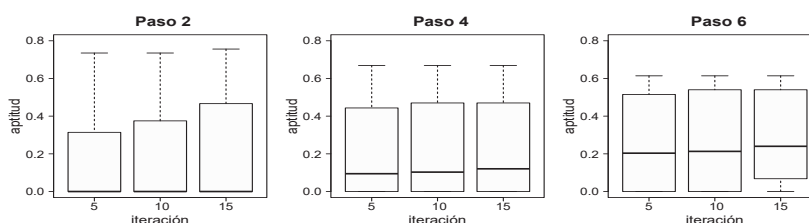


Figura 2. Distribución de la aptitud de la población, experimento C, con ESSIM-EA.

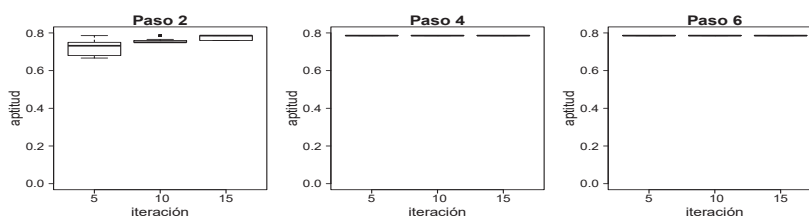


Figura 3. Distribución de la aptitud de la población, experimento C, con ESSIM-DE.

optimización de los individuos se realiza de acuerdo al paso de predicción que se esté considerando. Esta variante del método fue denominada ESSIM-DE(r), para distinguir la versión que incluye el operador de diversificación con reinicio de poblaciones, respecto de la versión original.

4. Experimentación y Resultados Obtenidos

Para validar la propuesta se realizaron experimentos con los tres casos de quemas controladas de la Tabla 1. Cada experimento fue efectuado con un total de 30 semillas diferentes, y los resultados obtenidos fueron promediados.

Los experimentos fueron configurados con los mismos parámetros para todas las características comunes relacionadas con el modelo de isla: 5 islas, 7 *workers* por isla. La configuración de migración fue establecida con un porcentaje de migración del 20% de los individuos de la población, y la migración se realiza en cada iteración. El criterio de terminación en el proceso evolutivo consiste en alcanzar un umbral de aptitud de 0,7 en cada paso. Este valor umbral se estableció teniendo en cuenta un valor de aptitud que representa una calidad aceptable de la predicción. El tamaño de cada población se definió como 200 individuos. Para ESSIM-DE y ESSIM-DE(r) se utilizó la misma configuración de los parámetros evolutivos: probabilidad de cruzamiento 0.3, factor de mutación 0.9, cruzamiento binomial.

La Fig. 4 muestra los resultados obtenidos. Los gráficos superiores de la figura corresponden a los promedios de los valores de aptitud obtenidos para cada paso de predicción y método, de los diferentes experimentos. El eje x de cada gráfico representa los diferentes pasos de predicción y el eje y el promedio de los valores de aptitud que resulta de evaluar el mapa obtenido de cada predicción respecto

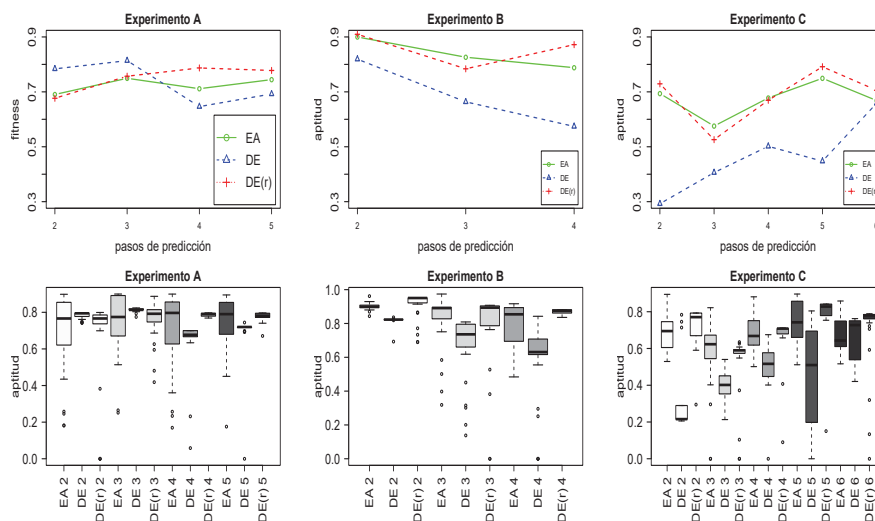


Figura 4. Experimentos A, B y C. **Arriba:** Promedios de aptitud para cada paso de predicción y método. **Abajo:** Distribución de la aptitud para cada paso de predicción y método.

del mapa real del avance del incendio. Para simplificar la notación, ESSIM-EA está representado por “EA”, ESSIM-DE por “DE” y el método ESSIM-DE con la propuesta de mejora se representa “DE(r)”. Los gráficos inferiores de la Fig. 4 representan la distribución de los 30 resultados obtenidos para cada semilla y en cada experimento, agrupados por colores (escala de grises) para cada método y paso de simulación.

Experimento A: En la Fig. 4, se puede apreciar del gráfico superior para el experimento A, que ESSIM-DE obtiene los mejores valores de aptitud para los dos primeros pasos de predicción. Sin embargo, la aptitud de ESSIM-DE decae a medida que el incendio avanza. Por su parte, ESSIM-DE(r) supera a ESSIM-EA en los pasos de predicción 3, 4 y 5 con valores de aptitud promedio cercanos a 0.8. También se observa, a partir del gráfico inferior para el experimento A, una baja distribución de los resultados de aptitud obtenidos por ESSIM-DE(r). Esta propiedad es deseable para los métodos que operan con metaheurísticas, puesto que es un indicador de la robustez del método al resolver de manera efectiva diferentes instancias del problema [6]. Por ambas características podemos concluir en este experimento que ESSIM-DE(r) es el método con mejor desempeño.

Experimento B: El experimento B consta de cuatro pasos de simulación. Se puede observar a partir del análisis del gráfico superior de la Fig. 4, para el experimento B, que ESSIM-DE y ESSIM-EA tienen una tendencia decreciente para los promedios de aptitud. Sin embargo este comportamiento no se refleja en el paso 4 para ESSIM-DE(r), el cual obtiene un valor promedio cercano a 0.9 (recordemos que un valor de aptitud 1 equivale a una predicción perfecta). Más aún, ESSIM-DE(r) obtiene muy baja distribución de los resultados en el cuarto

Tabla 2. Tiempo de ejecución promedio de cada método

Experimento	ESSIM-EA	ESSIM-DE	ESSIM-DE(r)
A	01:01:15	00:49:05	00:54:20
B	00:50:10	00:27:49	00:32:48
C	02:11:38	00:41:20	00:25:53

paso (ver gráfico inferior de la Fig. 4, para el experimento B), lo cual indica un excelente desempeño a nivel general.

Experimento C: Se puede observar a partir del gráfico superior de la Fig. 4, para el experimento C, que ESSIM-DE(r) mejora de manera significativa los promedios de aptitud obtenidos respecto de ESSIM-DE, y obtiene valores promedio superiores a ESSIM-EA para los pasos de simulación 2, 5 y 6. También es posible evidenciar a partir de los boxplots del gráfico inferior de la Fig. 4, para el experimento C, que ESSIM-DE(r) obtiene menor distribución de los resultados para dichos pasos. Por lo tanto, nuevamente se corrobora que la propuesta ESSIM-DE(r) es efectiva, mejorando el desempeño de la versión original de ESSIM-DE y superando los resultados obtenidos con ESSIM-EA.

La Tabla 2 muestra los tiempos de ejecución promedio obtenidos por los tres métodos, para cada experimento. En el caso A y B, la propuesta implementada con ESSIM-DE(r) provoca que la convergencia sea más lenta, lo que conduce a obtener tiempos de ejecución promedio mayores que la versión ESSIM-DE. Sin embargo, cualquiera de las dos versiones de ESSIM-DE obtienen mejor tiempo de respuesta que ESSIM-EA, reduciendo los tiempos hasta en un 80 %.

A partir del análisis realizado podemos verificar que existe una relación directa entre la capacidad de convergencia de la metaheurística utilizada como motor de búsqueda y la calidad de las predicciones. La propuesta de mejora ha demostrado ser efectiva para los casos de prueba considerados: ESSIM-DE(r) ha mejorado la calidad de las soluciones encontradas, en comparación con los resultados obtenidos tanto por ESSIM-EA como por la versión original de ESSIM-DE, con baja distribución de los resultados. Los tiempos de ejecución para ESSIM-DE(r) son similares e incluso menores a la versión original de ESSIM-DE. Todo esto indica que la propuesta de mejora ha sido efectiva para los experimentos considerados.

5. Conclusiones

En este trabajo se presentó una alternativa para mejorar la calidad de las soluciones encontradas por el método ESSIM-DE, a partir del análisis de la distribución de la población al inicio de cada paso de predicción. Se pudo corroborar que ESSIM-DE posee alta velocidad de convergencia en los instantes de tiempo iniciales del desarrollo del incendio. Por lo tanto, en los últimos pasos de predicción DE no puede aportar mejoras, lo que conduce a un decaimiento de la aptitud. Se propuso una variante para mitigar dicho efecto sin modificar el esquema de procesamiento de la metaheurística DE. Para ello se

decidió generar una población inicial con distribución uniforme al inicio de cada paso de predicción. Esta variante del método fue denominada ESSIM-DE(r) y ha demostrado mejorar la calidad de las predicciones, obteniendo baja distribución de los resultados para los experimentos realizados.

Esta propuesta considera la generación de un espacio de búsqueda nuevo al inicio de cada paso de predicción. Sin embargo, en algunos casos puede resultar conveniente que ciertos individuos de la población persistan a lo largo de la simulación, con el objetivo de mantener algunas de las condiciones del entorno a través de los diferentes pasos de predicción. En este sentido, la propuesta que hemos presentado no permite que se conserven dichas características, dado que en cada paso se regenera la población completa. Por este motivo, se propone como trabajos a futuro analizar otras alternativas para salir del estancamiento sin tener que generar un espacio de búsqueda completamente diferente. El objetivo es estudiar y definir un modelo matemático que permita definir el grado de distribución que debería tener la población al inicio de cada paso de predicción.

Agradecimientos. Este trabajo ha sido financiado por UTN bajo el proyecto EIUTIME0003939TC y EIUTNME0003952, y por el Ministerio de Economía y Competitividad de España bajo el proyecto TIN2014-53234-C2-1-R. El primer autor agradece a CONICET por la beca doctoral otorgada.

Referencias

1. Méndez Garabetti M., Bianchini G., Caymes Scutari P., Tardivo M.L.: Sistema Estadístico Evolutivo con Modelo de Islas: Calibración de parámetros evolutivos. En: Int. Symposium on Innovation and Technology, pp. 77-81. Lima (2015)
2. Méndez Garabetti M., Bianchini G., Caymes Scutari P., Tardivo M.L.: ESS-IM applied to Forest Fire Spread Prediction: Parameters Tuning for a Heterogeneous Configuration. En: Int. Symposium SCCC 2016, pp. 514-525. Valparaíso (2016)
3. Tardivo M.L., Caymes-Scutari P., Bianchini G., Méndez-Garabetti M.: Ajuste de Parámetros Evolutivos para el Método Paralelo de Reducción de Incertidumbre ESSIM-DE. En: Congreso Argentino de Ciencias de la Computación, pp. 4-13. San Luis (2016)
4. Tardivo M.L., Caymes-Scutari P., Bianchini G., Méndez-Garabetti M.: Reducción de incertidumbre con Evolución Diferencial en la predicción de incendios forestales: sintonización y análisis de parámetros. En: Mecánica Computacional, vol. 34, pp. 2903-2917. AMCA (2016)
5. Tardivo M.L., Caymes-Scutari P., Bianchini G., Méndez-Garabetti M., Cortés A.: Three evolutionary statistical parallel methods for uncertainty reduction in wildland fire prediction. En: Int. Conf. HPCS 2016, pp. 721-728. Innsbruck (2016)
6. Talbi E.: Metaheuristics: From Design to Implementation. John Wiley & Sons (2009)
7. Viegas D.X.: Project Spread - Forest Fire Spread Prevention and Mitigation. Recuperado de <http://www.algosystems.gr/spread/>
8. Price K., Storn R., Lampinen J.: Differential Evolution - A practical approach to global optimization. Springer-Verlag New York, Inc. (2005)

Survival of the mutable: architecture of adaptive reactive agents

Renata Luiza Stange^{1,2}, Paulo Roberto Massa Cereda², and João José Neto²

¹ Universidade Tecnológica Federal do Paraná

Av. Prof^a. Laura Pacheco Bastos, 800, Industrial, Guarapuava, PR – Brasil

² Escola Politécnica, Departamento de Engenharia de

Computação e Sistemas Digitais, Universidade de São Paulo

Av. Prof. Luciano Gualberto, s/n, Travessa 3, 158, São Paulo, SP – Brasil

rlstange@usp.br, paulo.cereda@usp.br, jjneto@usp.br

Abstract. An agent is defined as any device that perceives a certain environment through stimuli and acts upon it as to achieve a certain goal. There is a plethora of theories, architectures and languages in the literature aiming at how much an agent may be improved at performing a task. However, the majority of them focuses on the internal agent function itself instead of adopting a macroscopic, broader view of what the term “intelligent” means in the long run. In this paper we take a bio-inspired route and describe how the simplest reactive agent can be boosted towards improvements at performing complex tasks by making it mutable. We provide a mathematical framework to support such features. Conceptually, the addition of a mutability layer does not break the existing paradigms and allows hybrid approaches as a means to achieve better results.

1 Introduction

Like other concepts of computation, specially the ones targeted in different areas, the definition of what an agent is does not have universally accepted consensus amongst the scientific community [11,1]. We mainly focus on Artificial Intelligence and therefore adopt both definition and classification of agents as seen in [25]. Russell and Norvig [25] define an agent simply as “something that acts”. According to this definition, it is possible to glimpse a computer program as an agent. However, there are scenarios in which a computational agent is expected to be rational and operate autonomously, perceiving its environment and eventually being susceptible to change [11].

Change, in this context, is a disputable concept when discussing behavioral actions [8,20]. In general, there is no clear indication of how such operation affects the agent behavior for subsequent interactions with the environment besides of a narrowed, error-prone local inspection [20]. A single misplaced action might mutilate the agent function and permanently compromise the entire task resolution. The more safety guards added to detect and avoid error propagation, the more complex and unfeasible the agent will be in terms of maintenance and

debugging [9]. Change might convey the idea of arbitrary actions taken from a confined, limited perspective.

The phrase “survival of the fittest” was coined by Herbert Spencer [27] as a way of describing the mechanism of natural selection, in which the biological concept of fitness is defined as reproductive success. Similarly, we pinpoint mutability as the key element for a bio-inspired agent model, such that the change concept is, from now on, bound to evolving the entire entity and not only its corresponding internal function [18].

Credited to theoretical physicist Stephen Hawking, the famous quote “intelligence is the ability to adapt to change” points us towards a better, stricter term to denote successive evolutions in an intelligent agent model: adaptation. However, it does not suffice being adaptable; as an agent inevitably has to operate autonomously, model adaptations must be triggered from within as opposed to explicit external calls [28]. Hence, the agent must be adaptive, in which the model spontaneously evolves without external interference, based solely on perceptions, stimuli and history. Biological entities are heavily based on such concept. In [3], Beer argues that the Artificial Intelligence area has overlooked the importance of adaptive behavior as a crucial substrate for intelligent behavior. Since the adaptive model is expected to occasionally evolve, we chose the simplest, most straightforward agent type from Russell and Norvig’s classification [25] as a proof of concept to demonstrate that the complexity of an agent may be greatly reduced while retaining intelligent behavior. In this paper, we aim at providing an architecture for such agents.

As to provide a consistent, coherent architecture for adaptive reactive agents, we rely on a mathematical formalism proposed by José Neto [13,12]. This particular approach has the clear advantage of providing a macroscopic model view, omitting superfluous details and characteristics [6]. Similar formalisms, such as cellular automata and recursive adaptable grammars, are interesting alternatives for designing bio-inspired models and may be employed as well.

2 Mathematical background

This section formally introduces the mathematical formalism proposed by José Neto [13]. Observe that the theory relies on an adaptive mechanism enclosing a non-adaptive rule-driven device, such that the latter may be enhanced in order to accommodate an adaptive behavior while preserving its integrity and original properties [12].

Definition 1 (rule-driven device). *A rule-driven device is defined as $ND = (C, NR, S, c_0, A, NA)$, such that ND is a rule-driven device, C is the set of all possible configurations, $c_0 \in C$ is the initial configuration, S is the set of all possible input stimuli, $\epsilon \in S$, $A \subseteq C$ is the subset of all accepting configurations (respectively, $F = C - A$ is the subset of all rejecting configurations), NA is the set of all possible output stimuli of ND as a side effect of rule applications, $\epsilon \in NA$, and NR is the set of rules defining ND as a relation $NR \subseteq C \times S \times C \times NA$.*

Definition 2 (rule). A rule $r \in NR$ is defined as $r = (c_i, s, c_j, z)$, $c_i, c_j \in C$, $s \in S$ and $z \in NA$, indicating that, as response to a stimulus s , r changes the current configuration c_i to c_j , processes s and generates z as output [13]. A rule $r = (c_i, s, c_j, z)$ is said to be compatible with the current configuration c if and only if $c_i = c$ and s is either empty or equals the current input stimulus; in this case, the application of a rule r moves the device to a configuration c_j , denoted by $c_i \Rightarrow^s c_j$, and adds z to the output stream.

Definition 3 (acceptance of an input stimuli stream by a rule-driven device). An input stimuli stream $w = w_1 w_2 \dots w_n$, $w_k \in S - \{\epsilon\}$, $k = 1, \dots, n$, $n \geq 0$, is accepted by a device ND when $c_0 \Rightarrow^{w_1} c_1 \Rightarrow^{w_2} \dots \Rightarrow^{w_n} c$ (in short, $c_0 \Rightarrow^w c$), and $c \in A$. Respectively, w is rejected by ND when $c \in F$. The language described by a rule-driven device ND is represented by $L(ND) = \{w \in S^* \mid c_0 \Rightarrow^w c, c \in A\}$.

Definition 4 (adaptive rule-driven device). A rule-driven device $AD = (ND_0, AM)$, such that ND_0 is a device and AM is an adaptive mechanism, is said to be adaptive when, for all operation steps $k \geq 0$ (k is the value of an internal counter T starting in zero and incremented by one each time a non-null adaptive action is executed), AD follows the behavior of an underlying device ND_k until the start of an operation step $k + 1$ triggered by a non-null adaptive action, modifying the current rule set; in short, the execution of a non-null adaptive action in an operation step $k \geq 0$ makes the adaptive device AD evolve from an underlying device ND_k to ND_{k+1} .

Definition 5 (operation of an adaptive device). An adaptive device AD starts its operation in configuration c_0 , with the initial format defined as $AD_0 = (C_0, AR_0, S, c_0, A, NA, BA, AA)$. In step k , an input stimulus move AD to the next configuration and starts the operation step $k + 1$ if and only if a non-adaptive rule is executed; thus, being the device AD in step k , with $AD_k = (C_k, AR_k, S, c_k, A, NA, BA, AA)$, the execution of a non-null adaptive action leads to $AD_{k+1} = (C_{k+1}, AR_{k+1}, S, c_{k+1}, A, NA, BA, AA)$, in which $AD = (ND_0, AM)$ is an adaptive device with a starting underlying device ND_0 and an adaptive mechanism AM , ND_k is an underlying device of AD in an operation step k , NR_k is the set of non-adaptive rules of ND_k , C_k is the set of all possible configurations for ND in an operation step k , $c_k \in C_k$ is the starting configuration in an operation step k , S is the set of all possible input stimuli of AD , $A \subseteq C$ is the subset of accepting configurations (respectively, $F = C - A$ is the subset of rejecting configurations), BA and AA are sets of adaptive actions (both containing the null action, $\epsilon \in BA \cap AA$), NA , with $\epsilon \in NA$, is the set of all output stimuli of AD as side effect of rule applications, AR_k is the set of adaptive rules defined as a relation $AR_k \subseteq BA \times C \times S \times C \times NA \times AA$, with AR_0 defining the starting behavior of AD , AR is the set of all possible adaptive rules for AD , NR is the set of all possible underlying non-adaptive rules of AD , and AM is an adaptive mechanism, $AM \subseteq BA \times NR \times AA$, to be applied in an operation step k for each rule in $NR_k \subseteq NR$.

Definition 6 (adaptive rules). Adaptive rules $ar \in AR_k$ are defined as $ar = (ba, c_i, s, c_j, z, aa)$ indicating that, as response to an input stimulus $s \in S$, ar initially executes the prior adaptive action $ba \in BA$; the execution of ba is canceled if this action removes ar from AR_k ; otherwise, the underlying non-adaptive rule $nr = (c_i, s, c_j, z)$, $nr \in NR_k$ is applied and, finally, the post adaptive action $aa \in AA$ is applied [13].

Definition 7 (adaptive function). Adaptive actions may be defined as abstractions named adaptive functions, similar to function calls in programming languages [13]. The specification of an adaptive function must include the following elements: (a) a symbolic name, (b) formal parameters which will refer to values supplied as arguments, (c) variables which will hold values of applications of elementary adaptive actions of inspection, (d) generators that refer to new value references on each usage, and (e) the body of the function itself.

Definition 8 (elementary adaptive actions). Three types of elementary actions are defined in order to perform tests on the rule set or modify existing rules, namely:

- (i) *inspection:* the elementary action does not modify the current rule set, but allows inspection on such set and querying rules that match a certain pattern. It employs the form $?(pattern)$.
- (ii) *removal:* the elementary action removes rules that match a certain pattern from the current rule set. It employs the form $-(pattern)$. If no rule matches the pattern, nothing is done.
- (iii) *insertion:* the elementary action adds a rule that match a certain pattern to the rule set. It employs the form $+(pattern)$. If the rule already exists in the rule set, nothing is done.

Such elementary adaptive actions may be used in the body of an adaptive function, including rule patterns that use formal parameters, variables and generators available in the function scope.

According to Cereda and José Neto [6,5], adaptive devices offer conveniences of a more compact model representation and better organization, as each underlying device covers a specific context at a time.

3 Adaptive reactive agents

An agent is defined as a device that perceives its environment through sensors and acts on this environment through actuators [25]. Perception refers to the perceptual stimuli of the agent and a sequence of perceptions consists of the history of all perceptions so far. The choice of action may depend only on the actual perception or sequence of perceptions obtained up to a given instant in time [25]. Similarly, an adaptive reactive agent also perceives its environment through sensors and acts in the environment through actuators, according to a

function f that determines its internal behavior; however, the rule set of f may spontaneously adapt over time, in response to operating history and perceptions, such as a conventional biological entity. Function f maps a set of perceptions into an action [25], such that $f: 2^P \mapsto AC$, where P is the set of perceptions and AC is the set of actions. The rule set adaptation occurs through the adaptive mechanism, triggered by the execution of adaptive rules [28].

An ordinary reactive agent acts upon the current perception, ignoring history. Such approach has the clear advantage of providing an intelligible, modular and efficient representation through $\langle condition \mapsto action \rangle$ rules [25]. However, simplicity comes at the cost of having no explicit knowledge representation, besides the absence of symbolic internal environmental representation and action memory [28].

Definition 9 (reactive agent). *A reactive agent is defined as $RA = (Q, P, AC, \Gamma, \delta)$, such that Q is the set of internal states, P is the set of perceptions, AC is the set of actions, Γ is the mapping, $\Gamma: Q \times 2^P \mapsto Q$, and δ is a function that maps rules into actions, $\delta: P \mapsto AC$.*

Reactive agents have a limiting characteristic: decision-making acts only on the current perception, implying the environment must be fully observable, otherwise the agent will potentially fail [25]. The inclusion of an adaptive layer in a reactive agent allows handling predicted yet unexpected situations in the current context. Adaptive actions evolve the agent through time and accommodate new contexts from environment observations [6,7].

Definition 10 (adaptive reactive agent). *An adaptive reactive agent is defined as $AA = (A_0, AM)$, such that A_0 is a reactive agent and AM is an adaptive mechanism. For all operation steps $k \geq 0$ (k is the value of a built-in counter T started at zero and incremented by one unit every time a non-null adaptive action is executed), AA follows the behavior determined by the underlying reactive agent A_k until a non-null adaptive action is executed, which starts the $k + 1$ operation step through adapting the rule set of the agent function.*

The architecture of an adaptive reactive agent presented in Figure 1 is heavily inspired by Russell and Norvig's model [25]. For convenience, we included an additional component containing learning modules in order to reduce complexity and verbosity. However, it is important to observe that the model in itself is powerful enough to handle any sort of computational task; the introduction of a new component is merely for convenience and expressiveness purposes [24].

According to Figure 1, the adaptive mechanism is represented as an adaptive action handler. Such handler intercepts requests from the underlying agent function and adapts the rule set according to patterns established in its action set and optionally in the learning modules. The underlying reactive agent evolves over time based on adaptations on its rule set, triggered by decision-making actions in the associated handler. Note that the adaptive layer enclosing the agent retains the original reactive property; however, the agent itself acts by context during its life cycle, one evolutionary reactive step at a time.

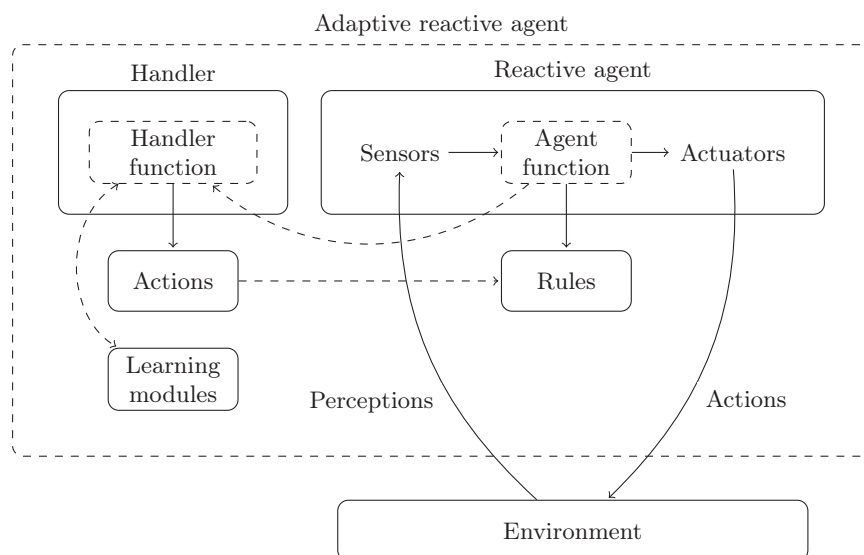


Fig. 1. Architecture of an adaptive reactive agent, heavily inspired by Russell and Norvig's model [25].

The first module employs associative learning, such that the agent is able to learn, memorize and reconstruct a pattern by associating the various parts of this pattern with one another [19,29]. This process involves the creation of a cognitive map. From a biological view, a cognitive map is the way certain beings memorize the information they gain about the spatial organization of their environment and how they make use of such information to navigate from one point to another [19]. Cognitive maps are usually represented as graphs with nodes that are computational elements. From this perspective, a cognitive map may be enhanced with higher level adaptation as well [26].

The second module employs reinforcement learning, such that the agent is able to recognize and favor certain behaviors that yield rewards rather than those that yield punishments [19,4]. Ethologists classify this characteristic as motivational system. The reinforcement process involves the production of goal seeking sequences through $\langle object \mapsto goal \rangle$ associations [4,10]. Booker [4] mentions the response (of his models) to regularities hidden behind the equivocal nature of sensory cues and use of information afforded by the environment with respect to goals.

New contexts provide incremental resolution by dividing a task into smaller subtasks [17,2]. Solving a task in a simpler abstraction space may be used to solve similar tasks at higher abstraction levels [15]. The process is then repeated until the original task is solved in its original abstraction space [23]. Such resolution strategy may produce significant search space reductions [21,22,16].

4 Experiments and discussions

In order to evaluate the concept of a reactive adaptive agent, we wrote a case study using the traditional *wumpus world* problem, a variant of a computer game [30] presented by Genesereth [25] as a testing environment for artificial intelligence techniques. The wumpus world consists of a two-dimensional grid containing a number of holes, a monster and an agent. The agent begins its iteration in grid position (1,1) and its task is set to avoid the monster and the holes, find the gold and leave the environment by reaching the same position it initiated. The agent can perceive a breeze in positions adjacent to the holes, a stench in positions adjacent to the monster, and a glow in positions adjacent to the gold. Variations of the wumpus world are available, in which the agent is armed with one or more arrows, being able to shoot the monster when found [14]. Figure 2 illustrates the possible contexts handled by the reactive adaptive agent, according to the execution of the corresponding adaptive actions.

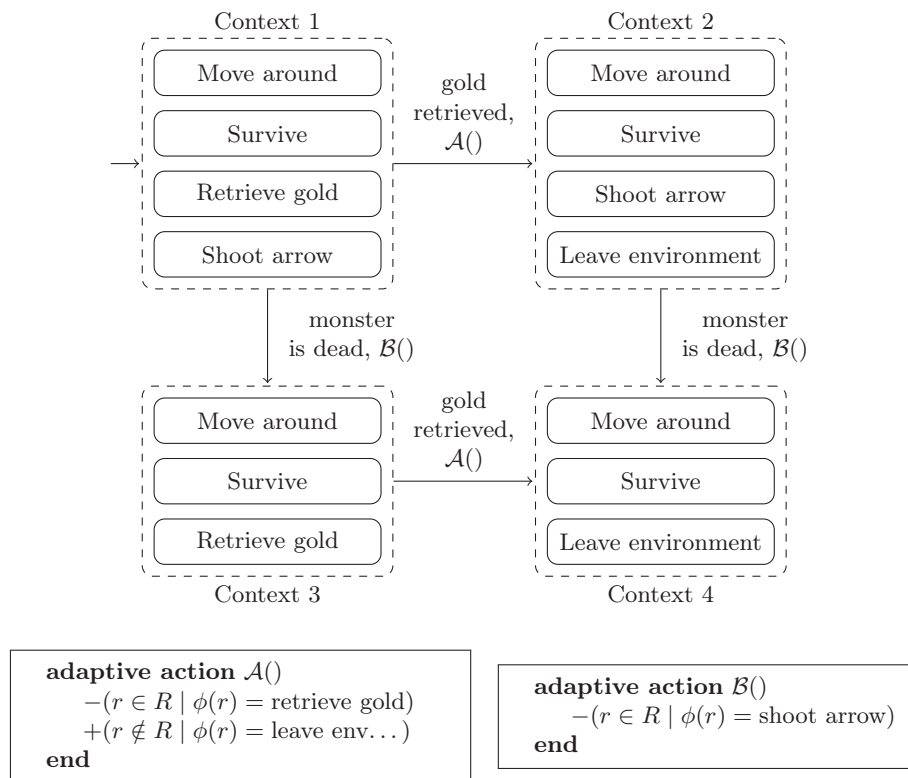


Fig. 2. Contexts handled by the reactive adaptive agent for the wumpus world problem.

The agent interaction in the wumpus world was simulated using an agent specification and simulation software developed by Jill Zimmerman³. We generated 14 random environments of size 4×4 with obstacles, gold and one monster. The simulation software measured the agent interaction with the environment and calculated a final score based on its overall performance [25]. We simulated the interaction of four types of agents – reactive, model-based, goal-based and adaptive reactive – in the generated environments and evaluated their global performances. We also introduced a second adaptive reactive agent enhanced with learning modules. The simulation was repeated 10 times for each ordered pair $\langle agent, environment \rangle$, totaling 700 executions. Results are shown in Figure 3.

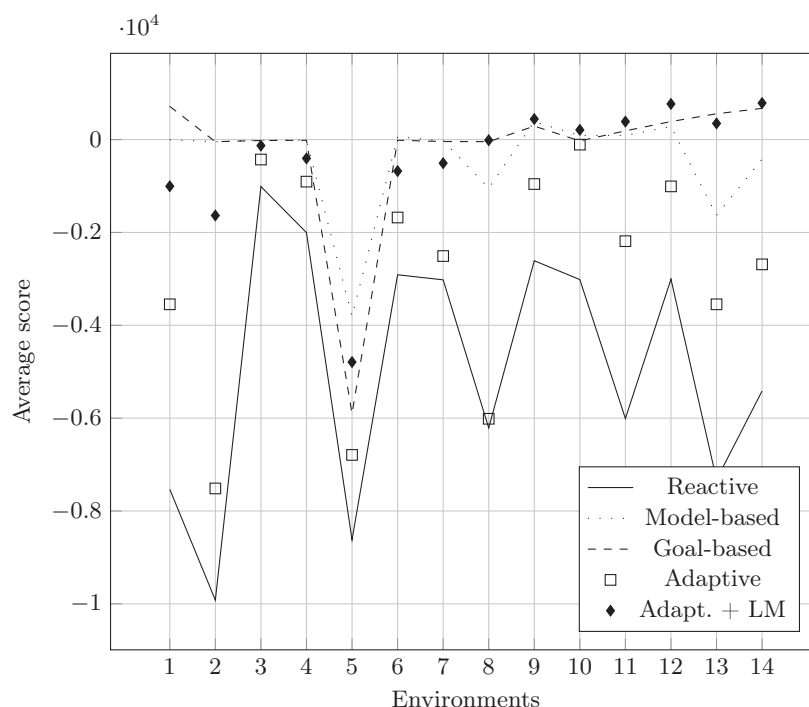


Fig. 3. Average score of 5 agents in 14 wumpus world environments.

According to Figure 3, agents had a similar overall execution for the same environments. In particular, the ones based on model and goals had significantly close performances. The reactive adaptive agent demonstrated a relatively better performance when compared to the ordinary reactive agent and, in some environments, had scores close to those based on model and goals. Most notably, when enhanced with learning modules, the adaptive reactive agent had a sig-

³ Available at <http://phoenix.goucher.edu/~jillz/cs340/>.

nificant performance boost, even surpassing the classical approaches on certain environments.

5 Final remarks

In this paper, we presented a mutable approach to the classic reactive agent through a bio-inspired model, while preserving its original properties. The key element towards intelligence is adaptation. New contexts bring practical benefits such as division of work, abstraction, reuse and maintenance. From a biological point of view, conservation of energy is achieved through a compact cognitive representation, since the current context holds no superfluous knowledge.

Through mutability, the organization of rules by context grants simplification of the original task, dividing it into smaller subtasks. Additionally, execution becomes more efficient in the long run since a context will only be activated if it is actually needed. Existing techniques may be combined [6,28] in order to specifically address each subtask.

Adaptivity is an interesting approach towards a natural, bio-inspired design of agents, such that the underlying behavior is flexibly adjusted to contingencies that arises in its interaction with an unknown environment. Mutability is a crucial feature for survival, and the understanding of adaptive behavior in biological beings may yield interesting results in the computational domain.

References

1. Ahmad, A., Ahmad, M.S., Yusof, M.Z.: An exploratory review of software agents. International Symposium on Information Technology (2008)
2. Anzai, Y., Simon, H.A.: The theory of learning by doing. *Psychological Review* (86), 124–140 (1979)
3. Beer, R.D.: *Intelligence as Adaptive Behavior: an Experiment in Computational Neuroethology*. Academic Press Inc, Boston (1990)
4. Booker, L.B.: Instinct as an inductive bias for learning behavioral sequences. In: *From Animals to Animats: Proceedings of the First International Conference on Simulation of Adaptive Behavior*. The MIT Press (1991)
5. Cereda, P.R.M., José Neto, J.: Towards performance-focused implementations of adaptive devices. *Procedia Computer Science* 109, 1164–1169 (2017)
6. Cereda, P.R.M., José Neto, J.: Adaptive data mining: Preliminary studies. *IEEE Latin America Transactions* 12(7), 1258–1270 (October 2014)
7. Cereda, P.R.M., José Neto, J.: Utilizando linguagens de programação orientadas a objetos para codificar programas adaptativos. In: *Memórias do IX Workshop de Tecnologia Adaptativa – WTA 2015*. pp. 2–9 (2015)
8. Chiel, H.J., Beer, R.D.: Simulation of adaptive behavior. *Current Opinion in Neurobiology* (1), 605–609 (1991)
9. Grillner, S., Buchanan, J.T., Lansner, A.: Simulation of the segmental burst generating network for locomotion in lamprey. *Neuroscience Letters* (89), 31–35 (1988)
10. Holland, J., Holyoak, K., Nisbett, R., Thagard, P.: *Induction: Processes of Inference, Learning and Discovery*. The MIT Press (1986)

11. Jennings, N.R., Sycara, K., Wooldridge, M.: A roadmap of agent research and development. *Autonomous Agents and Multi-Agent Systems* 1, 7–38 (1998)
12. José Neto, J.: Adaptive automata for context-dependent languages. *SIGPLAN Notices* 29(9), 115–124 (1994)
13. José Neto, J.: Adaptive rule-driven devices: general formulation and case study. In: *International Conference on Implementation and Application of Automata* (2001)
14. Khan, G.M., Miller, J.F., Halliday, D.M.: Advances in modeling adaptive and cognitive systems, chap. Intelligent agents capable of developing memory of their environment, pp. 77–114. *UEFS* (2010)
15. Knoblock, C.: Learning abstraction hierarchies for problem solving. In: *Proceedings of the Eighth National Conference on Artificial Intelligence*. pp. 923–928 (1990)
16. Knoblock, C.: Search reduction in hierarchical problem solving. In: *AAAI'91 Proceedings*. pp. 686–691 (1991)
17. Knoblock, C.: *Generating Abstraction Hierarchies: An Automated Approach to Reducing Search in Planning*. Springer US (1993)
18. Kristan, W.B., McGirr, S.J., Simpson, G.V.: Behavioural and mechanosensory neurone responses to skin stimulation in leeches. *Journal of Experimental Biology* (96), 143–160 (1982)
19. Matarić, M.: Navigating with a rat brain: A neurobiologically-inspired model for robot spatial representation. In: *From Animals to Animats: Proceedings of the First International Conference on Simulation of Adaptive Behavior*. The MIT Press (1991)
20. Meyer, J.A., Guillot, A.: Simulation of adaptive behavior in animals: review and prospect. In: *From Animals to Animats: Proceedings of the First International Conference on Simulation of Adaptive Behavior*. MIT Press (1991)
21. Minsky, M.: *Computers and Thought*, chap. Steps toward artificial intelligence, pp. 406–450. McGraw-Hill, New York, NY (1963)
22. Newell, A., Simon, H.A.: *Contemporary Approach to Creative Thinking*, chap. The processes of creative thinking, pp. 63–119. Atherton Press (1962)
23. Pólya, G.: *How to Solve It*. Princeton University Press (1945)
24. Rocha, R.L.A., José Neto, J.: Autômato adaptativo, limites e complexidade em comparação com a Máquina de Turing. In: *Proceedings of the Second Congress of Logic Applied to Technology*. pp. 33–48 (2000)
25. Russell, S.J., Norvig, P.: *Artificial Intelligence: A Modern Approach*. Pearson Education, 2 edn. (2003)
26. Silva Filho, R.I., Rocha, R.L.A.: Adaptive and Natural Computing Algorithms: 10th International Conference, ICANNGA 2011, Ljubljana, Slovenia, April 14–16, 2011, *Proceedings, Part II*, chap. Adaptive Finite Automaton: A New Algebraic Approach, pp. 275–284. Springer Berlin Heidelberg (2011)
27. Spencer, H.: *The Principles of Biology*, vol. 1. John Childs and Son, London (1864)
28. Stange, R.L., Cereda, P.R.M., José Neto, J.: Agentes adaptativos reativos: formalização e estudo de caso. In: *Memórias do XI Workshop de Tecnologia Adaptativa – WTA 2017*. pp. 63–71. São Paulo (2017)
29. Stange, R.L., José Neto, J.: Learning decision rules using adaptive technologies: a hybrid approach based on sequential covering. *Procedia Computer Science* 109, 1188–1193 (2017)
30. Yob, G.: Hunt the wumpus. *People's Computer Company* 2(1) (1973)

Un formalismo para persuasión con DeLP

Guillermo Aguirre, Marcelo Errecalde¹, Guillermo Simari²

¹ Lab. de Investigación y Desarrollo en Inteligencia Computacional (LIDIC),
{gaguirre,merreca}@uns1.edu.ar

Departamento de Informática. Universidad Nacional de San Luis
Ejército de los Andes 950 - Local 106. (D5700HHW) - San Luis - Argentina

² Laboratorio de Investigación y Desarrollo en Inteligencia Artificial (LIDIA),
grs@cs.uns.edu.ar

Depto. de Ciencias e Ingeniería de la Computación. Univ. Nacional del Sur,
Av. Alem 1253, (B8000CPB) Bahía Blanca, Argentina

Resumen La comunicación entre agentes requiere estar regulada y debe cumplir ciertos formalismos para facilitar la tarea de cada uno de los agentes involucrados. Bajo esos formalismos, la persuasión permite que dos agentes se pongan de acuerdo sobre alguna situación particular. El acuerdo se logra mediante los protocolos de persuasión, los cuales entre otras cosas requieren de dos lenguajes para formar las frases, por un lado, un lenguaje de comunicación general y por otro, un lenguaje específico del contexto del diálogo. Es necesario que este último lenguaje resuelva situaciones de conflicto en favor de quien haga sus planteos con argumentos más sólidos, para lo cual es conveniente el uso de algún lenguaje de programación lógica. En este trabajo se propone el uso de DeLP como dicho lenguaje, ya que posee las características requeridas, y mediante un programa de ejemplo se muestra que esta propuesta es viable.

Palabras claves juegos de diálogo, argumentación, DeLP.

1. Introducción

La comunicación entre los seres humanos consiste en el intercambio estructurado de frases y cuando se respetan las reglas propias del entorno, se consigue transmitir ideas de una manera racional. Cada frase se emite por una determinada razón, la que puede o no ser alcanzada por distintos motivos. Esta característica imprecisa de los diálogos ha dado origen a numerosos estudios: *Cómo hacer las cosas con palabras* [1] y *Teoría de acción comunicativa* [14]. A pesar de los diversos enfoques, existe coincidencia en considerar a las frases cortas como *actos del habla* porque pueden cambiar la realidad del mismo modo que las acciones físicas. Lo que resulta evidente es la necesidad de establecer contextos o sistemas que regulen los distintos diálogos posibles, especialmente cuando los participantes son agentes computacionales. Este trabajo se concentra en una clase particular de comunicación que ocurre habitualmente, una parte sostiene algo mediante argumentos que la otra parte pone a prueba.

El término “diálogo de persuasión” es utilizado por Walton & Krabber [16] como parte de su influyente clasificación de los diálogos en seis tipos de acuerdo a su finalidad. Mientras que la *persuasión* procura resolver una diferencia de opinión, la *negociación* trata de resolver un conflicto de intereses mediante un trato, la *búsqueda de información* alienta la transferencia de información, la *deliberación* procura alcanzar una decisión sobre un curso de acción, una *consulta* pretende “incrementar el conocimiento y el acuerdo” y la *disputa* es el sustituto verbal de la pelea. De acuerdo a esta clasificación es posible que los diálogos de una clase se conviertan en otro tipo. En particular, un diálogo de cualquier tipo puede transformarse en uno de persuasión cuando surge un conflicto de opinión. Por ejemplo en la búsqueda de información, puede surgir un conflicto de opinión sobre la credibilidad de una fuente de información, en una deliberación los participantes pueden disentir sobre los probables resultados de los planes o acciones y en una negociación los actores involucrados pueden no coincidir sobre las razones por las que una propuesta resulta de interés para alguno de ellos.

La estructura de comportamiento requerida en los diálogos se asemeja a las reglas de un juego; un tipo de juegos practicado entre oradores, quienes deben presentar diversas sentencias de acuerdo a reglas estrictas que han ido evolucionando en forma de distintos procedimientos. Por ejemplo, en [4,6,15] se plantean algunos juegos que modelan el razonamiento rebatible como una disputa entre proponente y oponente de una proposición. El proponente comienza con un argumento para la proposición inicial, después de lo cual, cada uno debe atacar algún argumento previo del otro participante con un contraargumento de fuerza suficiente. Los ataques pueden estar dirigidos directamente a la conclusión, lo que se conoce como *refutar* (en inglés rebutting), o bien estar dirigidos a las premisas que soportan la conclusión, *socavar* (en inglés undercutting) [5]. La proposición inicial resulta probada si el proponente tiene una estrategia ganadora, es decir, si él puede hacer que el oponente se quede sin movidas para hacer cualquier ataque. Claramente, este entorno dialéctico se corresponde bien con las aplicaciones de sistemas de diálogo mencionadas anteriormente.

Este trabajo sigue el sistema formal para persuasión presentando en [8] y [9] y plantea crear una instancia del protocolo para los juegos de diálogo propuesto por Prakken pero usando DeLP [2,3]. El fundamento de esta idea son las similitudes existentes entre el sistema argumentativo usado en la propuesta original (el que denominaremos PraTor [10]) y DeLP. En particular el hecho que ambos sistemas usan semántica fijas [11], una condición requerida en [7, pag. 70].

Ejemplo 1 *Para mostrar el uso de persuasión en el mundo real se adaptó un ejemplo presentado en [13]. El problema es el siguiente: un par de amigos viajan en auto y uno de ellos sugiere, gracias la información del contexto que obtiene dinámicamente del GPS del vehículo, parar en un hotel cercano. ¿Mediante qué elementos se podrán sostener argumentos a favor de la sugerencia?*

El trabajo se organiza así, la sección 2 plantea el uso de persuasión para lograr acuerdos entre agentes racionales y se presenta un ejemplo. Las características generales del sistema de argumentación DeLP se presentan en la Sección 3,

ahí mismo está el Ejemplo 1 como un programa DeLP. Luego en la Sección 4 se explican los principales elementos del protocolo para persuasión y se muestra cómo se puede instanciar el formalismo empleando el lenguaje DeLP. En la Sección 5 se presentan algunas conclusiones y futuras líneas de investigación.

2. Persuasión

En los diálogos de persuasión, la situación inicial es un conflicto de opinión lo que impide alcanzar una efectiva comunicación; para superar esta circunstancia se procura hallar una solución basada en la comunicación, una que sea dialógica. En esta búsqueda se encuentran los sistemas de diálogo que se caracterizan por establecer una serie de reglas sobre el uso de los actos del habla, permitiendo que los participantes intercambien afirmaciones o reclamos, planteen desafíos sobre las afirmaciones de otro participante³, acepten las afirmaciones del otro, puedan retirar las afirmaciones propias hechas anteriormente o presenten argumentos a favor de ellas. Estos sistemas de diálogo han sido abordados de una manera formal desde mediados del siglo pasado. La idea que la argumentación está relacionada con el diálogo, por lo tanto con las obligaciones y los *compromisos* en el diálogo, es tan vieja como la lógica misma.

Ejemplo 2 *En la Figura 1 se presenta un posible diálogo entre los agentes involucrados en la situación planteada en el Ejemplo 1.*

A:Sugiero parar en este hotel. (*Hace un reclamo*)
 B:¿Por qué parar en este hotel? (*Consulta sobre los fundamentos del reclamo*)
 A:Ya debemos parar y este hotel está cerca. (*Brinda fundamentos del reclamo*)
 B:¿Por qué debemos parar? (*Consulta sobre los fundamentos de un sub-argumento*)
 A:Hemos manejado durante muchas horas. (*Brinda fundamentos del reclamo*)
 B:Pero hemos hecho muchas paradas. (*Establece un contraargumento*)
 A:Debemos parar debido a que es de noche. (*Establece otro contraargumento*)
 B:Si es cierto, debemos parar. (*Acepta el sub-argumento anterior*)
 Pero no en este hotel, esta zona es peligrosa. (*Establece un nuevo contraargumento*)
 A:¿Por qué esta zona es peligrosa? (*Consulta los fundamentos del contraargumento*)
 B:Debido a que hay muchos robos. (*Brinda fundamentos del contraargumento*)
 A:De acuerdo, no fue buena la sugerencia de parar aquí. (*Retira el reclamo inicial*)
 Mejor no paremos en este hotel. (*Acepta el nuevo contraargumento*)

Figura 1. Ejemplo de diálogo entre agentes racionales

Se establecen algunas restricciones sobre las frases admitidas durante la comunicación, de esta manera se facilita tanto la interpretación como la alternancia

³ Lo que en términos legales se conoce como reconvencción: cuando el demandado inicia acciones contra el demandante.

de los turnos. Cada participante de la comunicación contará con un *almacén de compromisos*, el cual representa el punto de vista públicamente declarado por el agente sobre una proposición; esta perspectiva no necesita coincidir con las creencias internas del agente. Pero sí es necesario que el contenido del almacén de compromisos mantenga consistencia.

3. Argumentación en DeLP

Un programa en DeLP[2,3] es $P = \langle \Theta, \Omega, \Delta \rangle$, un conjunto de hechos, reglas estrictas, y reglas rebatibles. En ocasiones se usa $\Pi = \Theta \cup \Omega$. Los hechos de Θ , son literales fijos identificados como L_i . En los dos tipos de reglas, el literal a la izquierda de la flecha es la *cabeza* y los literales que están a la derecha de la flecha son el *cuerpo*. Las reglas estrictas junto con los hechos representan el conocimiento incuestionable. Las reglas rebatibles sostienen conocimiento tentativo: razones para creer en *cuerpo* proveen razones para creer en *cabeza*. Los elementos del conjunto de reglas estrictas Ω se caracterizan como $L_0 \leftarrow L_1, \dots, L_k$, por su parte, las reglas rebatibles de Δ se muestran como $L_0 \prec L_1, \dots, L_k$, puede ocurrir que el cuerpo de estas reglas sea vacío; en ese caso “se supone L_0 ”.

Para explicar cómo funciona DeLP se puede comenzar diciendo que a partir de un programa P se hace una derivación rebatible de literales. Con $P \vdash L$ se indica que el programa deriva rebatiblemente el literal L , el último de una secuencia de literales fijos. Cada elemento de la secuencia es un hecho o la cabeza de una regla (estricta o rebatible) en la cual todos los literales del cuerpo han aparecido antes en la secuencia. Considerando la amplia diversidad de alternativas, es posible derivar literales *contradictorios* (por ejemplo L y \bar{L}).

Luego se continúa con el *procedimiento de prueba*, que es un mecanismo global mediante el cual se determina cuáles son los literales *aceptados* a partir de un programa particular. El mecanismo de aceptación mencionado sirve para seleccionar ciertos *argumentos* adecuados a un literal específico. Para cualquier programa P , un argumento es una estructura con dos elementos: A que es un conjunto de reglas rebatibles en P y L que es un literal obtenido a partir de una derivación rebatible de P . También se debe cumplir lo siguiente:

1. $\Pi \cup A \vdash L$
2. $\Pi \cup A$ no es contradictorio
3. A es mínimo.

Para continuar con la explicación, es necesario aclarar lo siguiente: $\langle B, Q \rangle$ es un *sub-argumento* de $\langle A, L \rangle$ si $Q \subseteq A$. Hecha la aclaración, se introduce un concepto importante. El argumento $\langle A_1, L_1 \rangle$ *ataca* a otro argumento $\langle A_2, L_2 \rangle$ en el literal L , si $\langle A, L \rangle$ es un sub-argumento de $\langle A_2, L_2 \rangle$, tal que L_1 y L son contradictorios o están en desacuerdo.

El hecho que un programa pueda derivar literales contradictorios da origen al ataque entre sus correspondientes argumentos. El ataque entre los argumentos se resuelve mediante comparación, este procedimiento de comparación es el que determina si un argumento derrota a otro.

Se usa $\langle A_1, L_1 \rangle \succ \langle A, L \rangle$ para indicar la preferencia estricta del argumento que aparece a la izquierda del operador \succ . Para profundizar sobre la preferencia entre argumentos revisar las referencias [12,3,13].

Es posible ahora formalizar la derrota entre argumentos: $\langle A_1, L_1 \rangle$ derrota $\langle A_2, L_2 \rangle$ en L , si $\langle A, L \rangle$ es un sub-argumento de $\langle A_2, L_2 \rangle$ y L es el punto de *contra-argumentación* del ataque de $\langle A_1, L_1 \rangle$ a $\langle A_2, L_2 \rangle$. Además se debe cumplir una de las dos alternativas:

- $\langle A_1, L_1 \rangle \succ \langle A, L \rangle$. Es decir, $\langle A_1, L_1 \rangle$ es un derrotador propio de $\langle A, L \rangle$.
- $\langle A_1, L_1 \rangle \asymp \langle A, L \rangle$ ⁴. Donde $\langle A_1, L_1 \rangle$ es llamado un *derrotador por bloqueo*.

Un argumento que no es derrotado brinda razones para creer en el literal que lo integra. Resulta interesante, entonces, considerar lo que ocurre cuando hay una secuencia de derrotas entre los argumentos de un programa. Es decir, cuando un primer argumento derrota al segundo pero un tercer argumento derrota al primero. En ese caso el literal del segundo argumento se mantiene en vigencia como no derrotado. La situación anterior caracteriza lo que se llama una *línea de argumentación*.

Dado $\langle A_1, L_1 \rangle$, la secuencia $\Lambda = [\langle A_1, L_1 \rangle, \dots, \langle A_{i+1}, L_{i+1} \rangle]$ donde cada $\langle A_j, L_j \rangle$ derrota al anterior, es una línea de argumentación. Los argumentos en las posiciones impares tienen el rol de *soporte* del literal del primer argumento. En ese mismo sentido, los de posiciones pares son llamados argumentos de *interferencia*. Además Λ debe cumplir las siguientes condiciones:

1. Λ es finita,
2. El conjunto de argumentos de soporte en Λ es no contradictorio y el conjunto de argumentos de interferencia en Λ también es no contradictorio,
3. Ningún argumento A_j en Λ es un sub-argumento de otro A_i con $i < j$.
4. Todo derrotador por bloqueo A_i en Λ solo puede ser seguido por un derrotador propio A_{i+1} en Λ .

Como cada argumento puede tener varios derrotadores, la ramificación que ocurre en las líneas de argumentación origina un árbol. Un *árbol de dialéctica* $T_{\langle A_0, L_0 \rangle}$ se construye partiendo de $\langle A_0, L_0 \rangle$ como raíz. Todo nodo N es interior si tiene al menos un derrotador, por lo cual, un nodo es una hoja solamente si no tiene derrotadores.

El *marcado* de un árbol de dialéctica es el resultado de un proceso que consiste en marcar o etiquetar cada nodo como derrotado “D” o no derrotado “U”. Se trabaja de la siguiente forma: los nodos hojas son marcados como “U”. Un nodo interno N será marcado como “D” si tiene al menos un hijo marcado como “U”, N será “U” si todos sus hijos están marcados como “D”.

Un Árbol marcado $T_{\langle A_0, L_0 \rangle}^*$ resulta de interés porque el argumento $\langle A_0, L_0 \rangle$ está *garantizado* si su marca es “U”. Todo argumento que está garantizado es especialmente útil porque ningún otro argumento del programa puede derrotarlo; eso quiere decir que resultará victorioso al ser usado en cualquier disputa. Si $\langle A_0, L_0 \rangle$ está garantizado, entonces el literal que lo acompaña L_0 , también lo está.

⁴ El operador \asymp se usa cuando los argumentos son indiferentes o incomparables.

Ejemplo 3 En la Figura 2 se muestra el programa DeLP que corresponde a la realidad presentada en el Ejemplo 1.

$$\Psi_h = \left\{ \begin{array}{l} mParadas \\ mHManejando \\ deNoche \\ hCercano(h1) \\ estrellas(h1,5) \\ zRobos(h1) \\ cVehicular \leftarrow tLento \end{array} \right\} \Delta_h = \left\{ \begin{array}{l} sugerir(X) \prec bHotel(X), hCercano(X) \\ sugerir(X) \prec sParar, hCercano(X) \\ \sim sugerir(X) \prec zPeligrosa(X) \\ zPeligrosa(X) \prec zRobos(X) \\ \sim zPeligrosa(X) \prec zPolicias(X) \\ bHotel(X) \prec estrellas(X,S), S \geq 3 \\ \sim bHotel(X) \prec estrellas(X,S), S < 3 \\ \sim sParar \prec mParadas \\ sParar \prec mHManejando \\ sParar \prec deNoche \end{array} \right\}$$

Figura 2. Programa DeLP para sugerir un hotel

Observe que el conjunto Ψ_h tiene seis hechos y una regla estricta. Estos hechos representan información sobre la situación actual y que puede ir variando durante el viaje. Hasta el momento se sabe que: se realizaron muchas paradas (mParadas), el conductor ha manejado muchas horas (mHManejando), ya es de noche (deNoche), “h1” es un hotel cercano (hCercano), “h1” es un hotel de cinco estrellas (estrellas(h1,5)), y “h1” está en una zona de robos (zRobos(h1)). La regla estricta representa que “hay congestión vehicular cuando el tránsito es muy lento”.

El conjunto Δ_h contiene varias reglas rebatibles. Las primeras dos reglas representan razones para sugerir un hotel: “si es cercano y es un buen hotel” o “si es un hotel cercano y hay una razón que sugiere parar”. Las últimas tres reglas representan razones a favor y en contra de sugerir dejar de conducir para descansar, siendo que “si el conductor lleva muchas horas manejando o está manejando de noche éstas son razones para sugerir una parada”, mientras que “si durante el viaje el conductor ha realizado varias paradas, es una razón en contra de sugerir que el conductor pare”. Observe que la sexta y séptima reglas rebatibles son utilizadas para establecer si un hotel es bueno. La tercer regla expresa que “si un hotel está ubicado en una zona peligrosa entonces hay una razón tentativa para no sugerirlo”. Finalmente, la cuarta y quinta reglas pueden ser leídas de la siguiente manera: “un hotel está ubicado en una zona peligrosa si en la zona normalmente se reportan muchos casos de robos”, y por lo contrario “si la zona generalmente cuenta con presencia policial, entonces el hotel no está ubicado en una zona peligrosa”.

4. Juego de diálogo

Cada intervención de los participantes de un diálogo se denomina *movida* y consiste en algún acto del habla formado a partir de elementos del lenguaje de comunicación L_c y del lenguaje del tópico L_t . L_c está relacionado con aspectos como la asignación de los turnos para que los interlocutores hagan sus movidas, la vinculación de cada respuesta con la movida a la cual va dirigida o diferenciar claramente cuando una respuesta es un ataque o una admisión; es decir, que este lenguaje está involucrado con el desarrollo general del diálogo. Por su parte el L_t debe resolver los ataques entre las movidas y establecer de una manera racional cuando una movida derrota a otra. Con estos lenguajes y otras reglas más, el protocolo llega a especificar cómo se determina el ganador actual, cuándo es el turno de cada participante, cuándo se ha llegado a la terminación del diálogo, cuál es el efecto de cada frase sobre los compromisos de los participantes, etc. Para consultar más detalles acerca de los elementos del protocolo ver [8] y [9].

El principal interés de este marco es enfocarse en los diálogos, mientras se deja suficiente libertad como para realizar réplicas alternativas y para posponer las réplicas. Esta libertad se consigue mediante dos características principales del sistema. Primero, \mathcal{L}_c tiene una estructura de respuesta explícita, donde cada movida o bien ataca o admite a la movida objetivo. Un \mathcal{L}_c de ejemplo⁵ con este formato se muestra en la Tabla 1. Segundo, el ganador se define para cada uno de los diálogos, ya sea que esté terminado o no; el resultado es determinado por el *estado dialógico* de la raíz del *árbol de diálogo*. Durante el diálogo las movidas pueden ser atacadas o admitidas de distintas formas por el otro participante, es así como se abren distintas posibilidades durante la comunicación; algo similar al árbol de dialéctica pero en este caso cada nodo en lugar de ser un argumento, es una movida como las que aparecen en la Tabla 1. El estado dialógico de una movida se define recursivamente como sigue: para determinar el estado del diálogo, se hace uso de las características del árbol de diálogos generado por la estructura de respuesta del \mathcal{L}_c , ya que dicho árbol mantiene el estado de las distintas movidas. Una movida está *adentro* si es admitida o si todas las réplicas que la atacan están *afuera*. (Esto implica que una movida sin réplicas está *adentro*). Y una movida está *afuera* si tiene una réplica que esté *adentro*. Entonces un diálogo está siendo ganado (actualmente) por el proponente si su movida inicial está *adentro*, mientras que está (actualmente) siendo ganada por el oponente en otro caso.

El lenguaje de comunicación propuesto es el mostrado en la Tabla 1 y está restringido por el requerimiento de que cada movida debe ser *relevante*. Esto hace que el protocolo sea de *respuesta inmediata*, lo cual implica que cada turno consista de cero o más admisiones seguidas por un ataque. Dentro de estos límites la posposición de respuesta está permitida, a veces incluso indefinidamente.

⁵ La movida *se supone φ* no aparece en la propuesta original de Prakken, corresponde a una regla rebatible con cuerpo vacío en DeLP.

Acto	Ataques	Admisiones
reclamo φ	ζ por qué φ ?	concedo φ
φ dado S	ζ por qué ψ ? ($\psi \in S$) φ' dado S' (φ' dado S' derrota φ dado S) se supone φ' ($\Rightarrow \varphi'$ derrota φ dado S)	concedo ψ ($\psi \in S$) concedo φ
ζ por qué φ ?	φ dado S se supone φ	retiro φ
se supone φ	φ' dado S se supone φ'	concedo φ
concedo φ		
retiro φ		

Tabla 1. Un ejemplo de lenguaje de comunicación

4.1. Administración de los compromisos

A medida que progresa el diálogo se producen cambios en los respectivos almacenes de compromisos de los agentes. El conocimiento del agente no tiene nada que ver con los compromisos que van siendo asumidos durante el diálogo. Dentro de cada almacén, los compromisos deben mantener consistencia. Algunas de las locuciones de la Tabla 1 agregan o quitan proposiciones, a continuación se analizan los casos. Cada *reclamo* implica agregar la proposición al almacén de compromisos del agente que lo emitió. Cuando un agente *concede* un reclamo de otro agente, el agente que admite debe agregar la proposición del reclamo. Al *retirar* la proposición hecha en un reclamo previo, se la debe quitar del almacen de compromisos del emisor. Al usar el acto " φ dado S " se deben incorporar el consecuente y las premisas que aparecen en el argumento involucrado.

Ejemplo 4 *Continuación del ejemplo 2. Usando un lenguaje de comunicación con respuesta explícita como el propuesto, el diálogo entre los agente A y B, sería similar al mostrado en la Tabla 2.*

$P_1[-]$: <i>reclamo</i> sugerir(h1)	$O_2[P_1]$: ζ por qué sugerir(h1)?
$P_3[O_2]$: sugerir(h1) <i>dado</i> sParar, hCercano(h1)	$O_4[P_3]$: ζ por qué sParar?
$P_5[O_4]$: sParar <i>dado</i> mHManejando	$O_6[P_5]$: \neg sParar <i>dado</i> mParadas
$P_7[O_6]$: sParar <i>dado</i> deNoche	$O_8[P_7]$: <i>concedo</i> sParar
$P_{10}[O_9]$: ζ por qué zPeligrosa(h1)?	$O_9[P_1]$: \neg sugerir(h1) <i>dado</i> zPeligrosa(h1)
$P_{12}[O_2]$: <i>retiro</i> sugerir(h1)	$O_{11}[P_{10}]$: zPeligrosa(h1) <i>dado</i> zRobos(h1)
$P_{13}[O_9]$: <i>concedo</i> \neg sugerir(h1)	

Tabla 2. Protocolo relevante para el diálogo ejemplo

Ejemplo 5 Al finalizar el diálogo del Ejemplo 4 los compromisos de cada agente son los siguientes:

Compromisos de $A = \{sParar, hCercano(h1), mHManejando, deNoche, \neg sugerir(h1)\}$

Compromisos de $B = \{sParar, mParadas, zPeligrosa(h1), zRobos(h1), \neg sugerir(h1)\}$

A lo largo del artículo se han presentado distintas situaciones de ataque y admisión entre movidas; para mostrar cómo se aplican estas ideas concretamente en el caso del diálogo del Ejemplo 4, se muestra a continuación el correspondiente árbol de diálogo.

Ejemplo 6 Como se observa en la Figura 3, la raíz del árbol de diálogo se muestra en línea puntos indicando que el oponente es el actual ganador del diálogo.

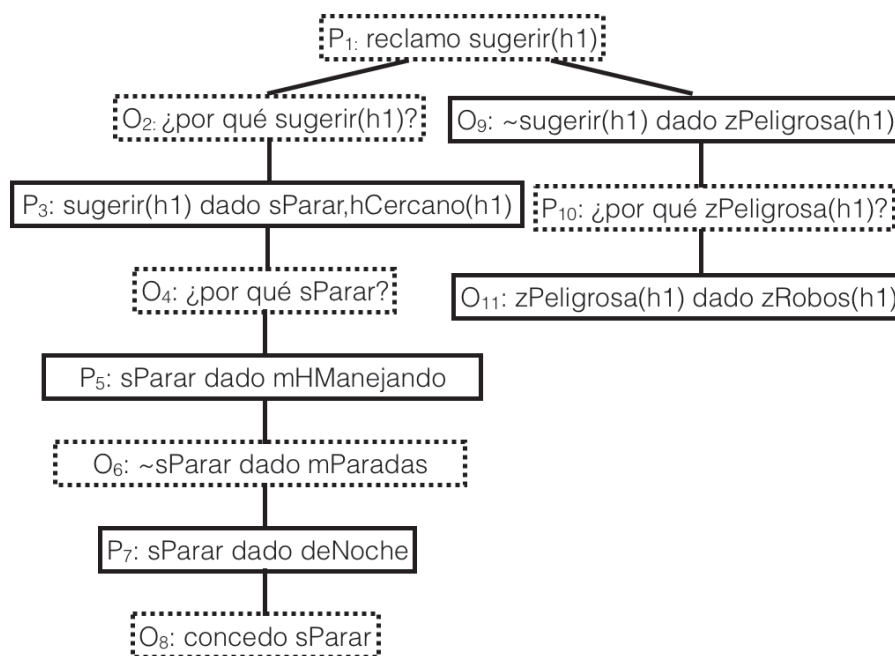


Figura 3. Árbol de diálogo

5. Consideraciones Finales

PraTor y DeLP son sistemas de argumentación que tienen suficientes similitudes como para sugerir el uso de cualquiera de los dos como lenguaje del tópicos en el sistema formal para diálogos de persuasión propuesto por Prakken. En este

trabajo no se han discutido en detalle las similitudes entre Prator y DeLP, simplemente se ha mostrado un ejemplo en el que DeLP se utiliza como lenguaje del t3pico particular de dicho ejemplo. Este trabajo busca destacar la flexibilidad del protocolo de persuasi3n que permite ser aplicado en una situaci3n particular empleando un sistema argumentativo diferente al usado en la propuesta original. De esta manera se refuerza la idea de que la propuesta de Prakken es robusta, ya que puede ser empleada tambi3n con DeLP. La diferencia entre DeLP y la programaci3n l3gica extendida con prioridades rebatibles, usada en la propuesta original, es que el primero ha sido implementado, por lo que es posible que los di3logos se ejecuten en el DeLP server.

La sugerencia planteada en este trabajo requiere una discusi3n m3s profunda y t3cnica que se realizar3 en futuros trabajos.

Los autores agradecen las observaciones realizadas por los evaluadores.

Referencias

1. Austin, J.L.: *How to Do Things with Words*. Harvard University Press, Cambridge, MA, Cambridge, MA (1962)
2. Garc3a, A.J.: *Defeasible Logic Programming: Definition, Operational Semantics and Parallelism*. Ph.D. thesis, Universidad Nacional del Sur, Bah3a (2000)
3. Garc3a, A.J., Simari, G.R.: *Defeasible logic programming: An argumentative approach*. *TPLP* 4(1-2), 95–138 (2004)
4. Loui, R.P.: *Process and policy: Resource-bounded nondemonstrative reasoning*. *Computational intelligence* 14(1), 1–38 (1998)
5. Pollock, J.: *Cognitive Carpentry: A Blueprint for how to Build a Person*. Bradford Bks, MIT Press (1995)
6. Prakken, H.: *Relating protocols for dynamic dispute with logics for defeasible argumentation*. *Synthese* 127(1-2), 187–219 (2001)
7. Prakken, H.: *Coherence and flexibility in dialogue games for argumentation*. *Journal of logic and computation* 15(6) (2005)
8. Prakken, H.: *Formal systems for persuasion dialogue*. *The knowledge engineering review* 21(02), 163–188 (2006)
9. Prakken, H.: *Models of persuasion dialogue*. In: Rahwan, I., Simari, G.R. (eds.) *Argumentation in artificial intelligence*, pp. 281–300. Springer (2009)
10. Prakken, H., Sartor, G.: *Argument-based extended logic programming with defeasible priorities*. *Journal of applied non-classical logics* 7(1-2), 25–75 (1997)
11. Prakken, H., Vreeswijk, G.: *Logics for Defeasible Argumentation*, pp. 219–318. Springer Netherlands, Dordrecht (2002)
12. Simari, G.R., Loui, R.P.: *A mathematical treatment of defeasible reasoning and its implementation*. *Artif. Intell.* 53(2-3), 125–157 (1992)
13. Teze, J.C.L.: *Formalizaci3n y Generalizaci3n del Manejo de Preferencias en Servicios de Razonamiento Rebatible*. Ph.D. thesis, U.N.S., Bah3a Blanca (2017)
14. Verharen, E.: *A Language-action Perspective on the Design of Cooperative Information Agents*. E.M. Verharen (1997)
15. Vreeswijk, G.A.: *The computational value of debate in defeasible reasoning*. *Argumentation* 9(2), 305–342 (1995)
16. Walton, D., Krabbe, E.: *Commitment in Dialog: Basic Concepts of Interpersonal Reasoning*. State University of New York Press, NJ (1995)

XVIII Workshop Procesamiento Distribuido y Paralelo (WPDP)

Coordinadores

Printista, Marcela - De Giusti, Laura - García Garino, Carlos

Comité Científico

Ardenghi, Jorge (Argentina)	Marín, Mauricio (Chile)
Balladini, Javier (Argentina)	Martín, María J. (España)
Bertogna, Leandro (Argentina)	Naiouf, Marcelo (Argentina)
Bianchini, Germán (Argentina)	Olcoz Herrero, Katzalin (España)
Cancela, Héctor (Uruguay)	Pardo, Xoan (España)
Caymes Scutari, Paola (Argentina)	Pacini, Elina (Argentina)
Chichizola, Franco (Argentina)	Piccoli, Fabiana (España)
De Giusti, Armando (Argentina)	Piñuel, Luis (España)
De Giusti, Laura (Argentina)	Printista, Marcela (Argentina)
Denham, Mónica (Argentina)	Rexachs, Dolores (España)
Doallo, Ramón (España)	Rodríguez León, Casiano (España)
Fabero, Juan C. (España)	Rodríguez de Souza, Josemar (Brasil)
García Garino, Carlos (Argentina)	Suppi, Remo (España)
Gil Costa, Verónica (Argentina)	Tinetti, Fernando (Argentina)
Iacono, Lucas (Argentina)	Tirado, Francisco (España)
Laguna, Carlos Alfonso (España)	Touriño, Juan (España)
Luque, Emilio (España)	

Análisis de la precisión de la predicción energética de RAPL en una aplicación de multiplicación de matrices en memoria compartida

Juan Manuel Paniego¹, Silvana Gallo^{1,3}, Martin Pi Puig¹, Franco Chichizola¹, Laura De Giusti¹, Armando De Giusti^{1,3}, Javier Balladini²

¹Instituto de Investigación en Informática III LIDI, Facultad de Informática, Universidad Nacional de La Plata, La Plata, Argentina

²Departamento de Ingeniería de Computadoras, Facultad de Informática, Universidad Nacional del Comahue, Neuquén, Argentina

³CONICET, Facultad de Informática, Universidad Nacional de La Plata, La Plata, Argentina

{jspaniego, sgallo, mpipuig, francoch, ldgiusti, degiusti}@lidi.info.unlp.edu.ar
javier.balladini@fi.uncoma.edu.ar

Resumen. Durante los últimos años, el consumo energético ha surgido como uno de los temas centrales en el desarrollo de aplicaciones para HPC. El enfoque tradicional del cómputo paralelo y distribuido ha cambiado la perspectiva desde la búsqueda de la mayor eficiencia computacional a un enfoque que equilibre las prestaciones obtenidas con el consumo energético. En consecuencia, se han implementado diferentes métricas y mecanismos de medición de consumo para poder lograr este balance. En el presente artículo el objetivo se centra en el monitoreo y comparación del consumo energético de una determinada aplicación a través de mediciones físicas y de una interface software basada en contadores de hardware. Se presenta una comparación de los valores de energía obtenidos por el modelo de Intel, RAPL en relación a las mediciones físicas obtenidas desde la alimentación del procesador. Dichas mediciones se aplican durante la ejecución de una multiplicación de matrices clásica. Los resultados muestran que, para la aplicación analizada, la potencia media demandada por el procesador a partir de valores predichos por RAPL, presenta un error de hasta un 22%.

Palabras claves: Consumo energético, Predicción de consumo, Potencia, Contadores de hardware, RAPL, Perf.

1 Introducción

El estudio en el área de Cómputo de Altas Prestaciones (o HPC, High Performance Computing) tiene como objetivo el diseño y desarrollo de soluciones a problemas que requieren grandes capacidades de cómputo, mediante técnicas y métodos que permitan

aprovechar toda la potencia de cómputo disponible para obtener el máximo speedup y eficiencia sobre la arquitectura.

Durante los últimos años y debido a las limitaciones existentes en los recursos energéticos, se ha puesto mayor énfasis en la investigación relacionada a las consecuencias del incremento de temperatura de los procesadores y la tendencia ha sido analizar el consumo energético sobre las diferentes aplicaciones, así como también desarrollar modelos de consumo para la obtención de predicciones y para planificar la ejecución en base a las métricas de consumo energético [2, 3].

Actualmente, la eficiencia en cuanto al consumo es una métrica de gran importancia e impacto que afecta desde sistemas móviles hasta grandes centros destinados a HPC. Estos últimos implican un enorme consumo de energía diariamente y, consecuentemente, los valores se incrementan con el paso del tiempo. Por tanto, el monitoreo de la disipación de potencia y el consumo de energía de los sistemas se ha convertido en un área de prioridad.

Muchos de los grandes clusters que existen en la actualidad se encuentran constituidos por máquinas que son de propósito general y, sobre todo, de un muy bajo costo monetario. Al interconectar decenas, cientos o miles de estos dispositivos, se obtiene una gran capacidad de cómputo, pero a su vez, un consumo total muy significativo, ya que no solo involucra los nodos del cluster, sino también el costo de mantenerlos refrigerados, por ejemplo. Por lo que la monitorización de la energía utilizada para cada uno de los computadores es un punto de partida prioritario para reducir dicha variable.

Este artículo se enfoca en el monitoreo y comparación del consumo energético de una aplicación de multiplicación de matrices a través de mediciones físicas y de una interface software basada en contadores de hardware.

Este artículo se organiza de la siguiente manera: en la Sección 2 se recopilan los trabajos relacionados en el ámbito del consumo energético. Luego, en la Sección 3 se presentan algunos conceptos básicos de la medición de consumo energético. Seguidamente, en la Sección 4 se detalla la metodología utilizada para las mediciones. Luego, en la Sección 5 se describe el trabajo experimental realizado, y en la Sección 6 se encuentran los resultados obtenidos. Finalmente, en la Sección 7 se comentan las conclusiones y las líneas de trabajo futuras.

2 Trabajos Relacionados

El consumo energético en HPC ha surgido como uno de los temas de investigación principales durante los últimos años debido a la necesidad de minimizar el impacto ambiental de la infraestructura que necesitan los centros de cómputo, grandes servidores, sistemas de almacenamiento y refrigeración. Para poder llevar a cabo la gestión de este consumo se han implementado numerosas aplicaciones e instrumentos de medición como los utilizados para este trabajo.

Cuando se desea tener una referencia del consumo de todo un sistema de cómputo, lo que se denomina *medición de grano grueso* se observan diferentes instrumentos que interceptan la alimentación y aplicaciones que llevan a cabo la extracción de los datos correspondientes, por ejemplo, las mediciones externas con pinzas amperimétricas y

osciloscopios [2, 3], o dispositivos como Watt's Up Pro combinados con aplicativos que toman los datos capturados y generan información representativa del escenario de prueba como PowerPack [11].

En otros escenarios, es necesario conocer el consumo de un componente del sistema de cómputo en particular (memoria, disco, red, GPU). Allí es necesaria una medición de *grano fino*. La implementación de las mediciones físicas de este artículo para el consumo de CPU se inspiró en [9, 21], que con el uso de sensores de efecto Hall permiten la extracción de la información del dispositivo en cuestión.

Otros trabajos han considerado la medición de grano fino mediante sensores como invasiva y muy compleja, por lo que han estudiado modelos de estimación en base a la información de contadores hardware. Tal es el caso de [4, 6, 7, 15, 17] que realizan la estimación de potencia del sistema completo o de diferentes componentes con mediciones de grano fino. Luego, otros modelos se enfocan hacia las aplicaciones, con el objetivo de evaluar a qué frecuencia ejecutarlas [18, 5] o cómo planificarlas [22]. También se han desarrollado modelos más complejos que incorporan heterogeneidad y contemplan el consumo de GPU [20, 24] y hasta modelos con cierto tipo de inteligencia mediante *machine learning* [23].

La estimación de la potencia a través de modelos de grano fino que proveen los fabricantes de procesadores como *Running Average Power Limit* (RAPL) en el caso de Intel [8, 14] y *Application Power Management* (APM) para AMD [1], ha despertado el interés en conocer si dichos valores se correlacionan con las medidas físicas reales. Por este motivo, la confiabilidad de los registros contadores y del modelo de consumo han sido cuestionados y existen numerosos trabajos que buscan validar la estimación contrastando los valores obtenidos por el modelo contra los métodos de medición física antes mencionados [9, 12, 13].

Existen trabajos que estiman la potencia de grano grueso (potencia de corriente alterna del sistema completo) a partir del modelo de potencia de grano fino provisto por el fabricante del procesador [16, 4].

3 Medición de consumo

Para llevar a cabo la medición de la energía consumida durante un periodo de tiempo determinado, existen varios enfoques, sin embargo, sólo dos son ampliamente utilizados.

Por un lado, es posible monitorear el consumo energético a través del uso de ciertas interfaces de *medición software*, las cuales se encargan de alimentar a un modelo de potencia refinado con diferentes variables del sistema para finalmente obtener una medición estimada del consumo actual del dispositivo. En el caso de los procesadores de Intel, dichas variables se obtienen a través de los registros contadores de hardware de consumo que poseen la mayoría de los procesadores diseñados durante los últimos años, más específicamente desde la microarquitectura SandyBridge [14].

Por otra parte, el control del consumo a través de *medición física* externa al sistema hace uso de diferentes herramientas especialmente diseñadas para ello. Se pueden encontrar trabajos donde se efectúa un muestreo de potencia del sistema completo a través de una pinza amperimétrica conectada a un osciloscopio digital [2, 3]. En otros

casos, se utilizan técnicas invasivas donde se añade sensores de corriente (por efecto Hall) en serie con las líneas de alimentación de los distintos dispositivos de una computadora [11].

3.1 Medición Software

Durante muchos años, los fabricantes de microprocesadores instalaron registros ocultos destinados a la depuración de problemas en el hardware, pero no fue hasta 1993 que se hizo pública su existencia y las posibilidades que proveen para conocer datos muy precisos acerca del estado del procesador. Los contadores de hardware son un conjunto de registros de propósito especial incorporados en la mayoría de los microprocesadores modernos cuya función es almacenar en forma acumulativa ciertos eventos relacionados al hardware del sistema. En los últimos diseños de Intel, se han incorporado eventos específicos para la medición de consumo.

Debido al potencial que ofrecen estos contadores, se han desarrollado herramientas que permiten especificar qué evento se desea monitorear ya sea para el profiling de aplicaciones ó para realizar ajustes durante la ejecución de algoritmos.

En el caso del profiling de aplicaciones, *perf* surge como una interfaz de GNU/Linux para visualizar un gran número de eventos utilizados en profiling y los hace accesibles desde línea de comandos. Los eventos disponibles dependen tanto de la arquitectura utilizada como del kernel Linux con el que se cuenta. Ante la complejidad de modificar o conocer los valores de los contadores en tiempo real, surge PAPI [19], una interfaz para acceder a los diferentes eventos y los registros contadores desde el código fuente de un algoritmo, de allí que surge también la posibilidad de alterar el comportamiento de una aplicación en base a los valores obtenidos mediante la herramienta.

Como se mencionó anteriormente, la incorporación de los eventos de consumo desencadenó la investigación relacionada a la estimación del consumo de los diferentes componentes de un sistema de cómputo, dando origen a modelos de estimación de consumo como RAPL [8]. RAPL es en sí un modelo de consumo que estima la energía utilizando contadores de hardware y modelos de E/S. En RAPL, existen dominios de medición, lo cual garantiza monitoreos de grano fino y pueden variar según el modelo de procesador.

El dominio Package (PKG0) comprende la potencia consumida por todo el socket. Luego, PP0 abarca todo el subsistema de cores y cachés, mientras que PP1 contempla el consumo de la circuitería externa a los cores, como por ejemplo la GPU *onboard*. Por último, el dominio DRAM genera información acerca de la energía consumida por la memoria principal. Como el dominio PKG0 agrupa a los subdominios PP0 y PP1, es el seleccionado para realizar la medición de energía de este trabajo.

3.2 Medición Física

Al momento de realizar el monitoreo de potencia a nivel físico se distinguen dos procedimientos: uno de grano grueso y otro de grano fino. En el primer caso, el método consiste en interceptar, mediante un instrumento adecuado, las variables necesarias para el cálculo de la potencia instantánea: la corriente y la tensión alterna. Luego, en la

solución de grano fino, se monitorea la corriente instantánea que atraviesa la alimentación de los principales dispositivos del sistema.

Este artículo centra su trabajo sobre el segundo procedimiento. Para ello, hace uso de varios dispositivos. En primer lugar, una placa de desarrollo Arduino UNO, basada en hardware libre y que cuenta con el procesador ATmega328P, 14 I/O digitales y 6 I/O analógicas y un reloj de 16MHz de frecuencia. Luego, un par de sensores ACS712, que se utilizarán para la medición de la corriente mediante efecto Hall invasivo. Esta clase de sensor cuenta con un rango de -5 a 5 Amperes y una resolución de 185mV/A, alimentado con 5V, puede apreciarse en la Figura 1a. Por otro lado, el equipo al cuál se desea realizar la medición debe estar provisto de una fuente con conexión ATX 12V de 4 pines (ver Figura 1b), que se utiliza para alimentar principalmente a la CPU mediante el conector visible en la Figura 1b.

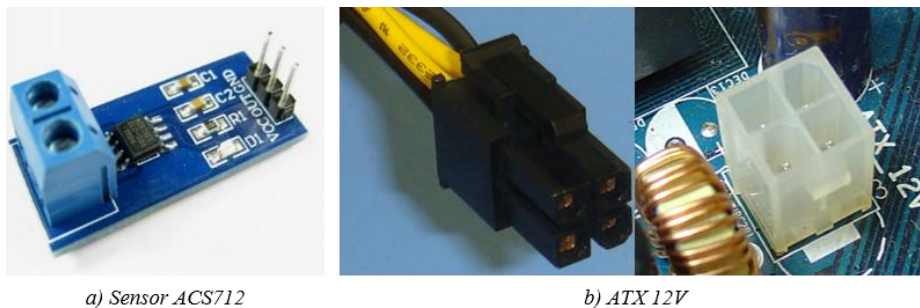


Fig. 1. Dispositivos necesarios para la medición física.

4 Implementación del Sistema de Medición

Para llevar a cabo la medición comparativa son necesarios dos equipos, un *equipo de estudio* en el cuál se realiza la ejecución del algoritmo objetivo de la medición y, por otra parte, el equipo que captura y permite la monitorización de las medidas físicas, al que se denomina *equipo de monitoreo*.

En la Figura 2 se muestra un esquema simple de la conexión. Los sensores anteriormente mencionados, se conectan en serie a la alimentación ATX 12V, luego la placa de desarrollo se encarga de la lectura de los valores y del envío de estas mediciones al equipo de monitoreo. Sobre este último, se ejecuta un script Python que se encarga de recibir los datos y almacenarlos en disco.

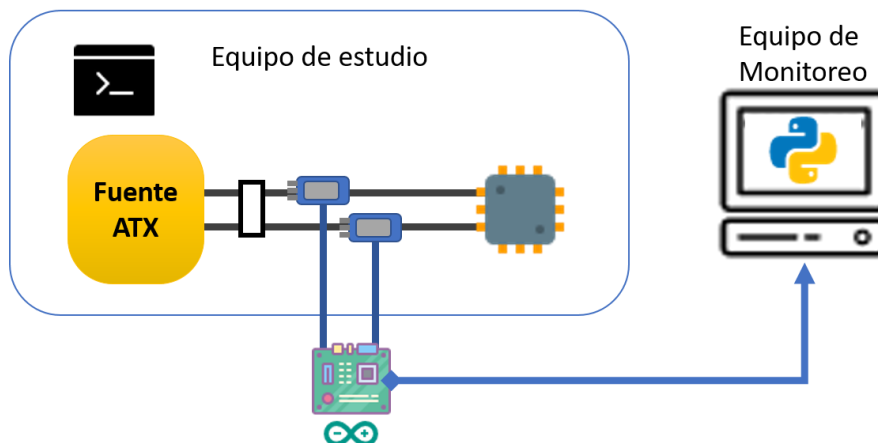


Fig. 2. Esquema de medición.

Finalmente, la medición a nivel software se realiza simultáneamente con el muestreo físico, a través de la herramienta *perf*. Se debe tener en cuenta que es posible realizar ambos muestreos simultáneamente debido a que *perf* es una herramienta sumamente liviana y, por tanto, no introduce un overhead significativo. La sincronización de ambos equipos fue llevada a cabo mediante el uso de un socket TCP.

5 Experimentación

Las mediciones fueron realizadas sobre el equipo denominado “Equipo de estudio” compuesto por un Motherboard Gigabyte GA-H81M-H, Procesador Intel Core I3-4170 (2 cores), 8GB DDR3 y con sistema operativo Ubuntu 14.04 (Kernel 3.16). El equipo con el cuál se realizaron las mediciones y monitoreo de los valores requiere únicamente de un intérprete Python para la obtención de los valores censados.

Por otro lado, se eligió la multiplicación de matrices (1), como caso de estudio, ya que representa un problema ampliamente tratado y que en muchos casos es parte de soluciones a problemas de mayor envergadura.

$$C_{i,j} = \sum_{k=1}^p A_{i,k} * B_{k,j} \quad (1)$$

Los experimentos se llevaron a cabo para la solución al algoritmo en cuestión en su implementación secuencial y multihilo a través de la librería OpenMP a una frecuencia de trabajo de 1.6GHz.

6 Resultados

En esta sección, se detallan los resultados correspondientes a las mediciones físicas y software. Se realiza un breve análisis gráfico y descriptivo de los datos obtenidos. Inicialmente, se muestran los resultados obtenidos para la ejecución de las soluciones secuenciales, variando la carga de trabajo. Es importante destacar que tanto la medición física como la medición software, se realizan con un script de ejecución que incluye un segmento de inactividad (contemplado en los resultados) mediante sleep antes y luego de la ejecución del algoritmo, de la siguiente manera:

```

_> Sleep 2s (Idle State)
_> Ejecución del programa
_> Sleep 2s (Idle State)

```

Un primer experimento (ver Figura 3), muestra la potencia a lo largo de la ejecución de la versión secuencial con una matriz de 2048x2048.

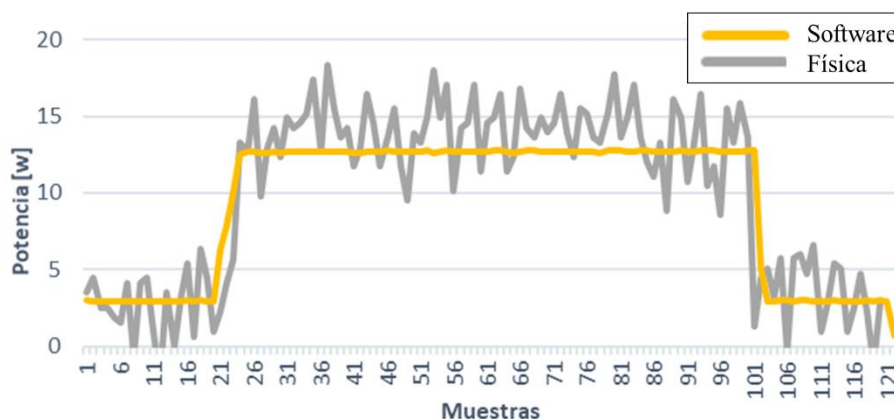


Fig. 3. Perfil de consumo, ejecución secuencial.

Se puede apreciar que la forma de las curvas es similar entre sí. Cabe aclarar que la medición física se observa “ruidosa” debido a que se muestran datos instantáneos (no es un promedio de diferentes muestras censadas).

A continuación, se muestran los valores de potencia promedio consumida por el procesador a lo largo de la ejecución. La Tabla 1 contiene los valores para la potencia activa de las aproximaciones física y software, al variar la carga de trabajo del algoritmo secuencial.

Tabla 1. Potencia activa (en Watts) para la solución secuencial.

	1024	2048	4096	8192
Medición software	12,35	12,70	13,07	12,90
Medición física	13,69	13,97	14,38	14,34

Es claro evidenciar que la medición física retorna un valor mayor al obtenido por la medición software. Esta desviación es lineal y presenta una variación cercana al 10% (ver Figura 4).

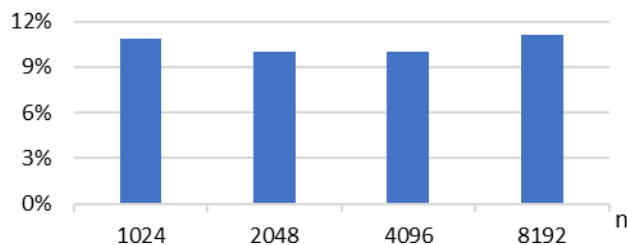


Fig. 4. Porcentaje de error para la implementación secuencial.

Por otro lado, las soluciones con OpenMP utilizando 2, 4 y 8 hilos arrojan los resultados que se pueden apreciar en la Figura 5 (a, b y c respectivamente).

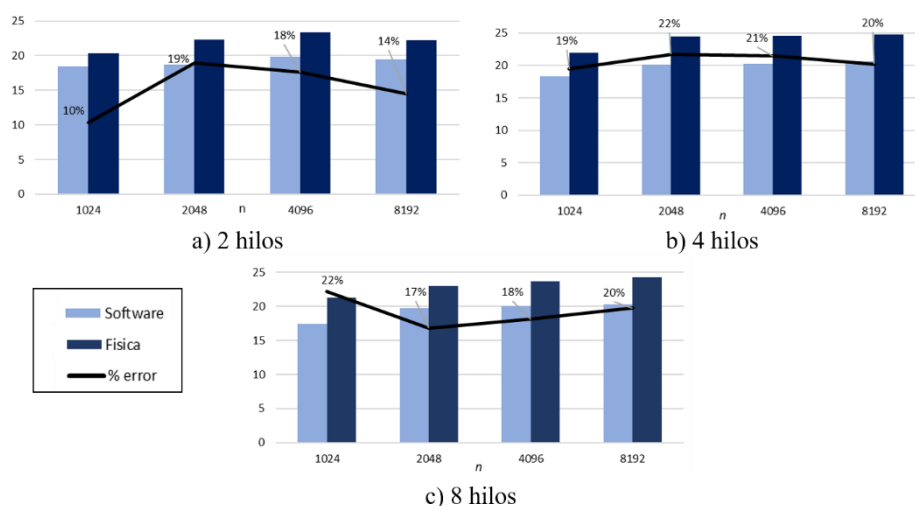


Fig. 5. Potencia media para las soluciones con hilos (en Watts). También se muestra el error entre la medición software y física.

Desde los valores obtenidos, puede deducirse un aumento de la diferencia entre mediciones de entre un 10% y 22%.

7 Conclusiones y Líneas de Trabajo Futuro

Este estudio se centró el monitoreo y comparación del consumo energético de una aplicación que resuelve multiplicaciones de matrices a través de mediciones físicas y de una interface software basada en contadores de hardware.

Se construyó un entorno de medición físico compuesto por un arduino con sensores de corriente que permitieron tomar mediciones reales de la potencia media demandada por el procesador y, conjuntamente se implementó un script capturador de datos en Python para poder almacenar dichas mediciones. Por otro lado, se utilizó *perf* para poder acceder a los valores del modelo RAPL para el procesador Intel evaluado.

De los resultados, se observó una desviación lineal entre los valores de potencia media de medición física y software, siendo la primera la mayor en la totalidad de los casos evaluados. Se cree que la diferencia entre las mediciones también se ve afectada por alguno o varios dispositivos hardware, de los cuales, su consumo por software, no se ha analizado. Además, se detectó un error variable de entre el 10% y hasta el 22%, en experimentaciones de multiplicaciones de matrices, variando el tamaño del problema y la cantidad de threads utilizados.

Finalmente, se plantea como trabajo futuro la detección y análisis de la brecha entre las mediciones software y física, así como también la correlación de los métodos de medición utilizados frente a la medición AC externa. Luego de validar contra la medición externa, resulta de gran interés evaluar el consumo de otros componentes en el sistema, como son la memoria RAM y la E/S que no han sido contemplados en este artículo.

Referencias

1. AMD, AMD Family 15th Processor BIOS and Kernel Developer Guide (2011)
2. Balladini J., Muresano R., Suppi R., Rexachs D., Luque E.: Methodology for predicting the energy consumption of SPMD application on virtualized environments. En *Journal of Computer Science & Technology*. pp. 130-136. 13. 3 (2013)
3. Balladini J., Rucci E., De Giusti A., Naiouf M., Suppi R., Rexachs D., Luque E.: Power Characterisation of Shared-Memory HPC Systems. En *Computer Science & Technology Series. XVIII Argentine Congress of Computer Science Selected Papers*. pp. 53-65 (2013)
4. Bircher W., John L.: Complete system power estimation using processor performance events. *IEEE Tran. on Comp.* 61. 4 (2012)
5. C. Lively, X. Wu, V. Taylor, S. Moore, H. Chang, C. Su, Cameron K.: Power-Aware Predictive Models of Hybrid (MPI/OpenMP) Scientific Applications on Multicore System. *Computer Science – Res. and Dev.* 27. 4 (2012)
6. Chen X., Xu C., Dick R., Mao Z.: Performance and Power Modeling in a Multi-Programmed Multi-Core Environment. *DAC2010* (2010)
7. Contreras G., Martonosi M.: Power Prediction for Intel XScale Processors Using Performance Monitoring Unit Events. *ISLPED'05* (2005)
8. David H., Gorbatov E.: RAPL: Memory Power Estimation and Capping. In: *2010 ACM/IEEE International Symposium on Low-Power Electronics and Design (ISLPED)*. pp. 189-194 (2010)
9. Desrochers S., Paradis C., Weaver V. M.: A validation of DRAM RAPL Power Measurements. *Second International Symposium on Memory Systems* (2016)
10. Diouri M. E. M., Dolz M.: Assessing Power Monitoring Approaches for Energy and Power Analysis of Computers. *Sustainable Computing: Informatics and Systems*, Elsevier. pp. 68-82. 4. 2 (2014)
11. Ge R., Feng X., Song S., Chang H. C., Li D., K. W. Cameron: PowerPack: Energy Profiling and Analysis of High-Performance Systems and Applications. *IEEE Transactions on Parallel and Distributed Systems*. pp. 658-671. 21. 5 (2010)

12. Hackenberg D., Ilsche I., Schone R., Molka D., Schmidt M., Nagel W. E.: Power Measurement Techniques on Standard Compute Nodes: A Quantitative Comparison. IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS) (2013)
13. Hahnel M., Dobel B., Volp M., Hartigl H.: Measuring Energy Consumption for Short Code Paths Using RAPL. Technische Universität Dresden. ACM SIGMETRICS Performance Evaluation Review Volume 40. pp. 13-17 (2012)
14. Intel Architecture Software Developer's Manual, volumen 3: System Programming Guide, (2009)
15. Isci C., Martonosi M.: Runtime Power Monitoring in High-end Processors: Methodology and Empirical Data. 36th IEEE/ACM International Symposium on Microarchitecture (2003)
16. Khan K. N., Ou Z., Hirki M., Nurminen J. K., Niemi T.: How much power does your server consume? Estimating wall socket power using RAPL measurements. Computer Science - Research and Development. pp. 207-214 (2016)
17. Lim M., Porterfield A., Fowler R.: SoftPower: Finegrain Power Estimations Using Performance Counters. 19th Int'l Symp High Perf. Dis. Comp. (HPDC '10) (2010)
18. Lively C., Taylor V., Wu X., Chang H., Su C., Cameron K., Moore S., Terpstra D.: E-AMOM: An Energy-Aware Modeling and Optimization Methodology for Scientific Applications on Multicore Systems, Computer Science – Res. and Dev. 29. 3 (2014)
19. Mucci P., Browne S., Deane C., Ho G.: PAPI: A Portable Interface to Hardware Performance Counters. En: Proceedings of the Department of Defense HPCMP Users Group Conference. pp. 7-10 (1999)
20. Nagasaka H., Maruyama N., Nukada A., Endo T., Matsuoka S.: Statistical Power Modeling of GPU Kernels Using Performance Counters, Intern. Green Computing Conference (2010)
21. Picariello F., Rapuano S., Villano U.: Evaluation of Power Consumption of Workstation Computers using Benchmarking. En: Proceedings 12th IMEKO TC10 Workshop on Technical Diagnostics: New Perspective in Measurements, Tools and Techniques for Industrial Applications University of Sannio, Italy. pp. 242-247 (2013)
22. Singh K., Bhadhauria M., McKee S. A.: Real Time Power Estimation and Thread Scheduling via Performance Counters. Workshop on Design, Architecture and Simulation of Chip Multi-Processors (2008)
23. Song S., Su C., Rountree B., Cameron K.: A Simplified and Accurate Model of Power-Performance Efficiency on Emergent GPU Architectures. IPDPS2013 (2013)
24. Tsafack Chetsa G.L., Lefevre L., Pierson J.M., Stolf P., Da Costa G.: Exploiting Performance Counters to Predict and Improve Energy Performance of HPC Systems. Future Gen. Computer Systems. 36 (2014)

Blocked All-Pairs Shortest Paths Algorithm on Intel Xeon Phi KNL Processor: A Case Study

Enzo Rucci¹, Armando De Giusti¹, and Marcelo Naiouf²

¹ III-LIDI, CONICET, Facultad de Informática, Universidad Nacional de La Plata
La Plata (1900), Buenos Aires, Argentina
Email: {erucci,degiusti}@lidi.info.unlp.edu.ar

² III-LIDI, Facultad de Informática, Universidad Nacional de La Plata
La Plata (1900), Buenos Aires, Argentina
Email: {mnaiouf}@lidi.info.unlp.edu.ar

Abstract. Manycores are consolidating in HPC community as a way of improving performance while keeping power efficiency. Knights Landing is the recently released second generation of Intel Xeon Phi architecture. While optimizing applications on CPUs, GPUs and first Xeon Phi's has been largely studied in the last years, the new features in Knights Landing processors require the revision of programming and optimization techniques for these devices. In this work, we selected the Floyd-Warshall algorithm as a representative case study of graph and memory-bound applications. Starting from the default serial version, we show how data, thread and compiler level optimizations help the parallel implementation to reach 338 GFLOPS.

Keywords: Xeon Phi, Knights Landing, Floyd-Warshall

1 Introduction

The power consumption problem represents one of the major obstacles for Exascale systems design. As a consequence, the scientific community is searching for different ways to improve power efficiency of High Performance Computing (HPC) systems [8]. One recent trend to increase compute power and, at the same time, limit power consumption of these systems lies in adding accelerators, like NVIDIA/AMD graphic processing units (GPUs), or Intel Many Integrated Core (MIC) co-processors. These manycore devices are capable of achieving better FLOPS/Watt ratios than traditional CPUs. For example, the number of Top500 [2] systems using accelerator technology grew from 54 in June 2013 to 91 in June 2017. In the same period, the number of systems based on accelerators increased from 55 to 90 on the Green500 list [1].

Recently, Intel has presented the second generation of its MIC architecture (branded Xeon Phi), codenamed Knights Landing (KNL). Among the main differences of KNL regarding its predecessor Knights Corner (KNC), we can find the incorporation of AVX-512 extensions, a remarkable number of vector units increment, a new on-package high-bandwidth memory (HBM) and the ability

to operate as a standalone processor. Even though optimizing applications on CPUs, GPUs and KNC Xeon Phi's has been largely studied in the last years, accelerating applications on KNL processors is still a pending task due to its recent commercialization. In that sense, the new features in KNL processors require the revision of programming and optimization techniques for these devices.

In this work, we selected the Floyd-Warshall (FW) algorithm as a representative case study of graph and memory-bound applications. This algorithm finds the shortest paths between all pairs of vertices in a graph and occurs in domains of communication networking [14], traffic routing [12], bioinformatics [16], among others. FW is both computationally and spatially expensive since it requires $O(n^3)$ operations and $O(n^2)$ memory space, where n is the number of vertices in a graph. Starting from the default serial version, we show how data, thread and compiler level optimizations help the parallel implementation to reach 338 GFLOPS.

The rest of the present paper is organized as follows. Section 2 briefly introduces the Intel Xeon Phi KNL architecture while Section 3 presents the FW algorithm. Section 4 describes our implementation. In Section 5 we analyze performance results while Section 6 discusses related works. Finally, Section 7 outlines conclusions and future lines of work.

2 Intel Xeon Phi Knights Landing

KNL is the second generation of the Intel Xeon Phi family and the first capable of operating as a standalone processor. The KNL architecture is based on a set of *Tiles* (up to 36) interconnected by a 2D mesh. Each Tile includes 2 cores based on the out-of-order Intel's Atom micro-architecture (4 threads per core), 2 Vector Processing Units (VPUs) and a shared L2 cache of 1 MB. These VPUs not only implement the new 512-bit AVX-512 ISA but they are also compatible with prior vector ISA's such as SSE x and AVX x . AVX-512 provides 512-bit SIMD support, 32 logical registers, 8 new mask registers for vector predication, and gather and scatter instructions to support loading and storing sparse data. As each AVX-512 instruction can perform 8 double-precision (DP) operations (8 FLOPS) or 16 single-precision (SP) operations (16 FLOPS), the peak performance is over 1.5 TFLOPS in DP and 3 TFLOPS in SP, more than two times higher than that of the KNC. It is also more energy efficient than its predecessor [17].

Other significant feature of the KNL architecture is the inclusion of an in-package HBM called MCDRAM. This special memory offers 3 operating modes: *Cache*, *Flat* and *Hybrid*. In *Cache* mode, the MCDRAM is used like an L3 cache, caching data from the DDR4 level of memory. Even though application code remains unchanged, the MCDRAM can suffer lower performance rates. In *Flat* mode, the MCDRAM has a physical addressable space offering the highest bandwidth and lowest latency. However, software modifications may be required in order to use both the DDR and the MCDRAM in the same application. Finally, in the *Hybrid mode*, HBM is divided in two parts: one part in *Cache mode* and one in *Flat mode* [5].

From a software perspective, KNL supports parallel programming models used traditionally on HPC systems such as OpenMP or MPI. This fact represents a strength of this platform since it simplifies code development and improves portability over other alternatives based on accelerator specific programming languages such as CUDA or OpenCL. However, to achieve high performance, programmers should attend to:

- the efficient exploitation of the memory hierarchy, especially when handling large datasets, and
- how to structure the computations to take advantage of the VPU.

Automatic vectorization is obviously the easiest programming way to exploit VPUs. However, in most cases the compiler is unable to generate SIMD binary code since it can not detect free data dependences into loops. In that sense, SIMD instructions are supported in KNL processors through the use of guided compilation or hand-tuned codification with intrinsic instructions [17]. On one hand, in guided vectorization, the programmer indicates the compiler (through the insertion of tags) which loops are independent and their memory pattern access. In this way, the compiler is able to generate SIMD binary code preserving the program portability. On the other hand, intrinsic vectorization usually involves rewriting most of the corresponding algorithm. The programmer gains in control at the cost of losing portability. Moreover, this approach also suggests the inhibition of other compiler loop-level optimizations. Nevertheless, it is the only way to exploit parallelism in some applications with no regular access patterns or loop data dependencies which can be hidden by recomputing techniques [6].

3 Floyd-Warshall Algorithm

The FW algorithm uses a dynamic programming approach to compute the all-pairs shortest-paths problem on a directed graph [7, 20]. This algorithm takes as input a $N \times N$ distance matrix D , where $D_{i,j}$ is initialized with the original distance from node i to node j . FW runs for N iterations and at k -th iteration it evaluates all the possible paths between each pair of vertices from i to j through the intermediate vertex k . As a result, FW produces an updated matrix D , where $D_{i,j}$ now contains the shortest distance between nodes i and j . Besides, an additional matrix P is generated when the reconstruction of the shortest path is required. $P_{i,j}$ contains the most recently added intermediate node between i and j . Figure 1 exhibits the naive FW algorithm.

4 Implementation

In this section, we address the optimizations performed on the Intel Xeon Phi KNL processor. First of all, we developed a serial implementation following the naive version described in Figure 1, as this implementation will work as baseline. Next, we optimized the serial version considering data locality and data level parallelism. Finally, we introduced thread level parallelism exploiting the OpenMP programming model to obtain a multi-threaded implementation.

4.1 Data Locality

To improve data locality, the FW algorithm can be blocked [19]. Unfortunately, the three loops can not be interchanged in free manner due to the data dependencies from one iteration to the next in the k -loop (just i and j loops can be done in any order). However, under certain conditions, the k -loop can be put inside the i -loop and j -loop, making blocking possible. The distance matrix D is partitioned into blocks of size $BS \times BS$, so that there are $(N/BS)^2$ blocks. The computations involve $R = N/BS$ rounds and each round is divided into four phases based on the data dependency among the blocks:

1. Update the block k,k ($D^{k,k}$) because it is self-dependent.
2. Update the remaining blocks of the k -th row because each of these blocks depends on itself and the previously computed $D^{k,k}$.
3. Update the remaining blocks of the k -th column because each of these blocks depends on itself and the previously computed $D^{k,k}$.
4. Update the rest of the matrix blocks as each of them depends on the k -th block of its row and the k -th block of its column.

In this way, we satisfy all dependencies from this algorithm. Figure 2 shows a schematic representation of a round computation and the data dependences among the blocks while Figure 3 presents the corresponding pseudo-code.

4.2 Data Level Parallelism

The innermost loop of FW_BLOCK code block from Figure 3 is clearly the most computationally expensive part of the algorithm. In that sense, this loop is the best candidate for vectorization. The loop body is composed of an *if* statement that involves one addition, one comparison and (may be) two assign operations. Unfortunately, the compiler detects false dependencies in that loop and is not able to generate SIMD binary code. For that reason, we have explored two SIMD exploitation approaches: (1) guided vectorization through the usage

```

1: for k = 0..N-1 do
2:   for i = 0..N-1 do
3:     for j = 0..N-1 do
4:       if  $D_{i,k} + D_{k,j} < D_{i,j}$  then
5:          $D_{i,j} = D_{i,k} + D_{k,j}$ 
6:          $P_{i,j} = k$ 
7:       end if
8:     end for
9:   end for
10: end for

```

Fig. 1: Naive Floyd-Warshall Algorithm

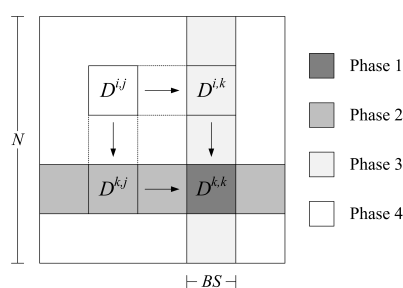


Fig. 2: Schematic representation of the blocked Floyd-Warshall Algorithm


```

1: function FW_BLOCK ( $D^1, D^2, D^3, P, base$ )
2:   for  $k = 0..BS-1$  do
3:     for  $i = 0..BS-1$  do
4:       for  $j = 0..BS-1$  do
5:         if  $D^2_{ik} + D^3_{kj} < D^1_{ij}$  then
6:            $D^1_{ij} = D^2_{ik} + D^3_{kj}$ 
7:            $P_{ij} = base+k$ 
8:         end if
9:       end for
10:    end for
11:  end for
12: end function

13: for  $k = 0..R-1$  do
14:    $b = k \times BS$ 
15:   # Phase 1
16:   FW_BLOCK( $D^{k,k}, D^{k,k}, D^{k,k}, P^{k,k}, b$ )
17:   # Phase 2
18:   for  $j = 0..R-1, j \neq k$  do
19:     FW_BLOCK( $D^{k,j}, D^{k,k}, D^{k,j}, P^{k,j}, b$ )
20:   end for
21:   # Phase 3
22:   for  $i = 0..R-1, i \neq k$  do
23:     FW_BLOCK( $D^{i,k}, D^{i,k}, D^{i,k}, P^{i,k}, b$ )
24:   end for
25:   # Phase 4
26:   for  $i, j = 0..R-1, i, j \neq k$  do
27:     FW_BLOCK( $D^{ij}, D^{i,k}, D^{k,j}, P^{ij}, b$ )
28:   end for
29: end for

```

Fig. 3: Blocked Floyd-Warshall algorithm.

```

1: for  $k = 0..BS-1$  do
2:   for  $i = 0..BS-1$  do
3:     #pragma unroll
4:     #pragma omp simd
5:     for  $j = 0..BS-1$  do
6:       if ( $D^2[i][k] + D^3[k][j] < D^1[i][j]$ ) then
7:          $D^1[i][j] = D^2[i][k] + D^3[k][j]$ ;
8:          $P[i][j] = base+k$ ;
9:       end if
10:    end for
11:  end for
12: end for

```

Fig. 4: Pseudo-code for FW_BLOCK implementation using guided vectorization

```

1: for  $k = 0..BS-1$  do
2:    $bk = \_mm512\_set1\_epi32(base+k)$ ;
3:   for  $i = 0..BS-1$  do
4:      $dik = \_mm512\_set1\_ps(D^2[i][k])$ ;
5:     #pragma unroll
6:     for  $j = 0..BS-1$  by 16 do
7:        $dij = \_mm512\_load\_ps(&D^1[i][j])$ ;
8:        $dkj = \_mm512\_load\_ps(&D^3[k][j])$ ;
9:        $sum = \_mm512\_add\_ps(dik, dkj)$ ;
10:       $mask = \_mm512\_cmp\_ps\_mask(sum, dij, CMP_LT_OS)$ ;
11:       $\_mm512\_mask\_store\_ps(&D^1[i][j], mask, sum)$ ;
12:       $\_mm512\_mask\_store\_epi32(&P[i][j], mask, bk)$ ;
13:    end for
14:  end for
15: end for

```

Fig. 5: Pseudo-code for FW_BLOCK implementation using intrinsic vectorization

of the OpenMP 4.0 *simd* directive and (2) intrinsic vectorization employing the AVX-512 extensions. The guided approach simply consists of inserting the *simd* directive to the innermost loop of FW_BLOCK code block (line 4). On the opposite sense, the intrinsic approach consists of rewriting the entire loop body. Figures 4 and 5 show the pseudo-code for FW_BLOCK implementation using guided and manual vectorization, respectively. In order to accelerate SIMD computation with 512-bit vectors, we have carefully managed the memory allocations so that distance and path matrices are 64-byte aligned. In the guided approach, this also requires adding the *aligned* clause to the *simd* directive.

4.3 Loop Unrolling

Loop unrolling is another optimization technique that helped us to improve the code performance. Fully unrolling the innermost loop of FW_BLOCK code block

was found to work well. Unrolling the i -loop of the same code block once was also found to work well.

4.4 Thread Level Parallelism

To exploit parallelism across multiple cores, we have implemented a multi-threaded version of FW algorithm based on OpenMP programming model. A *parallel* construct is inserted before the loop of line 13 in Figure 3 to create a parallel block. To respect data dependencies among the block computations, the work-sharing constructs must be carefully inserted. At each round, phase 1 must be computed before the rest. So a *single* construct is inserted to enclose line 16. Next, phases 2 and 3 must be computed before phase 4. As these blocks are independent among them, a *for* directive is inserted before the loops of lines 18 and 22. Besides, a *nowait* clause is added to the phase 2 loop to alleviate the thread idling. Finally, another *for* construct is inserted before the loop of line 26 to distribute the remaining blocks among the threads.

5 Experimental Results

5.1 Experimental Design

All tests have been performed on an Intel server running CentOS 7.2 equipped with a Xeon Phi 7250 processor 68-core 1.40GHz (4 hw thread per core and 16GB MCDRAM memory) and 48GB main memory. The processor was run in *Flat* memory mode and *Quadrant* cluster mode.

We have used Intel's ICC compiler (version 17.0.1.132) with the $-O3$ optimization level. To generate explicit AVX2 and AVX-512 instructions, we employed the $-xAVX2$ and $-xMIC-AVX512$ flags, respectively. Also, we used the *numactl* utility to exploit MCDRAM memory (no source code modification is required). Besides, different workloads were tested: $N = \{4096, 8192, 16384, 32768, 65536\}$.

5.2 Performance Results

First, we evaluated the performance improvements of the different optimization techniques applied to the naive serial version, such as blocking (*blocked*), data level parallelism (*simd*, *simd (AVX2)* and *simd (AVX-512)*), aligned access (*aligned*) and loop unrolling (*unrolled*). Table 1 shows the execution time (in seconds) of the different serial versions when $N=4096$. As it can be observed, blocking optimization reduces execution time by 5%. Regarding the block size, 256×256 was found to work best. In the most memory demanding case of each round (phase 4), four blocks are loaded into the cache (3 distance blocks and 1 path block). The four blocks requires $4 \times 256 \times 256 \times 4$ bytes = 1024 KB = 1MB, which is exactly the L2 cache size.

As stated in Section 4.2, the compiler is not able to generate SIMD binary code by itself in the blocked version. Adding the corresponding *simd* constructs

Table 1: Execution time (in seconds) of the different optimization techniques applied to the naive serial version when $N=4096$.

<i>naive</i>	<i>blocked</i>	<i>simd</i>	<i>simd (AVX2)</i>	<i>simd (AVX-512)</i>	<i>aligned</i>	<i>unrolled</i>
602.8	572.66	204.52	100.47	36.95	33.28	22.95

to the blocked version reduced the execution time from 572.66 to 204.52 seconds, which represents a speedup of $2.8\times$. However, AVX-512 instructions can perform 16 SP operations at the same time. After inspecting the code at assembly level, we realized that the compiler generates SSE x instructions by default. As SSE x can perform 4 SP operations at the same time, the $2.8\times$ speedup has more sense since not all the code can be vectorized. Next, we re-compiled the code including the `-xAVX2` and `-xMIC-AVX512` flags to force the compiler to generate AVX2 and AVX-512 SIMD instructions, respectively. AVX2 extensions accelerated the blocked version by a factor of $5.8\times$ while AVX-512 instructions achieved a speedup of $15.5\times$. So, it is clear that this application benefits from larger SIMD width. In relation to the other optimization techniques employed, we have found that the *simd (AVX-512)* implementation runs $1.11\times$ faster when aligning memory accesses in AVX-512 computations (*aligned*). Additionally, applying the loop unrolling optimization to the *aligned* version led to higher performance, gaining a $1.45\times$ speedup. In summary, we achieve a $26.3\times$ speedup over the naive serial version through the combination of the different optimizations described.

Taking the optimized serial version, we developed a multi-threaded implementation as described in Section 4.4. Figure 6 shows the performance (in terms of GFLOPS) for the different affinity types used varying the number of threads when $N=8192$. As expected, *compact* affinity produced the worst results since it favours using all threads on a core before using other cores. *Scatter* and *balanced* affinities presented similar performances improving the *none* counterpart. As the KNL processor used in this study has all its cores in the same package, *scatter* and *balanced* affinities distribute the threads in the same manner when one thread per core is assigned. Regarding the number of threads, using a single thread per core is enough to get maximal performance (except in *compact* affinity). This behavior is opposed to the KNC generation where two or more threads per core were required to achieve high performance. However, it should not be a surprise since the KNL cores were designed to optimize single thread performance including out-of-order pipelines and two VPUs per core.

It is important to remark that, unlike the optimized serial version, the parallel implementation used a smaller block size since it delivered higher performance. A smaller block size allowed a finer-grain workload distribution and decreased thread idling, especially when the number of threads was larger than the number of blocks in phases 2 and 3. Another reason to decrease block size was that the L2 available space is now shared between the threads in a tile, contrary to the single threaded case. In particular, $BS=64$ was found to work best.

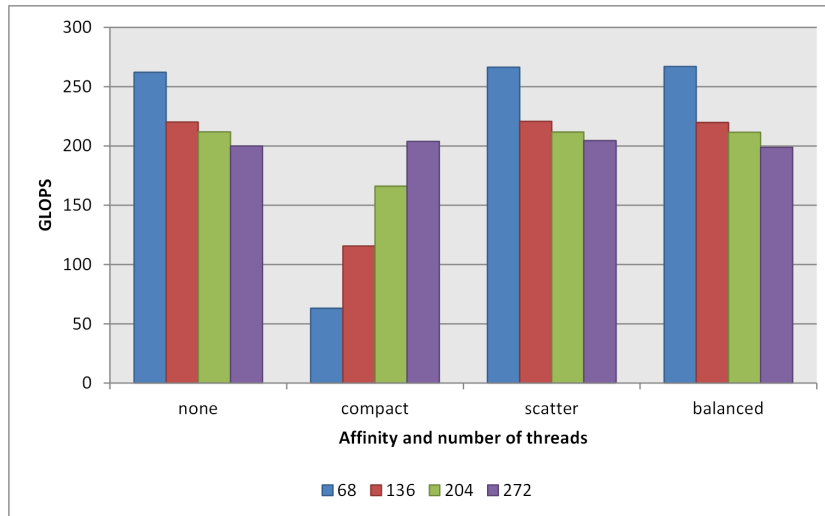


Fig. 6: Performance for the different affinity types used varying the number of threads when $N=8192$.

Figure 7 illustrates performance evolution varying workload and MCDRAM exploitation for the different vectorization approaches. For small workloads ($N = 8192$), the performance improvement is little ($\sim 1.1\times$). However, MCDRAM memory presents remarkable speedups for greater workloads, even when the dataset largely exceeds the MCDRAM size ($N = 655536$). In particular, MCDRAM exploitation achieves an average speedup of $9.8\times$ and a maximum speedup of $15.5\times$. In this way, we can see how MCDRAM usage is an efficient strategy for bandwidth-sensitive applications.

In relation to the vectorization approach, we can appreciate that guided vectorization leads to slightly better performance than the intrinsic counterpart, running up to $1.03\times$ faster. The best performances are 330 and 338 GFLOPS for the intrinsic and guided versions, respectively. After analyzing the assembly code, we realized that this difference is caused by the prefetching instructions introduced by the compiler when guided vectorization is used. Unfortunately, the compiler disables automatic prefetching when code is manually vectorized.

6 Related Works

Despite its recent commercialization, there are some works that evaluate KNL processors. In that sense, we highlight [18] that presents a study of the performance differences observed when using the three MCDRAM configurations available in combination with the three possible memory access or cluster modes. Also, Barnes et al. [3] discussed the lessons learned from optimizing a number of different high-performance applications and kernels. Besides, Haidar et al. [9]

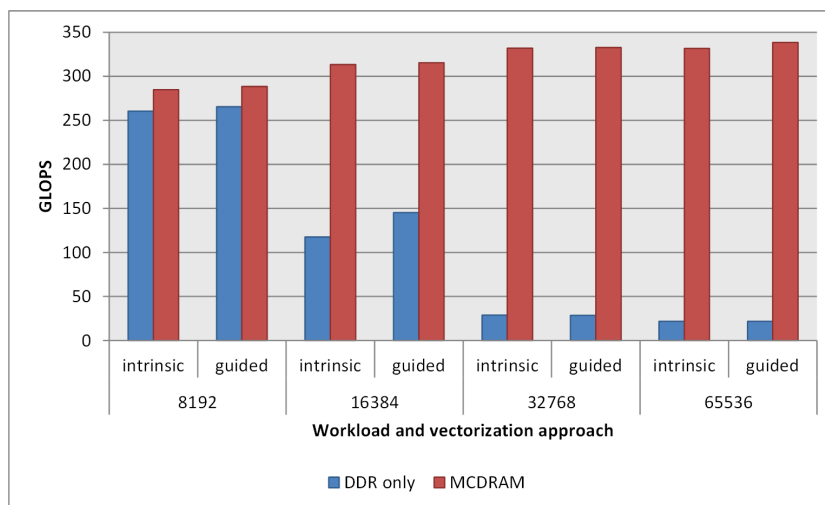


Fig. 7: Performance evolution varying workload and the MCDRAM exploitation.

proposed and evaluated several optimization techniques for different matrix factorizations methods on many-core systems.

Obtaining high-performance in graph algorithms is usually a difficult task since they tend to suffer from irregular dependencies and large space requirements. Regarding FW algorithm, there are many works proposed to solve the all-pairs shortest paths problem on different hardware architectures. However, to the best of the authors knowledge, there are no related works with KNL processors. Han and Kang [10] demonstrated that exploiting SSE2 instructions led to $2.3\times$ - $5.2\times$ speedups over a blocked version. Bondhugula et al. [4] proposed a tiled parallel implementation using Field Programmable Gate Arrays. In the field of GPUs, we highlight the work of Katz and Kider [13], who proposed a shared memory cache efficient implementation to handle graph sizes that are inherently larger than the DRAM memory available on the device. Also, Matsumoto et al. [15] presented a blocked algorithm for hybrid CPU-GPU systems aimed to minimize host-device communication. Finally, Hou et al. [11] evaluated different optimization techniques for Xeon Phi KNC coprocessor. Just as this study, they found that blocking and vectorization are key aspects in this problem to achieve high performance. Also, guided vectorization led to better results than the manual approach, but with larger performance differences. Contrary to this work, their implementation benefited from using more than one thread per core. However, as stated before, there are significant architectural differences between these platforms that support this behavior.

7 Conclusions

KNL is the second generation of Xeon Phi family and features new technologies in SIMD execution and memory access. In this paper, we have evaluated a set of programming and optimization techniques for these processors taking the FW algorithm as a representative case study of graph and memory-bound applications. Among the main contributions of this research we can summarize:

- Blocking technique not only improved performance but also allowed us to apply a coarse-grain workload distribution in the parallel implementation.
- SIMD exploitation was crucial to achieve top performance. In particular, the serial version run $2.9\times$, $6\times$ and $15.5\times$ faster using the SSE, AVX2 and AVX-512 extensions, respectively.
- Aligning memory accesses and loop unrolling also showed significant speedups.
- A single thread per core was enough to get maximal performance. In addition, *scatter* and *balanced* affinities provided extra performance.
- Besides keeping portability, guided vectorization led to slightly better performance than the intrinsic counterpart, running upto $1.03\times$ faster.
- MCDRAM usage demonstrated to be an efficient strategy to tolerate high-bandwidth demands with practically null programmer intervention, even when the dataset largely exceeded the MCDRAM size. In particular, it produced an average speedup of $9.8\times$ and a maximum speedup of $15.5\times$

As future work, we consider evaluating programming and optimization techniques in other cluster and memory modes as a way to extract more performance.

Acknowledgments The authors thank the ArTeCS Group from Universidad Complutense de Madrid for letting use their Xeon Phi KNL system.

References

1. Green500 Supercomputer Ranking, <https://www.green500.org/>
2. Top500 Supercomputer Ranking, <https://www.top500.org/>
3. Barnes, T., et al.: Evaluating and Optimizing the NERSC Workload on Knights Landing. In: Proceedings of the 7th International Workshop on Performance Modeling, Benchmarking and Simulation of High Performance Computing Systems. pp. 43–53. PMBS '16, IEEE Press, Piscataway, NJ, USA (2016)
4. Bondhugula, U., Devulapalli, A., Dinan, J., Fernando, J., Wyckoff, P., Stahlberg, E., Sadayappan, P.: Hardware/software integration for fpga-based all-pairs shortest-paths. In: 2006 14th Annual IEEE Symposium on Field-Programmable Custom Computing Machines. pp. 152–164 (April 2006)
5. Codreanu, V., Rodriguez, J., Saastad, O.W.: Best Practice Guide - Knights Landing (2017), <http://www.prace-ri.eu/IMG/pdf/Best-Practice-Guide-Knights-Landing.pdf>
6. Culler, D.E., Gupta, A., Singh, J.P.: Parallel Computer Architecture: A Hardware/Software Approach. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1st edn. (1997)

7. Floyd, R.W.: Algorithm 97: Shortest path. *Commun. ACM* 5(6), 345– (Jun 1962)
8. Giles, M.B., Reguly, I.: Trends in high-performance computing for engineering calculations. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 372(2022) (2014)
9. Haidar, A., Tomov, S., Arturov, K., Guney, M., Story, S., Dongarra, J.: LU, QR, and Cholesky factorizations: Programming model, performance analysis and optimization techniques for the Intel Knights Landing Xeon Phi. In: 2016 IEEE High Performance Extreme Computing Conference (HPEC). pp. 1–7 (Sept 2016)
10. Han, S., Kang, S.: Optimizing all-pairs shortest-path algorithm using vector instructions (2005)
11. Hou, K., Wang, H., c. Feng, W.: Delivering parallel programmability to the masses via the intel mic ecosystem: A case study. In: 2014 43rd International Conference on Parallel Processing Workshops. pp. 273–282 (Sept 2014)
12. Jalali, S., Noroozi, M.: Determination of the optimal escape routes of underground mine networks in emergency cases. *Safety Science* 47(8), 1077 – 1082 (2009)
13. Katz, G.J., Kider, Jr, J.T.: All-pairs shortest-paths for large graphs on the gpu. In: Proceedings of the 23rd ACM SIGGRAPH/EUROGRAPHICS Symposium on Graphics Hardware. pp. 47–55. GH '08, Eurographics Association, Aire-la-Ville, Switzerland, Switzerland (2008)
14. Khan, P., Konar, G., Chakraborty, N.: Modification of floyd-warshall's algorithm for shortest path routing in wireless sensor networks. In: 2014 Annual IEEE India Conference (INDICON). pp. 1–6 (Dec 2014)
15. Matsumoto, K., Nakasato, N., Sedukhin, S.G.: Blocked all-pairs shortest paths algorithm for hybrid cpu-gpu system. In: 2011 IEEE International Conference on High Performance Computing and Communications. pp. 145–152 (Sept 2011)
16. Nakaya, A., Goto, S., Kanehisa, M.: Extraction of correlated gene clusters by multiple graph comparison. *Genome Informatics* 12, 44–53 (2001)
17. Reinders, J., Jeffers, J., Sodani, A.: Intel Xeon Phi Processor High Performance Programming Knights Landing Edition. Morgan Kaufmann Publishers Inc., Boston, MA, USA (2016)
18. Rosales, C., Cazes, J., Milfeld, K., Gómez-Iglesias, A., Koesterke, L., Huang, L., Vienne, J.: A Comparative Study of Application Performance and Scalability on the Intel Knights Landing Processor, pp. 307–318. Springer International Publishing, Cham (2016)
19. Venkataraman, G., Sahni, S., Mukhopadhyaya, S.: A Blocked All-Pairs Shortest-Paths Algorithm, pp. 419–432. Springer Berlin Heidelberg (2000)
20. Warshall, S.: A theorem on boolean matrices. *J. ACM* 9(1), 11–12 (Jan 1962)

Cloud Robotics: Auto Rover 4WD y Cuadricóptero controlados remotamente desde AWS

Armando De Giusti^{1,2}, Ismael Rodríguez¹, Manuel Costanzo, Marcos Boggia

¹ Instituto de Investigación en Informática LIDI (III-LIDI), Facultad de Informática, Universidad Nacional de La Plata – Comisión de Investigaciones Científicas de la Provincia de Buenos Aires, Argentina.

² Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) Argentina

{degiusti, ismael} @lidi.info.unlp.edu.ar, manuelcostanzo22@gmail.com, marcos_boggia@hotmail.com

Resumen. Este trabajo presenta el despliegue de un sistema multi-robot, integrado por un chasis de auto Rover de 4 ruedas y un cuadricóptero Parrot Bebop, conectados al Cloud público de Amazon Web Services. Se detallan los prototipos desarrollados, como así también el protocolo de comunicación utilizado con el servicio de AWS IoT, y los algoritmos implementados para simular el control remoto de un vehículo de 4 ruedas y un dispositivo de vuelo, ambos, no tripulados.

Palabras Clave: Cloud Robotics, Cloud Computing, Internet of Things, AWS, MQTT, NODE-RED, Cuadricóptero, Drone, Rover 4WD, IoT.

1 Introducción

Los avances en el paradigma de Cloud Computing han provocado un factor disruptivo de las TI en la industria tecnológica. Según el Instituto Nacional de Estándares y Tecnologías del Departamento de Comercio de los EEUU (NIST), como en varias publicaciones de diversos autores, se ha definido a Cloud Computing como: “*un paradigma informático de cómputo distribuido, que proporciona grandes conjuntos de recursos virtuales (como ser hardware, plataformas de desarrollo, almacenamiento y/o aplicaciones), fácilmente accesibles y utilizables por medio de una interfaz de administración web. Estos recursos son proporcionados como servicios (del inglés, “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). El acceso a los recursos se realiza bajo la demanda de los usuarios, en base a un modelo de autoservicio*” [1].

El modelo de Cloud Computing presenta las características y beneficios siguientes: recursos disponibles bajo demanda, escalabilidad y elasticidad, aprovisionamiento automático de recursos y autoservicio.

Dicho paradigma, brinda al menos tres modelos de despliegue: Cloud Público, Cloud Privado y Cloud Híbrido [2] [3].

Por otro lado, considerando la gran influencia de la robótica en la sociedad y la diversidad de servicios ofrecidos por robots, nos encontramos que los mismos presentan grandes limitaciones en consumo de energía, poder de cómputo, capacidad de almacenamiento, toma de decisiones, tareas cognitivas, etc.

En el año 2010, comenzaron a surgir proyectos de investigación (ej.: RoboEarth [4]), que integraban las tecnologías de Cloud con los sistemas de robots. Es así, que James Kuffner propone el concepto de Cloud Robotics, basado en combinar las tecnologías de robots con el paradigma de Cloud Computing [5].

La idea de Cloud Robotics permite, por medio de aplicaciones, tratar los datos de los componentes de hardware del robot (sensores, actuadores, cámaras, microcontroladores, memoria, etc.), sin importar las limitaciones de cómputo de las placas de desarrollo con microcontroladores y la capacidad de almacenamiento de las mismas [6]. En otras palabras, este concepto permite a los robots obtener resultados de tareas de cómputo intensivo, tales como: procesamiento de imágenes, reconocimiento de voz, determinación de rutas, confección de mapas, acciones cognitivas, etc., sin tratamiento local, sino en el Cloud.

Este paradigma brinda la capacidad de establecer escenarios para sistemas de multi-robot, donde cada robot se integra de un hardware mínimo; una placa microcontroladora con conectividad WiFi permite al robot comunicarse con el Cloud y otros robots. Los datos de sensores y/o sistemas de adquisición de imágenes se procesan en el Cloud y los actuadores de cada robot llevarán a cabo las operaciones necesarias [7].

El propósito de investigación del presente trabajo es el despliegue de un sistema de multi-robot, inicialmente con dos robots conectados al Cloud Público de Amazon Web Services con el fin de simular el control remoto de un vehículo de 4 ruedas y un dispositivo de vuelo, ambos, no tripulados. Para alcanzar tales objetivos, se determinó el método de comunicación con el Cloud; se registraron los robots como dispositivos o “cosas” en el servicio de AWS IoT [8]; también, se confeccionaron los prototipos de cada robot y se desarrollaron los algoritmos que permiten controlar las acciones a ejecutar por cada robot.

El presente trabajo está estructurado de la manera siguiente: en la Sección 2, se introducen algunos conceptos elementales. A continuación, la Sección 3, presenta los prototipos confeccionados y sus componentes. En la Sección 4, se describe el trabajo experimental realizado y los resultados obtenidos. Por último, la Sección 5, expone las líneas de trabajo futuro en relación a esta investigación.

2 Algunos conceptos elementales

En esta Sección, se mencionan algunos conceptos empleados en la investigación realizada.

▪ Amazon Web Services

Es una plataforma de Cloud Público, que provee “Infraestructura como servicio”; a través de la tecnología de virtualización, brinda un gran conjunto de recursos de cómputo, como almacenamiento y capacidad de procesamiento, que pueden ser solicitados a demanda, como así también, se adecuan dinámicamente en tamaño conforme la necesidad del consumidor. Los servicios más destacados de AWS son “Elastic Compute Cloud” (EC2), “Simple Storage Service” (S3) y en nuestro caso utilizaremos los servicios de “AWS Internet of Things” (AWS IoT) [9].

▪ Internet of Things (IoT)

Es un nuevo paradigma cuya definición deriva de considerar “objetos” o “cosas” conectadas a Internet por medio de redes Wireless, sean estas de tecnología WiFi o 4G LTE, utilizando protocolos de comunicación estándar. Dichas “cosas” pueden considerarse etiquetas de identificación por radio frecuencia (RFID), sensores, actuadores, teléfonos móviles, placas de desarrollos, etc. [10].

▪ Protocolo MQTT

Message Queuing Telemetry Transport [11] es un protocolo de comunicación ligero, especialmente diseñado para tolerar conexiones intermitentes y reducir los requisitos de ancho de banda de la red. Desde finales del año 2014, fue presentado como un estándar abierto OASIS [12] y se ha convertido en el protocolo por excelencia para IoT. Soporta comunicación segura con TLS. Maneja tres niveles de calidad de servicio: QoS 0: A lo sumo una vez la entrega del mensaje. QoS 1: Al menos una vez la entrega del mensaje. QoS 2: Exactamente una vez la entrega del mensaje.

Este protocolo, implementa la comunicación de mensajes por medio de la publicación/suscripción sobre un tópico específico (canal de comunicación), como se puede observar en la figura siguiente:

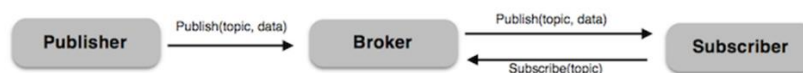


Figura 1: MQTT: publicación/suscripción.

▪ AWS IoT

Es un servicio que proporciona AWS con el fin de conectar, administrar y operar grandes conjuntos de dispositivos o cosas. Ofrece mecanismos de conectividad y seguridad para la transmisión de datos. También, permite que los datos, una vez enviados a la plataforma, puedan ser procesados por las aplicaciones de análisis masivo de datos (Elastic MapReduce), análisis predictivo para aprendizaje automático (Amazon Machine Learning), almacenamiento en bases de datos, etc. [13].

AWS IoT permite conectar fácilmente dispositivos al Cloud y a otros dispositivos. El servicio es compatible con los protocolos: HTTP, WebSockets y

MQTT. Amazon ha optado por este último, por ende AWS IoT utiliza la especificación del protocolo MQTT v.3.1.1 con QoS 0 y 1.

Cada dispositivo debe ser registrado como una “cosa” en AWS IoT, para lo cual se emitirá un certificado y un par de llaves privada-pública para el mismo. El certificado y la llave privada, junto con el certificado de la Entidad Certificantes (CA) de Amazon, deberán ser instalados en el dispositivo con el fin que este pueda conectarse al servicio de AWS IoT previa autenticación necesaria.

- **Node-RED**

Es una herramienta de programación visual, que permite programar algoritmos basados en flujos para IoT, sin la necesidad de escribir código. Provee de un editor de flujo basado en navegador WEB, que brinda una amplia paleta de nodos con diversa funcionalidad. El flujo se confecciona conectando los nodos entre sí. La funcionalidad de cada nodo está desarrollada sobre Node.js. Los flujos creados son almacenados utilizando JSON, lo cual facilita la importación o exportación para compartir los mismos [14].

3 Prototipos y componentes

Con el fin de simular el control remoto de un vehículo de 4 ruedas y un dispositivo de vuelo, no tripulados, se han implementado dos prototipos de robots que interactúan con el Cloud público de AWS, vía conexión WiFi y Ethernet; de aquí en adelante se denomina al primer robot, como “Rover 4WD” y al segundo, “Cuadricóptero”.

Asimismo, se ha desplegado una instancia de AWS EC2, para ejecutar los algoritmos que permiten controlar remotamente los robots mencionados.

3.1 Robot Rover 4WD

El robot Rover 4WD, emula un vehículo de 4 ruedas, que recibe, a través del Cloud, las acciones que debe ejecutar. Para la confección de tal robot, se han utilizado los componentes siguientes:

- **Chasis Rover de 4 ruedas**

Consiste en dos placas de acrílico, con cuatro motores DC de 3v (dos del lado derecho: R1, R2, y dos del izquierdo: L1, L2), cuatro ruedas y un codificador de velocidad.



Figura 2: Chasis Rover de 4 ruedas.

- **Placa de desarrollo Raspberry**

Placa Raspberry Pi 3 con conexión WiFi, almacenamiento en memoria MicroSD de 32Gb y alimentación de energía de 5v.

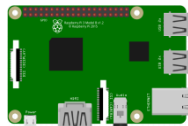


Figura 3: Placa Raspberry Pi 3.

- **Placa de desarrollo Arduino**

Placa Arduino UNO, con sistema de alimentación de energía de 5v. Esta placa gestiona los sensores ultrasónicos para detectar obstáculos.



Figura 4: Placa Arduino UNO.

- **Puente H**

Módulo L298 para controlar la dirección de los motores; por medio de PWM, la velocidad de los mismos.



Figura 5: Puente H Módulo L298.

- **Sensor ultrasónico HC-SR04**

Detector de proximidad o detector de obstáculos. Cuantifica el delay de la señal, es decir, el tiempo que demora la señal en regresar el eco.

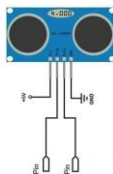


Figura 6: Sensor ultrasónico HC-SR04.

- **Prototipo Rover 4WD ensamblado**



Figura 7: Robot Rover 4WD.

3.2 Cuadricóptero

El prototipo del robot cuadricóptero está integrado por una placa de desarrollo Raspberry Pi 3 y un Drone Parrot Bebop. El cuadricóptero está conectado vía Ethernet al Cloud de AWS por medio de la placa de desarrollo Raspberry Pi 3. La conexión entre la placa de desarrollo Raspberry y el Drone se realiza vía WiFi, estableciendo una conexión punto a punto segura, evitando así perder el control del dron.

- **Placa de desarrollo Raspberry**

Placa Raspberry Pi 3 con conexión WiFi, almacenamiento en memoria MicroSD de 16Gb y alimentación de energía de 5v.

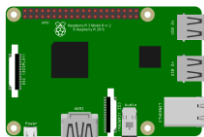


Figura 9: Placa Raspberry Pi 3.

- **Drone Parrot Bebop**

Cuadricóptero de uso civil, de origen Francés, ligero, compacto, con cámara de video Full HD y autonomía de vuelo de 25 minutos con 2 baterías de 1200 mAh.



Figura 10: Drone Parrot Bebop.

3.3 Instancia AWS EC2 – Aplicación para control remoto

Consiste en el despliegue de una instancia de AWS EC2 “t2.micro”, con 1 VCPU y 1GB de RAM; Sistema Operativo Linux distribución Ubuntu Server 16.04 LTS [15] [16]. Dicha instancia provee una interfaz WEB, desarrollada con HTML + JavaScript + Python, con el fin de controlar remotamente al sistema de multi-robots.

4 Trabajo experimental

Para el trabajo de investigación se requirió desarrollar una plataforma web que permita controlar remotamente, vía internet, el sistema de multi-robot que emula un vehículo de 4 ruedas y un cuadricóptero, no tripulados.

En la instancia de AWS, se configuró una interfaz WEB accesible por medio de su IP pública, lo que permite al usuario controlar el sistema multi-robot. Para el robot “Rover 4WD” se dispone de un teclado visual que permite gestionar las acciones sobre el mismo, por medio de los botones visuales o las teclas del teclado de una PC/Notebook: “arriba, abajo, izquierda y derecha” o con las teclas “A, W, S, D”; con este último caso, también se brinda la posibilidad de hacer girar al robot sobre su propio eje sin avanzar, esto se efectúa presionando las teclas “Q” o “E”. El uso de esta plataforma es similar al de un joystick, por ejemplo, para avanzar se mantiene presionada la tecla “arriba” (o la tecla “W”) y cuando se libera la presión de la misma, automáticamente el robot frena. Este algoritmo está desarrollado en JavaScript, integrado en la página HTML que ofrece la interfaz WEB. En segundo plano, la instancia de AWS envía los comandos al robot mediante el protocolo MQTT, publicando mensajes en el tópico “ROVER”, haciendo uso de la aplicación SDK de AWS IoT, implementada con Python. En la figura siguiente, se puede observar la interfaz WEB y la implementación del algoritmo:

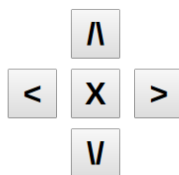


Figura 11: Interfaz WEB - Robot “Rover 4WD”.

```
<html> <head>
<script      src="{{url_for('static',
filename='js/rover.js')      }}"></script>
</head>
<body> <div align="center">
<div id="up">
<button id="bt_up">Λ</button> </div>
<div id="middle">
<button id="bt_left"><</button>
<button id="bt_stop">X</button>
<button id="bt_right">>>/button>
</div>
<div id="down">
<button id="bt_right">V</button> </div>
</div>
</body> </html>
```

Figura 12: Algoritmo - Robot “Rover 4WD”.

Para el caso del robot “Cuadrícóptero”, también se dispone de una interfaz WEB que ofrece un teclado visual que permite controlar las acciones del robot (Take Off, Land, Forward, Right, Left, Backward). Esta interfaz está implementada con node-red-dashboard [17], y en segundo plano, la instancia de AWS envía los comandos al

robot mediante el protocolo MQTT utilizando node-red-mqtt-topic, publicando los mensajes en el tópic “DRONE”. En la figura siguiente, se puede observar la interfaz WEB y la implementación de su algoritmo:

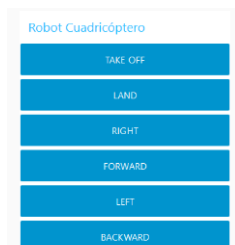


Figura 13: Interfaz WEB - “Cuadróptero”.

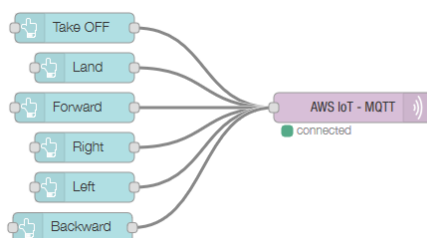


Figura 14: Algoritmo - “Cuadróptero”.

El robot “Rover 4WD”, dispone de dos placas de desarrollo: Raspberry Pi3 y Arduino UNO. La placa Raspberry Pi3, vía la conexión WiFi, se encarga de recibir los comandos enviados por el protocolo MQTT (suscripto al tópic “ROVER”), activando o desactivando los motores (R1, R2, L1, L2); la lógica de este algoritmo se encuentra implementada con un flujo de Node-Red. Además, consta de dos algoritmos, los cuales reciben y envían datos al Arduino UNO respectivamente. El algoritmo que consume datos del Arduino UNO, recibirá una notificación cuando el robot esté por colisionar y, en caso afirmativo, luego recibirá otra notificación cuando el robot ya no esté más en peligro, ya sea porque el obstáculo se movió o porque se cambió la dirección del robot. Esta notificación se envía a la aplicación Node-Red para que realice las acciones correspondientes a través de un socket local en la placa Raspberry Pi3.

Por otro lado, la placa de desarrollo Arduino UNO se encarga de controlar que el robot no colisione, haciendo uso de los sensores ultrasónicos. Para determinar qué sensor ultrasónico utilizar, se reciben datos de la placa Raspberry Pi3, con el fin de conocer la orientación y dirección en la que viaja el robot y así evaluar el lado en el que se deben comprobar las distancias con los obstáculos. Cuando se está por colisionar, se envía una señal a la Raspberry Pi3 para que detenga los motores; una vez que se evade el obstáculo, ya sea porque el mismo se movió y no hay más peligro o, porque se cambió la dirección del robot a una en donde no hay peligro de colisión, se envía una señal afirmativa para que pueda avanzar.

En la figura siguiente, se puede observar el flujo de Node-RED para el robot “Rover 4WD”:

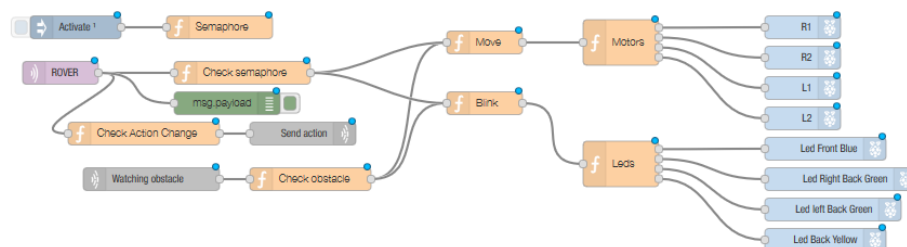


Figura 15: Flujo de Node-RED del robot “Rover 4WD”.

El robot “Cuadrícóptero”, dispone de una placa de desarrollo Raspberry Pi3 y un dispositivo de vuelo Drone Parrot Bebop. La placa Raspberry Pi3 establece una conexión punto a punto con el Drone Bebop, de forma segura, y vía conexión Ethernet en la placa, recibe los comandos enviados por el protocolo MQTT (subscrito al tópic “DRONE”); por cada comando recibido, se envía una orden específica al Drone Bebop, por medio del nodo de node-red-contrib-bebop2 [18].

En la figura siguiente, se puede observar el flujo de Node-RED para el robot “Cuadrícóptero”.



Figura 16: Flujo de Node-RED del robot “Cuadrícóptero”.

5 Conclusiones y Líneas de Trabajo Futuro

En esta investigación se ha implementado un sistema multi-robot conectado al Cloud. Se han diseñado los prototipos para simular un vehículo de 4 ruedas y un dispositivo de vuelo, no tripulados. Se han desarrollado los algoritmos que permitan a un usuario controlar remotamente los robots por medio de una interfaz WEB. Podemos concluir y reafirmar que Cloud Robotics es una tecnología que favorecerá el avance de los sistemas multi-robots reduciendo las limitaciones actuales.

Las líneas de trabajo a futuro relacionadas con este trabajo, incluyen la integración de una cámara al frente del robot “Rover 4WD” y la visualización de los videos en la interfaz WEB; así como también ampliar el espectro de visión de obstáculos adicionando más sensores ultrasónicos al chasis del robot.

Del mismo modo, es deseable ampliar el desarrollo de los algoritmos del nodo node-red-contrib-bebop2 para poder acceder a los datos que pueda proporcionar el robot “Cuadrícóptero”, como ser: video de la cámara incorporada, datos de vuelo, coordenadas de GPS, etc.

Por último, se espera adicionar uno o más robots Rover 4WD, adaptando la solución para simular una ciudad en la cual los sistemas multi-robot puedan trabajar colaborativamente.

Referencias

1. Mell P., Grance T.: “The NIST Definition of Cloud Computing”. Publicación Especial 800-145. Septiembre, 2011.
2. Pettoruti J., Rodríguez I., Chichizola F., De Giusti A.: “Análisis de la degradación de las comunicaciones en algoritmos de cómputo científico en un Cloud privado”. XII Workshop de Procesamiento Distribuido y Paralelo (WPDP) – XVIII Congreso Argentino de Ciencias de la Computación (CACIC2012). Bahía Blanca, Argentina, 2012.
3. Rodríguez, I., Pettoruti, J.E., Chichizola, F., De Giusti, A.: Despliegue de un Cloud Privado para entornos de cómputo científico. Proceedings del XI Workshop de Procesamiento Distribuido y Paralelo (WPDP) - XVII Congreso Argentino de Ciencias de la Computación (CACIC 2011). La Plata, Argentina. 2011.
4. RoboEarth. In: <http://www.roboearth.org>.
5. Kuffner J.: “Cloud-enabled robots”. IEEE-RAS International Conference on Humanoid Robot. Nashville, USA, 2010.
6. Wang L., Liu M., Meng M., Siegwart R.: “Towards Real-Time Multi-Sensor Information Retrieval in Cloud Robotic System”. IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI). Hamburgo, Alemania, 2012.
7. Turnbull L., et al.: “Cloud Robotics: Formation Control of a Multi Robot System Utilizing Cloud Infrastructure”. Proceedings of IEEE – Southeastcon. Jacksonville, USA, 2013.
8. AWS IoT, <https://aws.amazon.com/es/iot>.
9. Bermudez I., Traverso S., Mellia M., Munafò M.: “Exploring the cloud from passive measurements: The Amazon AWS case”. Proceedings of IEEE – INFOCOM, 2013.
10. Atzori L., Iera A., Morabito G.: “The Internet of Things: A survey”. ELSEVIER Journal Computer Networks. Volume 54, Issue 15. 2010.
11. Protocolo MQTT, <http://mqtt.org>.
12. OASIS MQTT, <https://www.oasis-open.org>.
13. Informe técnico: AWS Whitepapers. “Core Tenets of IoT”, <https://aws.amazon.com/es/whitepapers/>.
14. Node-RED, <https://nodered.org>.
15. AWS Instance Types: <https://aws.amazon.com/es/ec2/instance-types>.
16. AWS Marketplace: <https://aws.amazon.com/marketplace/pp/B00JV9JBDS>.
17. Node-Red-Dashboard, <https://github.com/node-red/node-red-dashboard>.
18. Node-Red-Contrib-Bebop2, <https://www.npmjs.com/package/node-red-contrib-bebop2>.

Despliegue de un Sistema Web Vertical Rígido como Sistema Web Elástico en la Nube

Fernando Invernizzi, Sebastián Ortiz, Carlos Núñez, and Cristian Cabanellas

Universidad Nacional de Asunción, Facultad Politécnica, Paraguay
{ferinver92,sebastianortizsantacruz}@gmail.com
{cnunez,cristian.cabanellas}@pol.una.py
<http://www.pol.una.py/>

Resumen La computación en la nube es una tecnología que se encuentra en crecimiento, la cual tiene la capacidad de escalar y aprovisionar recursos en forma dinámica sin la gestión de la complejidad subyacente. La necesidad de acceder a la información desde cualquier lugar y momento han llevado a repensar las formas en que se despliegan las aplicaciones actuales, esencialmente aquellas del tipo web. Los beneficios que ofrece la nube hacen de ella una candidata excelente para tomarla como solución. Entonces, ¿Cómo migramos desde el entorno tradicional?

Este trabajo implementa una metodología de migración extendiendo a un esquema conceptual base existente, presentando las adversidades durante el proceso y las conveniencias del mismo.

El estudio lanzó resultados alentadores, principalmente en inversión y eficiencia, representando una sólida justificación para la adopción del paradigma. Se espera que la presente investigación motive la migración de aplicaciones verticales rígidas a la nube.

Keywords: Computación en la nube, Metodología de migración, Aplicación nativa de la nube, Cloud Ready, IaaS, TICs

1. Introducción

Analizando cifras de mercado se comprueba que 65 % de las inversiones de TI se destinan al mantenimiento de infraestructuras y servicios (“Run the business spendings” expuesto en [1]). Esta es una de las principales razones por las que muchas organizaciones de TI muestran interés en una tecnología que les permita disminuir y hacer variables los costes de infraestructuras y mantenimiento [2].

Una vía para incrementar la productividad es optimizando la inversión en infraestructuras acudiendo a servicios en la nube [2]. Al notar su atractivo, la pregunta que viene inmediatamente es: ¿Cómo migramos allí?

Este trabajo tiene como objetivo general contribuir con la implementación de una metodología de migración de un sistema web convencional alojado en infraestructura tradicional, a una infraestructura de computación en la nube, tomando como base un esquema conceptual propuesto por AWS [3], de manera a comprobar empíricamente las limitaciones y conveniencia de la migración en términos de coste y beneficios.

La metodología propuesta está orientada a la adopción de una arquitectura *Cloud Ready*, es decir, una arquitectura nativa de la nube. La validación *Cloud Ready* es uno de los principales aportes del estudio.

El resto de este documento se estructura como sigue: los antecedentes y motivaciones en la sección 2. En la sección 3, se define *Cloud Ready* y el algoritmo de validación de aplicaciones. En la sección 4, se realiza la descripción de la metodología propuesta. En 5 se aplica la metodología a un caso de estudio. Finalmente, en la sección 6 se exponen las conclusiones del trabajo.

2. Antecedentes y Motivación

Normalmente se piensa que la migración a la nube no es más que la instalación en el entorno. Este concepto es erróneo ya que se confunde con una replicación, la cual no aprovecharía al máximo a la nube. Actualmente existen esquemas propuestos para migrar, pero los mismos sólo se presentan de manera conceptual.

En [4] se propone un modelo de cascada para la migración a la nube basado en el modelo cascada iterativo del Ciclo de Vida de Desarrollo Software (SDLC). El artículo destaca desafíos en el proceso sin incluir aspectos de implementación.

El framework CloudGenius [5] ofrece un proceso de migración y soporte para toma de decisiones, con un enfoque que considera solo el servidor web, además de asumir que primero se trabaja y elige una imagen de máquina virtual.

En [6], se concluye la clara necesidad de un proceso bien definido para la migración de aplicaciones a la nube y de los beneficios que se obtendrían.

[7] y [8] manifiestan la falta de apoyo para la migración a la nube en forma de frameworks, herramientas y métodos de ayuda para la toma de decisión.

En [9] los autores presentan una metodología para migración a la nube, pero para ambientes PaaS. Este trabajo pretende extender el mismo concentrándose en las aplicaciones mediante el aporte de una metodología orientada a la adopción de una arquitectura nativa de la nube *Cloud Ready*.

3. Cloud Ready

CloudGenius [5] y *Cloud Migration Research: A Systematic Review* [8] mencionan la falta de apoyo para la adaptación arquitectónica de sistemas que se buscan llevar a la nube, logrando aprovechar de la mejor manera el entorno.

Cloud Ready es aquel servicio diseñado para trabajar a través de internet [10]. Es una aplicación que puede ser efectivamente desplegada en la nube [11].

La validación *Cloud Ready*, aporte de este trabajo, permite una arquitectura óptima para la nube mediante el cumplimiento de los principios *Cloud Ready*,

3.1. Principios Cloud Ready

Los principios *Cloud Ready* son el conjunto de características que necesita un aplicativo para ser considerado poseedor de una arquitectura nativa de la nube.

Estos principios se elaboraron a partir de una serie de artículos y estudios [11][12][13][14][15][16][17][18]. A continuación se expone el conjunto formulado:

1. **No escribir la aplicación para una topología específica.** No suponer la existencia fija de nodos [11][14][18].
2. **No asumir que el sistema de archivos local es permanente.** No suponer el ciclo de vida de los archivos en el sistema de archivos [11][18].
3. **No mantener una sesión en la aplicación.** Una aplicación del tipo stateful limita la escalabilidad, se sugiere que sea stateless [11][12][13][18].
4. **No escribir logs en el sistema de archivos.** Con logs escritos localmente, se pierde información y trazabilidad si la máquina falla [11][12][13][14][18].
5. **No utilizar APIs de infraestructura ni características propias del sistema operativo.** Evitar utilizar “APIs de la infraestructura o de SO” ya que dificultan el monitoreo por parte del proveedor de nube [11][14].
6. **No utilizar protocolos oscuros.** Evitar protocolos que requieran de alguna configuración especial que pueda afectar la flexibilidad [11][12][13][18].
7. **No instalar manualmente la aplicación.** La instalación sencilla permite adoptar diferentes técnicas de automatización [11][12][13][14][15][17][18].
8. **Código base.** Cada aplicación desplegable se debe mantener en un código único bajo un control de revisión [12][13][18].
9. **Servicios de soporte.** Los servicios de soporte deben tratarse como recursos adjuntos, de consumo idéntico en los entornos de despliegue [12][13].
10. **Concurrencia.** Dos o más procesos son concurrentes cuando se ejecutan al mismo tiempo y no existen dependencias entre los mismos [12][13].
11. **Desechabilidad y robustez.** Es el manejo de entradas inválidas, permitiendo así confiabilidad y recuperación frente a accidentes [12][13][15][17][18].
12. **Procesos administrativos.** Procesos de “una sola vez” deberían ejecutarse en entornos idénticos a los procesos regulares de larga duración [12][13][18].
13. **Escalabilidad.** Los sistemas que se espera crezcan con el tiempo necesitan ser construidos sobre una arquitectura escalable [14][15][16][17].
14. **Procesos stateless.** Se debe tener componentes débilmente acoplados, que idealmente no compartan nada [16][17][18].
15. **Paralelización.** Busca analizar cómo superponer operaciones para mejorar el rendimiento y el uso de infraestructura al realizar una tarea concreta [16].
16. **Técnicas y estrategias de tolerancia a fallos.** Describe un sistema diseñado de modo a que en caso de fallo puede restablecerse y alertar sin la pérdida del servicio [14][15][16][18].
17. **Evitar adivinar la necesidad de capacidad.** Una aplicación orientada a la nube debe prepararse para usar recursos bajo demanda [14][16][18].
18. **Aplicar seguridad en todas las capas.** Se debe proteger la integridad y privacidad de la información almacenada en un sistema [14][15][18].

3.2. Validación Cloud Ready

El algoritmo de validación *Cloud Ready* (presentando en Alg. 1) utiliza los principios *Cloud Ready* para comprobar si una aplicación efectivamente cuenta con una arquitectura nativa de la nube.

Antes de iniciar, la validación requiere de una buena comprensión del aplicativo a migrar. Se sugiere su descomposición y análisis en detalle.

Por cada principio, se evalúa el cumplimiento de la arquitectura del sistema. Si cumple, se pasa al siguiente principio. En caso de no cumplirlo, se lo ajusta.

Algoritmo 1 Validación Cloud Ready

Entrada: Aplicativo A, Conjunto de Principios CP

Salida: Aplicativo con arquitectura Cloud Ready.

- 1: **para todo** P en CP **hacer**
 - 2: **si** A cumple P **entonces**
 - 3: obtener siguiente principio
 - 4: **si no**
 - 5: ajustar A para que cumpla P
 - 6: **fin si**
 - 7: **fin para**
-

El empleo ideal del algoritmo *Cloud Ready* ofrece un aplicativo con la arquitectura capaz de explotar de la mejor manera las características de la nube.

3.3. Extensión a la validación Cloud Ready en caso de limitaciones

En algunos casos, por diversos motivos (dinero, esfuerzo, capacidad, tiempo, entre otros), los principios no pueden ser alcanzados en su totalidad. Para estas situaciones, el proceso a seguir para determinar si es beneficioso continuar con la migración está compuesto por las siguientes tareas:

- **Recuento de principios**, tarea que busca obtener el porcentaje *Cloud Ready*, el cual se calcula a partir de la sumatoria de valoración de los principios. Cada principio se evalúa según el grado de cumplimiento que tenga (Alta, Media, Baja) y, a partir de ese grado, se suma su correspondiente valoración (1, 0.5, 0) al total.
- **Conclusión a partir del recuento**, tarea que busca determinar a partir del porcentaje *Cloud Ready*, si el aplicativo se encuentra apto para ser migrado. Si el porcentaje *Cloud Ready* es superior al 80 %, el aplicativo es apto para la migración, si se encuentra entre 60 % y 80 % es medianamente apto y, en caso de ser inferior al 60 %, no resulta ser apto.

Si se tuviera un aplicativo que no cumple plenamente con los 18 principios, y cuya sumatoria de valoración, arroje un total de 15, su porcentaje *Cloud Ready* estaría definido por: $15/18 \cdot 100 = 83,3 \%$, con lo que se puede concluir, que el aplicativo de igual modo se encuentra en condiciones para ser migrado.

4. Metodología propuesta de migración

El modelo conceptual de Amazon [3] fue la base para la implementación de la metodología de migración. La elección se debe a que su enfoque se corresponde con nuestras percepciones iniciales de la necesidad existente, envolviendo aspectos de caracterización, con la visión de explotar los beneficios de la nube.

La metodología propuesta se ocupa de comprender a la organización en la que uno se encuentra, conocer los diferentes sistemas junto a la complejidad de la arquitectura subyacente de cada una y su relación con los demás. Plantea una ruta para reconocer las virtudes del proveedor de nube, migrar el aplicativo según la estrategia mas conveniente y añadir servicios que aprovechen a la nube.

Este trabajo aporta lo siguiente:

- Evalúa y enfoca al aplicativo a una arquitectura *Cloud Ready*.
- Discute aspectos técnicos y limitaciones (tiempo, dinero) para la toma de decisiones durante la migración.
- Presenta una perspectiva general para la selección de los servicios ofrecidos por el proveedor de nube.
- Incluye un esquema de pruebas para validar los beneficios de la nube en el aplicativo migrado.

Como resultado, la metodología migra un aplicativo a la nube capaz de aprovechar al máximo de sus beneficios. La Fig. 1 presenta la metodología final.

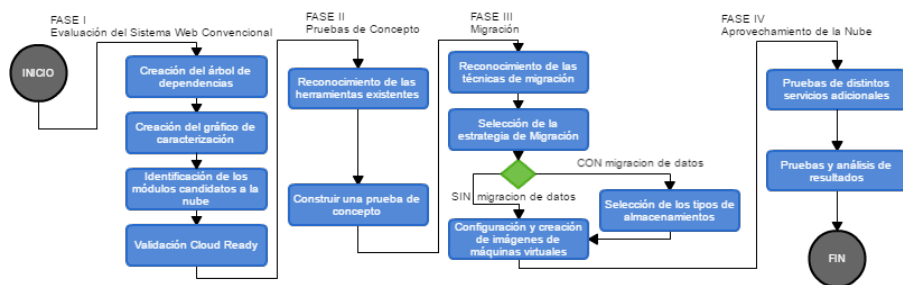


Figura 1. Metodología Propuesta

La metodología utiliza el proveedor AWS [19] para su ejemplificación ya que cuenta con importantes informes elaborados por Gartner y Forester que destacan su integridad, capacidades críticas y seguridad [20][21][22][23].

4.1. Evaluación del Sistema Web Convencional

La fase inicial busca conocer los sistemas de la organización, analizar las conexiones existentes, caracterizar a los componentes y, partiendo de la caracterización, identificar los mejores candidatos para la selección y ajuste del mismo.

Creación del árbol de dependencias

Esta etapa realiza un examen de las conexiones lógicas existentes entre los sistemas mediante la elaboración de un árbol de dependencias, el mismo permite identificar las aplicaciones y sus dependencias sobre otros componentes. Los nodos representan los componentes y las aristas las relaciones.

Creación del gráfico de caracterización

El gráfico de caracterización se resume en un diagrama de las aplicaciones de la organización. Se conforma por los módulos identificados en el árbol de dependencias, donde cada módulo se caracteriza según su papel (dependencias).

Identificación de los módulos candidatos a la nube

Los mejores candidatos para la nube, sin tener en cuenta modificaciones estructurales en su actual disposición, son aquellas aplicaciones con pocos componentes; aplicaciones con necesidad de escalar y que corren por sobre su capacidad. Definida la lista de candidatos, se debe priorizar la migración según convenga.

Validación Cloud Ready

Con el aplicativo seleccionado, lo siguiente es emplear la validación *Cloud Ready* (sección 3) sobre el mismo. El objetivo de la etapa es que el aplicativo a migrar posea una arquitectura capaz de aprovechar eficientemente a la nube.

4.2. Pruebas de Concepto

Esta fase se centra en conocer las características del proveedor y obtener confianza en sus servicios dando los primeros pasos sobre el mismo.

Reconocimiento de las herramientas existentes

Esta actividad consiste en familiarizarse con las herramientas del proveedor. Los servicios básicos a aprender deberían ser: creación de máquinas virtuales, gestión de imágenes virtuales, asignación de volúmenes de almacenamiento, carga de datos a la nube, gestión de base de datos, entre otras cosas.

Construir una prueba de concepto

Al conocer los servicios básicos del proveedor, se debe continuar con la construcción de una prueba de concepto. La prueba consiste en representar al aplicativo en la nube mediante los servicios reconocidos.

4.3. Migración

Esta etapa presenta las principales estrategias de migración de aplicaciones y los aspectos a tener en cuenta para la selección de tipos de almacenamiento.

Reconocimiento de las técnicas de migración

Una estrategia de migración representa un plan que permite trasladar a la nube un aplicativo. Las principales estrategias que se sugiere conocer y tener en cuenta para la migración son la híbrida, de reemplazo, forklift y cloudify [8] [24].

Selección de la estrategia

La selección implica decidir la estrategia de migración que más convenga. *Cloud Ready* sugiere la estrategia cloudify. Si la estrategia requiere de almacenamiento se debe continuar con la “Selección de los tipos de almacenamientos”, de otro modo con la “Configuración y creación de imágenes de máquinas virtuales”.

Selección de los tipos de almacenamientos

Para la selección del tipo de almacenamiento adecuado, se deben realizar las compensaciones justas entre las diversas dimensiones existentes (costo, disponibilidad, durabilidad, latencia, performance, capacidad de cache y consistencia).

Configuración y creación de imágenes de máquinas virtuales

La creación de imágenes virtuales y procesos de despliegue automatizados reducen el tiempo y esfuerzo de construcción de una aplicación [3]. Se recomienda servirse de herramientas de administración de despliegue y configuración.

4.4. Aprovechamiento de la Nube

Esta fase se centra en incorporar aptitudes del proveedor de la nube a la arquitectura migrada y la creación de un esquema de pruebas cuyo objetivo es el de determinar el éxito de la migración. Ambas tareas se encuentran fundamentadas sobre las cinco características esenciales definidas por el NIST [25].

Pruebas de distintos servicios adicionales

El apartado se centra en realizar pruebas sobre servicios que permitan fortalecer al aplicativo desplegado inicialmente con características de la nube. Se recomienda incluir servicios que potencien a las prestaciones (autoescalamiento, balanceo de carga y otros).

Pruebas y análisis de resultados

El esquema de pruebas para determinar el éxito de la migración es como sigue:

1. *Emparejamiento*: consiste en asociar los componentes de la arquitectura en el entorno tradicional con los del entorno en la nube.
2. *Definición de las funcionalidades de prueba*: la selección debe incluir a aquellas que causen más carga en el sistema y se utilicen con mayor frecuencia.
3. *Selección de la herramienta de pruebas*: la herramienta debe permitir inyectar cargas reales al sistema. Debe incluir características como concurrencia, definición de secuencia de tareas y la generación de reportes representativos.
4. *Evaluación*: se divide en la estimación del costo de los servicios utilizados y la medición del éxito de la migración con respecto a las características nube.
 - a) *Costo de servicios*: El objetivo es tener un estimativo del costo de los servicios utilizados en la nube y su rentabilidad en el tiempo.
 - b) *Características nube*: El éxito de la migración se evalúa a partir del comportamiento del aplicativo con respecto al cumplimiento de las características esenciales de la nube definidas por el NIST [25].

5. Caso de estudio

El entorno seleccionado para el estudio es el de una entidad financiera. Fue elegido por la cantidad de usuarios que maneja y la alta demanda de recursos.

El árbol de dependencias y el gráfico de caracterización permitieron detectar todos los módulos y conexiones dentro de la entidad. Ante la alta demanda del particular *módulo de transferencias*, se lo priorizó para buscar otorgarle flexibilidad y elasticidad mediante la nube. La validación *Cloud Ready* (sección 3) no pudo ser empleada idealmente por falta de tiempo y capital. De igual forma, se realizó el cálculo del porcentaje *Cloud Ready* obteniendo un 91,6%, por lo que se concluyó que el mismo se encontraba apto para la migración.

En las pruebas de concepto, se reconoció y estudió a los servicios *Elastic Compute Cloud* (EC2) [26], *Elastic Block Store* (EBS) [27], *Relational Database Service* (RDS) [28] y *Simple Storage Service* (S3) [29] ofrecidos por AWS.

Para la migración, la técnica utilizada fue *cloudify*, reemplazando todos los componentes por servicios de AWS. El detalle es como sigue: a) el módulo core instalado en una instancia EC2, b) anexo de volumen EBS a la instancia EC2 en la cual se situarán los archivos de configuración de arranque, c) base de datos migrada como una instancia RDS, y d) archivos estáticos y logs en un volumen S3. Por otra parte, la migración no utilizó herramientas de despliegue.

Para el aprovechamiento de la nube, se utilizó el Grupo de Autoescalado [30] y el *Elastic Load Balancing* (ELB) [31]. La calculadora de Costo Total de Propiedad (TCO) [32] con los parámetros que reflejan la infraestructura antes de la migración, estima un ahorro del 97% de los costos en 3 años. Finalmente, las figuras 2 y 3 comprueban el éxito de la migración con la obtención del autoservicio bajo demanda y la rápida elasticidad del aplicativo en la nube.



Figura 2. Autoservicio bajo demanda. El eje Y representa la *Carga de CPU (%)* y el eje X la cantidad de *Usuarios por segundo*. Luego del punto determinado por 150 usuarios por segundo y 62% de carga, la CPU disminuye debido a su autoescala en base a la demanda de los usuarios, por lo que queda demostrada la característica.

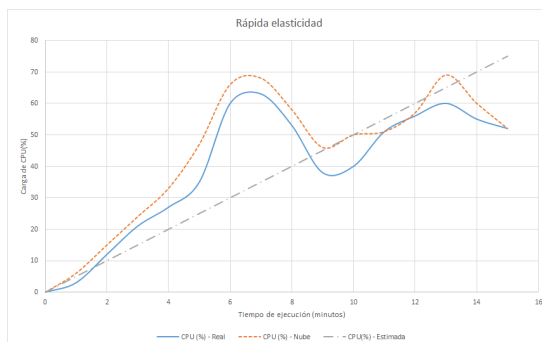


Figura 3. Rápida elasticidad. Se aprecian 3 curvas de la *Carga de CPU (%)*, la lisa representa a una instancia única, la de guión corto a la aplicación en la nube (con autoescalado y balanceo) y la de guión largo a la carga estimada. Como la curva de la aplicación en la nube es similar a las de instancia única y estimada, se comprueba la característica ya que, independientemente a la demanda, la adaptación es dinámica.

6. Conclusión

Las organizaciones encuentran atractiva a la nube por los múltiples beneficios que ofrece, principalmente con la facilidad de gestión de recursos, la reducción de costos mediante el modelo *Pay as you go* y el incremento de la productividad.

En la actualidad no se cuenta con una guía, a nuestro entender lo suficientemente completa y actualizada, que permita migrar a la nube aplicaciones que son desplegadas en entorno tradicionales. Este trabajo ofrece una metodología que se encarga de cubrir dicha necesidad.

La metodología implementada se encuentra basada en el esquema conceptual de Amazon [3], aunque con ciertas modificaciones debido al direccionamiento hacia sus productos y la falta de detalles técnicos para la toma de decisiones. La metodología propone la definición de un esquema *Cloud Ready*, permitiendo así obtener el mayor de los beneficios del nuevo entorno.

La utilización de la metodología en el caso de estudio arroja resultados alentadores, consolidando al proceso y los cambios hechos al esquema base adoptado, favoreciendo al entorno nube por sobre el tradicional haciendo notar principalmente la reducción de costos y la mejora del rendimiento frente a cargas variables.

Referencias

1. Gartner IT Key Metrics Data, 2012 IT ENTERPRISE, SUMMARY REPORT, <http://www.gartner.com/resId=1872515>.
2. Management Solutions, *La nube: oportunidades y retos para los integrantes de la cadena de valor*, 2012.
3. Jinesh Varia, *Migrating your Existing Applications to the AWS Cloud - A Phase-driven Approach to Cloud Migration*, October 2010.

4. Rashmi, Dr. Shabana Mehfuz, Dr.G.Sahoo, *A five-phased approach for the cloud migration*, ISSN 2250-2459, Volume 2, Issue 4, April 2012.
5. Michael Menzel, Rajiv Ranjan, *CloudGenius: Decision Support for Web Server Cloud Migration*, April 16–20, 2012, Lyon, France.
6. Rashmi Rai, Dr. Shabana Mehfuz, Dr. G. Sahoo, *Efficient Migration of Application to Clouds: Analysis and Comparison*, GSTF Journal on Computing (JoC), Vol. 3 No. 3, Dec 2013.
7. Raghavan P, Shiva Murthy G, *Migration of Legacy Application to Cloud Environment: A Survey*.
8. Pooyan Jamshidi, Aakash Ahmad, and Claus Pahl, *Cloud Migration Research: A Systematic Review*, IEEE.
9. Juan Marcelo Pintos, Carlos Núñez Castillo, Fabio López Pires, *Evaluation and Comparison Framework for Platform as a Service Providers*.
10. PCMAG CR, <http://www.pcmag.com/encyclopedia/term/67090/cloud-ready>.
11. Kyle Brown, Mike Capern, *Top 9 rules for cloud applications*, IBM, April 09, 2014.
12. Matt Stine, *Migrating to Cloud-Native Application Architectures (Twelve factor)*, O'Reilly.
13. The twelve-factor App, <https://12factor.net/>.
14. AWS Well-Architected Framework. November 2016.
15. Architecting for the Cloud - AWS Best Practices. February 2016.
16. Jinesh Varia, *Cloud Architectures*, June 2008.
17. Maram Mohammed Falatah, Omar Abdullah Batarf, *Cloud Scalability Considerations*, International Journal of Computer Science & Engineering Survey (IJCSES) Vol.5, No.4, August 2014.
18. James Hamilton - Windows Live Services Platform, *On Designing and Deploying Internet-Scale Services*.
19. Amazon Web Services, <https://aws.amazon.com/es/>.
20. Gartner: Magic Quadrant for Cloud Infrastructure as a Service, Lydia Leong, Mayo, 2015, <https://aws.amazon.com/es/resources/gartner-2015-mq-learn-more/>.
21. Gartner: Magic Quadrant for Public Cloud Storage Services, Lydia Leong, Junio, 2015, <http://aws.amazon.com/es/resources/gartner-storage-mq-learn-more/>.
22. Gartner: Critical Capabilities for Public Cloud Infrastructure as a Service, Lydia Leong, Octubre, 2015, <https://aws.amazon.com/es/resources/gartner-cc-learn-more/>.
23. Forrester Wave: Public Cloud Service Providers' Security, <https://aws.amazon.com/es/resources/forrester-security-wave-learn-more/>.
24. Tobias Binz, Frank Leymann, David Schumm, *CMotion: A Framework for Migration of Applications into and between Clouds*, Institute of Architecture of Application Systems, University of Stuttgart, Germany.
25. National Institute of Standards and Technology, *The NIST Definition of Cloud Computing*, NIST Special Publication 800-145.
26. Amazon EC2, <https://aws.amazon.com/es/ec2/>.
27. Amazon Elastic Block Store, <https://aws.amazon.com/es/ebs/>.
28. AWS Relational Database Service (RDS), <https://aws.amazon.com/es/rds/>.
29. Amazon S3, <https://aws.amazon.com/es/s3/>.
30. Amazon Auto Scaling, <https://aws.amazon.com/es/autoscaling/>.
31. AWS Elastic Load Balancing, <https://aws.amazon.com/es/elasticloadbalancing/>.
32. AWS Total Cost of Ownership (TCO) Calculator, <https://awstccalculator.com/>.

GPU Performance and Power Consumption Analysis: A DCT based denoising application

Martín Pi Puig¹, Laura De Giusti¹, Marcelo Naiouf¹, Armando De Giusti^{1,2}

¹Instituto de Investigación en Informática LIDI (III-LIDI), Facultad de Informática, Universidad Nacional de La Plata, 50 y 120 2do piso, La Plata, Argentina.

²CONICET – Consejo Nacional de Investigaciones Científicas y Técnicas, CABA, Argentina

{mpipuig, ldgiusti, mnaiouf, degiusti}@lidi.info.unlp.edu.ar

Abstract. It is known that energy and power consumption are becoming serious metrics in the design of high performance workstations because of heat dissipation problems. In the last years, GPU accelerators have been integrating many of these expensive systems despite they are embedding more and more transistors on their chips producing a quick increase of power consumption requirements. This paper analyzes an image processing application, in particular a Discrete Cosine Transform denoising algorithm, in terms of CPU and GPU performance and energy consumption. Specifically, we want to compare single-threaded and multithreaded CPU versions with a GPU version, and characterize the execution time, true instant power and average energy consumption to deflate the idea that GPUs are non-green computing devices.

Keywords: Denoising, Power, Energy, GPU, NVML, RAPL.

1. Introduction

The computing scenario has changed substantially since the introduction of accelerators, principally GPGPU (short for general purpose computing on graphics processing units). Therefore, the number of devices with GPUs and the amount of GPU accelerated applications increased more and more over the past years.

These devices have drawn the attention of the research community because they have a great computational power next to a high memory bandwidth, and are formidably suited for massively data parallel computation (Single Instruction Multiple Threads applications). Coming along with these features, the energy consumption of GPU containers like high performance workstations and personal computers became a real problem [1]. Some direct consequences of its higher power consumption are growing dissipation of heat, more complex cooling solutions, and noisier fans [2].

As a result, power dissipation must be reduced without losing computing performance. Further, the peak power of latest Nvidia and AMD GPUs is as high as 300W, while a typical CPU consumes only 80W at TDP. This does not indicate that

the GPU has lower energy efficiency since the increasing advantage in performance can offset the larger power consumption [3][4]. For instance, in the June 2017 Green 500 ranking, 8 of 10 top computer systems incorporate Nvidia accelerators.

Historically, GPGPU researching has focused primary on accelerating scientific applications [5] such as physical simulations, medical analysis, image and video processing. This work not only accelerates an image processing algorithm via GPU, it also makes a description of performance, power and energy consumption of CPU and GPU.

Generally, the energy quantification process relies on two approaches: hardware based and software based measurement. In first, a physical measuring device is attached between the power supply and the many-core device. Then, the electrical power can be computed by multiplying current and voltage. Also, energy consumption can be determined by integrating the power over total execution time [6]. On the other hand, latest high-end Nvidia GPUs provide the possibility to read the current power consumption by software through the Nvidia Management Library (NVML) [7]. In addition, Intel CPUs present a set of counters providing energy and power consumption information through Running Average Power Limit (RAPL) software.

In this case, we raise the image denoising issue, where zero-mean white Gaussian additive noise must be removed from a given set of images. The approach taken is based on a simple Discrete Cosine Transform (DCT) hard thresholding with a given scale factor over noise standard deviation.

Specifically, we implement a serial CPU version of the algorithm and two additional versions: a multithreaded CPU version and a GPU version, where the computation runs exclusively on the accelerator. Also, we evaluate their performance, instant power, and energy consumption.

The rest of the paper is organized as follows. Section 2 presents the background for the proposed work. Section 3 describes the main performance and power results and finally Section 4 makes a brief conclusion of obtained data.

2. Background

This section presents some background material in order to put our research in perspective, particularly the DCT denoising theory, a brief review of power consumption measuring scenarios and a description of the chosen experiments.

2.1 DCT Image denoising

This work addresses the classic denoising problem: an image is contaminated by noise in its acquisition or transmission. Generally, pictures taken with both digital cameras and conventional film cameras will pick up noise from a variety of sources. Further use of these images will often require that the noise be removed for aesthetic purposes as in artistic work or marketing, or for practical purposes such as computer vision.

Signal denoising aims to estimate the original image while retaining as much as possible the important figure features. Thus,

$$y(i, j) = x(i, j) + n(i, j), \quad (1)$$

where $y(i, j)$ is the examined value, $x(i, j)$ is the original value and $n(i, j)$ is the noise perturbation at pixel i, j . Besides, noise is often modeled as a zero-mean Gaussian process with a given standard deviation σ . This signal is independent of the original image and it commonly appears as high frequency coefficients, but useful signals appear as either low frequency or smoother details.

When we decompose a given signal using DCT, we are left with a set of spectral coefficients that correlates to details in the image. If these details are small enough, they can be omitted without substantially affecting the main picture quality. Additionally, these values are often those associated with noise. Therefore, by setting these coefficients to zero, we are essentially reducing/killing the noise. This becomes the basic concept behind thresholding: setting all frequency coefficients that are less than a specific threshold to zero and use these coefficients in an inverse DCT process to reconstruct the image [8].

So, the denoising process could be reduced to: apply DCT to (1) to get the spectral coefficients $Y(i, j)$. Eventually tweak these coefficients using hard thresholding and then take inverse-DCT to get the denoised image [9].

Then, the thresholding operation (2) can be modeled as

$$Y(i, j) = \begin{cases} Y(i, j), & \text{if } |Y(i, j)| > T \\ 0, & \text{if } |Y(i, j)| \leq T \end{cases} \quad (2)$$

where T is the threshold. For this implementation, a threshold value of $T = 3.5 * \sigma$ is selected. The original signal and the hard thresholding operation are portrayed in Fig. 1.

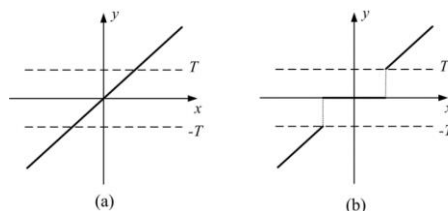


Fig. 1. Thresholding process, (a) original signal, (b) hard thresholding.

Since the present denoising algorithm uses a forward and inverse Fourier-related transform, specifically a DCT, this process becomes the primary computational module of the whole system. Thus, it has to be optimized to achieve an implementation as fast as possible.

There are several types of DCT. The most powerful and popular is two-dimensional symmetric variation of the transform that operates on 8×8 blocks (DCT 8×8). The DCT 8×8 is utilized in JPEG compression routines and has become a de-facto standard in image and video coding algorithms [10]. Because of GPU is a high segmented computational device, this block-division DCT approach is exploited in the present

work. Therefore, 8x8 blocks are processed in a sequential order by a single core in the serial denoising algorithm while the multithreaded and gpu versions makes a specific inter-block parallelization.

2.2 Power monitoring approach

On the power consumption measurement, we can discriminate two direct ways.

Firstly, a hardware-oriented solution, where an additional power capable device is attached between the computational equipment and the power supply. In particular, this method consists on sampling electrical current while voltage remains constant. However, voltage could be eventually measured by clamp probes. Then, power is computed by multiplying the two signals, and total energy is calculated as the integral of the power over the execution time. This approach might lead to several errors caused by clamp sampling variations and instant power approximation (by multiplication means).

This hardware technique can be invasive or not. In the invasive way, we measure the power consumption interfering the power supply lines generally through direct current sensors (Hall effect). In contrast, an extern system measurement can be carried out by intercepting the power supply input through alternating current and voltage tools.

On the other hand, energy consumption could be supervised across particular software interfaces. In this case, real time power information could be accessed through built in on-board sensors or by a specific power estimation model. In this last approach, the power scheme is feeded by hardware counters data.

In this work, we employs a software measurement approach, where CPU power consumption is monitored through Intel RAPL interface and GPU power information is gathered using Nvidia NVML.

RAPL provides a set of counters producing energy and power consumption information. It uses a software power model that estimates energy usage by querying hardware performance counters and I/O models [11] and results are available to the user via a model specific register (MSR). This power model has been validated by Intel in [12].

NVML is an API for monitoring and managing different Nvidia GPUs features like the ability to set/unset ECC (Error Correction Code), or to monitor memory usage, temperature, utilization rates, and more. Also, this library provides the ability to query power consumption at runtime through the built in power sensor.

2.3 Measurement description

For all implementations the measuring method consists on: running the computational algorithm in one/multiple threads and the RAPL/NVML code in another thread using Pthreads, which provides negligible overhead.

Then, the only communication between the power measuring threads and computational threads is a flag variable. Furthermore, power readings are stored in a global structure. Finally, RAPL and NVML threads stop when the shared flag is reset,

which is when the CPU/GPU finalize its execution. This brief description is showed in Fig. 2.

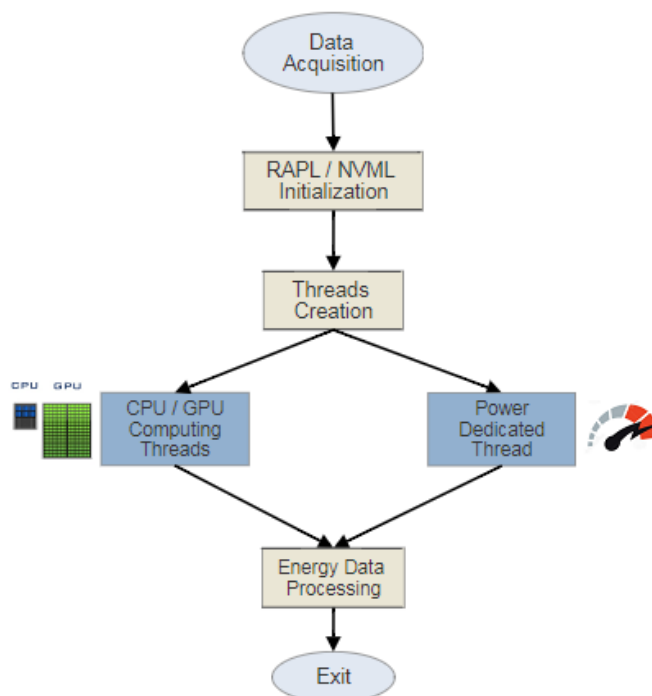


Fig. 2. Measurement method.

Besides, a critical parameter to define is the power sampling interval. Some previous research works on GPU power consumption have focused on this point. Lang and Runger et al. [6] shows that the optimal NVML sampling frequency is 50Hz (20ms). In contrast, Burtscher et al. [13] demonstrates that the maximum frequency supported by hardware is 66.7Hz (15ms). In addition, Kasichayanula et al. [1] and Weaver et al. [14] recommends a sampling frequency of 62.5Hz (16ms).

In this work, we use a NVML (GPU) and RAPL (CPU) sampling frequency of 62.5Hz. Therefore, we wrote our own tool to query the GPU sensor via NVML interface and to obtain estimated CPU power data through RAPL.

It is also necessary to mention that NVML power information refers to whole GPU board, including DC voltage converters, integrated circuits like video chip and bridge, memories, etc. The returned value is accurate to within a range of +/- 5 milliwatts. However, Intel RAPL provides entire CPU package power data (including cores, uncore circuit and DRAM memory) with +/- 1 milliwatt precision.

Generally, power consumption of commercial CPUs and GPUs can be described as the sum of static power, dynamic power and the impact of ascending temperature. Thus,

$$P = P_S + P_D + P_T, \quad (3)$$

where P_S is static power, P_D is dynamic power and P_T is the temperature effect on power. Static power depends on chip layout and circuit technology, and is independent of workload execution. Dynamic power results from transistors switching overhead. Heat also has an impact on power due to transistors current leakage increases with temperature [15].

In this work, static power is measured when no workload is executed and while none of the CPU and GPU resources are turned off. Furthermore, temperature contribution is considered negligible because of GPU algorithm executes in a few seconds. Despite CPU code runs a longer time, the measured temperature does not increase significantly.

Finally, as dynamic power depends on the specific workload to be processed, it is the chosen variable to analyze in results section.

2.4 Target platform

The hardware platform includes a computing server with an Intel Core i3-4170 (4th gen.) processor with a 3MB cache and 2 physical cores (4 threads) at 3.70GHz. This processor is attached to a Gigabyte H81M-H motherboard, which additionally holds an 8GB RAM memory. The server also contains a Nvidia Tesla C2075 scientific computing GPU equipped with 14 multiprocessors including 32 cuda cores each (448 total cores) and a 6GB GDDR5 global memory.

Moreover, our software configuration combines a 64-bit Ubuntu distribution (Linux kernel 3.2.0) with Nvidia Driver v331.62 and CUDA Toolkit 6.0.

2.5 Experiments

As mentioned before, we implement three versions of the DCT based denoising algorithm. In particular, a serial CPU version, a multithreaded CPU version and a GPU version. The multithreaded implementation exploits the OpenMP API [16] to interpret user directives in order to create and manage threads execution.

To increase performance, both CPU versions are compiled with and without g++ optimization flags. For detailed information, go to [17].

Therefore, the five scenarios are:

- Serial version
- Serial version (compiled with optimization flags)
- Multithreaded version
- Multithreaded version (compiled with optimization flags)
- GPU version

In particular, the GPU implementation uses a 2D 8x8 threads block configuration to optimize DCT/iDCT performance. Also, we exploit shared memory and texture memory, operate with asynchronous CPU/GPU memory copies (pinned memory) and apply a multi-stream based computation.

Therefore, we took a total of 10 image datasets where all of them have a sum of 50 RGB images each. Every set is characterized by a specific image size (square or not). Such dimensions are 1MP, 2MP, 4MP, 6MP, 8MP, 14MP, 20MP, 28MP, 34MP and

70MP, where MP (megapixels) refers to the millions of pixels in the picture. In order to simulate the real acquisition, we contaminate all images with additive Gaussian white noise with an input noise level equal to 20 db.

So, we process ten image datasets through the five experiments evaluating in each set the execution time, power and energy consumption. In addition, each experiment is executed several times to obtain more accurate information.

3. Results

3.1 Denoising capability

In order to reveal the solution effectiveness, we took an RGB image from the 34MP dataset and evaluate the visual and technical output of the denoising implementation. Fig. 3 displays the obtained data.

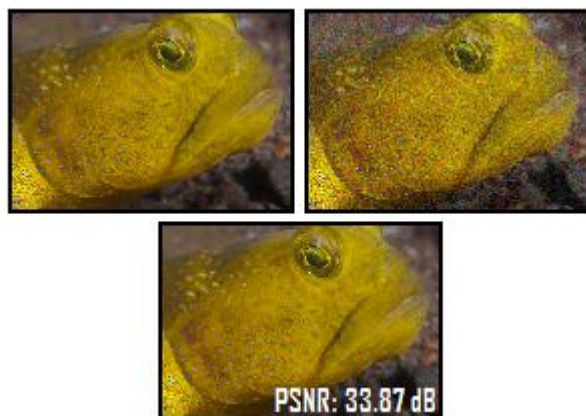


Fig. 3. From top to bottom, left to right: original, noisy and denoised images.

In this example, the algorithm returns a highly acceptable denoised sample with a PSNR value of 33.87 dB.

3.2 Performance and Energy analysis

In this section, we introduce the evaluated computation time and some power related measurements like instant dynamic power (load dependent observed power) and total energy consumption.

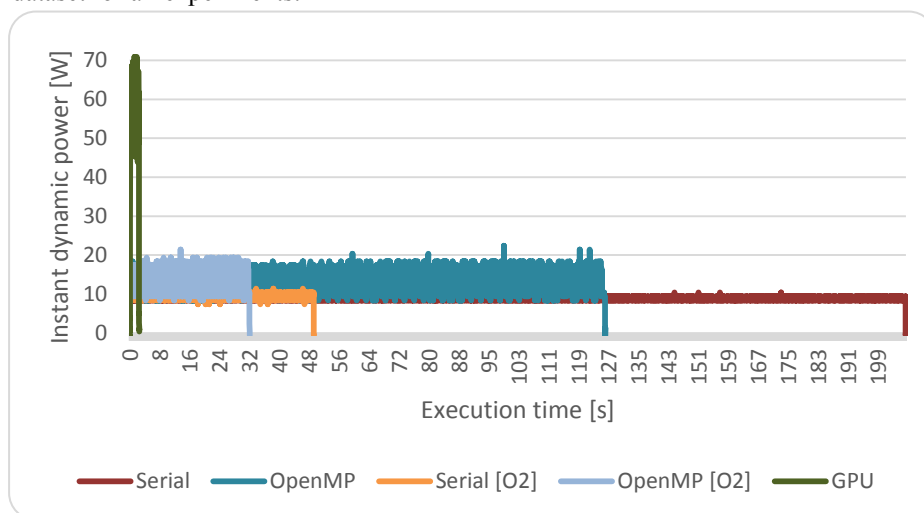
As explained in measurement section, total power fluctuation associates directly with dynamic power variation. Table I shows static, dynamic and total average power for the 6th image dataset. As we mentioned, temperature impact is considered negligible in this work.

Table I. Observed power consumption (Watts)

	CPU	GPU
Static	2,61	77,98
Dynamic	8,46	51,41
Total	11,09	129,42

Static power measurements correspond with CPU and GPU idle state and it is almost constant for all dataset. Then, dynamic values vary depending on the specific workload.

So as to calculate and display dynamic power data, we use the measured static power values shown above: 2,61W for the Intel i3 CPU and 77,98W for the Nvidia Tesla GPU. These numbers are the zero reference for CPU and GPU dynamic power plot. Fig. 4 presents the execution time and instant dynamic power of the 6th image dataset for all experiments.

**Fig. 4.** Instant dynamic power.

As shown in the graph, the zero reference (idle state) represents beginning and end of algorithm execution for all cases. Notice that serial versions corresponds to less power consumption while multithreaded versions raise this parameter. Moreover, the GPU implementation presents a huge difference in power consumption regarding serial and OpenMP versions.

In contrast, GPU execution time is minimal compared to CPU versions. Fig. 5 introduces this metric for all image datasets.

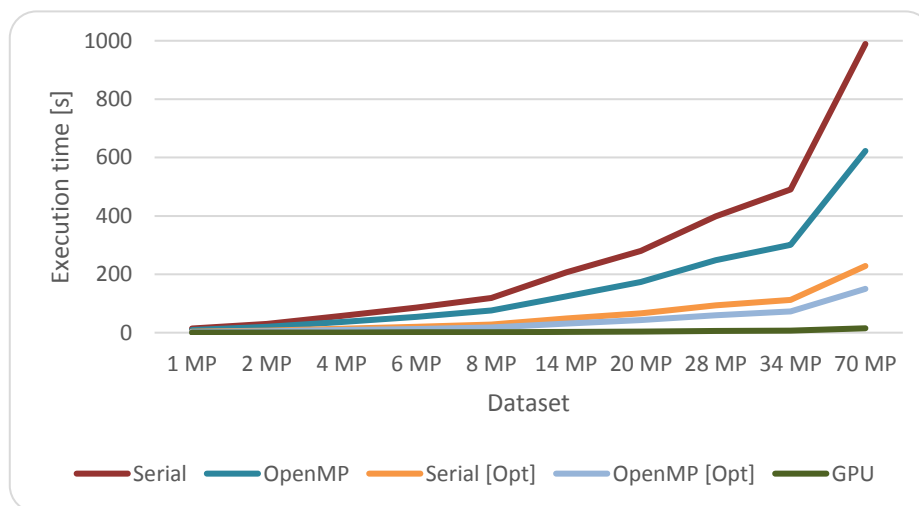


Fig. 5. Execution time.

As shown, CPU and GPU execution time increases with image size. In the first dataset, GPU outperforms multithreaded optimized CPU version by 10.2x and single-threaded optimized CPU version by 16.3x. Moreover, if these implementations are not built with compiler optimizations, the achieved speedup raises to 42.1x and 65.9x, respectively. These speedup values remain almost constant throughout all image datasets.

Having exposed instant power measurements and total execution time, we can now present the average energy consumption for all solutions. This metric can be computed by multiplying average power consumption (static + dynamic) over total execution time.

Whereas measured power consumption for all experiments remains almost constant, it is probably that energy data will present a nearly linear behavior over the performance one. Then, energy consumption and execution time curves will be proportional and the offset between them will be set by power consumption values.

Fig. 6 shows the average energy consumption for all experiments.

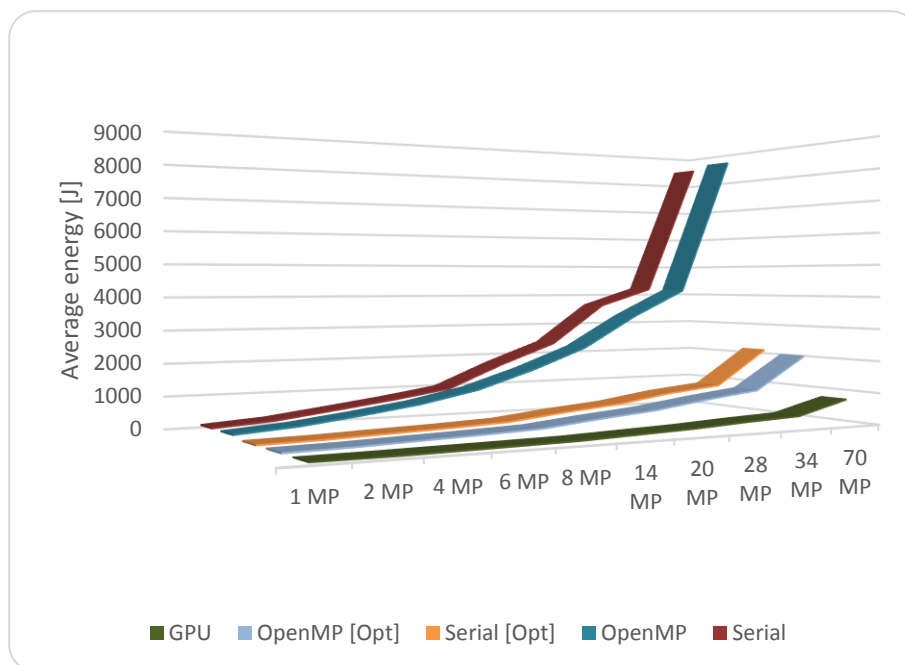


Fig. 6. Energy consumption.

GPU energy consumption is significantly lower than CPU solutions. For instance, in the 70MP dataset the multithreaded optimized CPU implementation returns an average energy value equal to 2068J while the GPU average energy consumption drops to 809J (2.5x). Further, GPU outperforms the single-threaded optimized version by 2.7x. Similarly, when using no compiler optimizations, the GPU beat CPU applications by 10.6x and 10.4x in total energy consumption.

4. Conclusion and future work

This article presents some measurements and a detailed analysis of the performance and energy consumption of a DCT based denoising algorithm. We implement three different codes including a single-threaded CPU version, a multithreaded CPU version and a GPU version. CPU implementations also include a compiler-optimized variant.

We showed that GPU accelerated code can outperform CPU serial and multithreaded programs in terms of performance and total energy consumption. Also, we expose that compiler based optimizations are a vital resource for lower CPU execution time and energy consumption. However, this optimized GPU implementation can offset CPU execution time by almost 66x. Despite GPU measured power is huge, applications can finish faster so the total energy consumption is notably less than CPU versions.

In future work, we plan to combine software based power data with physical measurements directly from CPU and GPU. Therefore, this accurate information combined with performance counters data can be the starting point to design a new model to predict CPU and GPU power consumption.

5. References

- [1] K. Kasichayanula, D. Terpstra. Power Aware Computing on GPUs. In: Symposium on Application Accelerators in High Performance Computing (2012).
- [2] X. Ma, M. Dong, L. Zhong, Z. Deng. Statistical Power Consumption Analysis and Modeling for GPU-based Computing. In: HotPower ACM SOSP Workshop Power Aware Computing and Systems (2009).
- [3] H. Nagasaka, N. Maruyama. Statistical Power Modeling of GPU Kernels Using Performance Counters. In: International Conference on Green Computing (2010).
- [4] R. Suda, Da Qi Ren. Accurate Measurements and Precise Modeling of Power Dissipation of CUDA Kernels toward Power Optimized High Performance CPU-GPU Computing. In: International Conference on Parallel and Distributed Computing, Applications and Technologies (2009).
- [5] S. Huang, S. Xiao, W. Feng. On the energy efficiency of Graphic Processing Units for scientific computing. In: IEEE Symposium on Parallel & Distributed Processing (2009).
- [6] J. Lang, G. Runger. High-Resolution Power Profiling of GPU Functions Using Low-Resolution Measurement. In: Euro-Par 2013 Parallel Processing Lecture Notes in Computer Science Volume 8097, Springer, pp 801–812 (2013).
- [7] NVML Reference Guide, <http://docs.nvidia.com/deploy/nvml-api/index.html>.
- [8] Signal denoising, <http://eeweb.poly.edu/iselesni/DoubleSoftware/signal.html>.
- [9] M. Biswas, H. Om. A New Soft-Thresholding Image Denoising Method. In: 2nd International Conference on Communication, Computing & Security (2012).
- [10] A. Obukhov, A. Kharlamov. Discrete Cosine Transform for 8x8 blocks with CUDA. Nvidia whitepaper (2008).
- [11] Running Average Power Limit – RAPL, <https://01.org/blogs/2014/running-average-power-limit-%E2%80%93rapl>.
- [12] E. Rotem, A. Naveh, D. Rajwan, A. Anathkrishnan, E. Weissmann, Power-management architecture of the Intel microarchitecture codenamed Sandy Bridge. In: IEEE Micro, vol. 32, no. 2, pp. 20–27, 2012.
- [13] M. Burtscher, I. Zecena, Z. Zong. Measuring GPU Power with the K20 Built-in Sensor. In: GPGPU-7 Proceedings of Workshop on General Purpose Processing Using GPUs (2014).
- [14] V. Weaver, M. Johnson, K. Kasichayanula, J. Ralph, P. Luszczek, D. Terpstra, S. Moore. Measuring Energy and Power with PAPI. In: 41st International Conference on Parallel Processing Workshops (2012).
- [15] D. Li, S. Byna, S. Chakradhar. Energy-Aware Workload Consolidation on GPU.
- [16] OpenMP Application Programming Interface, <http://www.openmp.org/wp-content/uploads/openmp-4.5.pdf>.
- [17] GCC Command Options: Options That Control Optimization, <https://gcc.gnu.org/onlinedocs/gcc-4.6.2/gcc/Optimize-Options.html>.

High Performance Computing for Tumor Propagation Agent-based Model

Ghazal Tashakor, Emilio Luque, Remo Suppi

Universitat Autònoma de Barcelona,
Department Computer Architecture & Operating Systems,
School of Engineering, Campus Bellaterra,
Cerdanyola del Vallès, Barcelona, 08193, Spain
Ghazal.Tashakor@caos.uab.cat, { Emilio.Luque, Remo.Suppi }@uab.cat

Abstract. Agent based modeling (ABM) and High Performance Computing (HPC) techniques are very popular in investigation and understanding cellular and molecular systems. The complex nature of these systems and the demand for emulation and comprehension at different levels in these models creates the expectation for new effective simulation strategies and tools. The present paper peruses the foresaid demands and the approaches for developing simulation in tumor model and its interactions using ABM and HPC. ABM allows the analysis of the actions and interactions of autonomous agents (cells in this case) to evaluate their effects on the system as a whole in order to re-create and predict the appearance of a complex phenomenon. This is a parametric model and it is necessary to explore the data model space to determine which combinations of adjustments cause the behaviors which are of interest. In this case, HPC is a useful tool to perform experiments in acceptable time.

Keywords: Agent Based Models (ABM), High Performance Simulation, Hpcnetlogo, HPC.

1 Introduction

Presumptive and dynamic modeling techniques in different simulators to generate complex biological models are promising research tools in oncology, but the need for computational techniques for large scale parameters and factors in complex biological models has not yet been addressed. Here in this paper, we perused a simulation model which has been created based on the agent-based system in Netlogo by Wilensky to emulate the development and progression of tumor and metastasis [1].

This model demonstrates the growth of a tumor and how it prevents chemical treatment. The tumor mass is recognized as a self-organized and self-regulated model of complexity and contains not only inherent properties of tumor cells (e.g., stem cells, growth, and migration) but also its interactions with both the immune and the vascular system. All types of self-organized immune systems rely on immune cells' capability to produce signaling molecules (e.g., cytokines, antibodies) in reaction to the environmental conflicts, which consequently are recognized and lead to the

dynamic changes of their factors and actions [2]. Besides, the blood vessel development, or angiogenesis, is controlled by the multiple molecules including vascular endothelial growth factor and endothelins. The tumor cells discharge both cytokines and angiogenic factors which have strong influence on both the local immune system and angiogenic responses of the immune and vascular systems [3].

The sum of the interactions between the tumor, the immune and the vascular local systems determines both the success and the rate of tumor progression. The complexity of these interactions during tumor progression creates more complex interactions that can be modeled and simulated using agent-based systems (ABS) to provide a dynamic representation of the tumor development to extract the emergent information and knowledge of the system.

Cancer research is generating a large amount of knowledge using information provided by the genome. This knowledge can be used, with the support of the computational techniques, in addition to those described in the laboratory and clinics that can be complex and slow, to create parametric models based on principles/rules at the cell level that allow us to understand the emergent behavior of the tumor growth/propagation at system level [4, 8].

The high-grade tumor (World Health Organization grade III & IV) expresses a cellular hierarchy with the self-renewing called tumorigenic cancer stem cells (CSCs) which is an attractive model to explain many aspects of tumor behavior and its different growth conditions[8].

Most researches and observation of the cancer stem cell growth has been focused on the discovery of mutations and variations of tumor suppressor genes with recessive loss of the function. Consequently, the design of the computational models, especially multi-agent models and using high performance computing resources and techniques, will allow the increase of the knowledge on this subject and will accelerate the research and facilitate the clinical experiments. Although it is a promising technique, it is very important to consider that the computational models must be validated to represent the real system and this may be a complex task, but once this step is done, it can be used to understand how CSCs emerge.

This paper is organized as follows; Section 2 describes the research about general tumor and cancer models. Section 3 presents the main concepts about the extended model used to demonstrate the possibilities of ABM and HPC. In section 4, our experimental environment, named HPCNetlogo, is described which is designed to facilitate the scientists doing the HPC experimentation without the complexity of high performance environment. Section 5 shows the results of the experimentation from HPC point of view (speedup, efficiency and scalability of the model) for different parametric executions to explore some parts of the possibilities of the model data space. Finally, section 6 presents the conclusion and open lines of this work.

2 State of the Art

There is a large set of references where their authors have used probabilistic or mathematical models to simulate the growth of active cancer cells using different

phases and cell cycle durations. These models can be classified as oriented toward the vasculature without considering the tumor mass or others that determine the expansion of the tumor in the presence of non-evolutionary angiogenesis. [9]

However, models have evolved and the tumor mass can be considered as a self-organized and self-regulated system that includes, in addition to the properties of the tumor cells (stem cells, growth, migration, etc.), interactions with the system and angiogenesis. This model, which is considered very complex, considers the self-organization of the immune system (to analyze the immune cells that produce signaling molecules) or the evolution of angiogenesis and how they produce dynamic changes in the system. [6, 8, 10-12]

Agent based modeling (ABM) is a discrete-based modeling that offers many possibilities and advantages over other methods for studying and simulating cancer development [6]. ABM allows modelers to study the largest influences on tumor behavior by changing factors and parameters and generating repetitions to obtain data stability (statistically). There are several types of ABM techniques which have been used in cancer research such as lattice-based/free/gas, cellular potts, and sub-cellular elements modelling methods [2, 6].

Wilensky's tumor model [1] in the NetLogo library is used as the base for the present case study. This model is designed as a self-organized ABM model that illustrates the growth of a tumor and how it resists chemical treatment. In this case the model consists of two kinds of cells: stem cells and transitory cells and includes controls to kill transitory cells that are younger, to kill a stem cell or moving cells or visual information about the total number of living cells or their trajectory.

Considering the complexity of interactions between tumor cells and vascularization during tumor progression, the ABM model provides a more dynamic representation of the tumor development. However, since the number of parameters and agents required to give a realistic approach increases, it is necessary to use high-performance computing to make an exploration of the model data space in the acceptable time. In addition, since the model includes stochastic variables, it is necessary to perform a large set of simulations to obtain statistically acceptable values which the HPC needs.

3 ABM Tumor Model

The design of the agent-based tumor model of the present work is based on the Wilensky model [1] and considers the premises stated by Hallmarks of Cancer [5].

In this model two kinds of cells are considered: stem cells and transitory cells. Each cell is represented by an agent and tumor cells can breed, move, or die where the time is measured in simulation steps (ticks). Initial tumor cells are blue and the cells change their color depending on their age (different colors from red palette). A stem cell can divide either asymmetrically or symmetrically during the mitosis. In this case study, we considered malignant tumors where a symmetric mitosis is followed. In this process the stem cell divides symmetrically in two stem cells and the first child of this division remains static, but the second cell moves to distant organs to generate a metastasis. In this model, we considered only metastasis (and not invasion) as the

mechanism of the spreading cancer cells, reaching it to an organ different from the original tumor using blood or lymphatic paths. In the model, the metastasis is red and made of cells that die young. As the tumor propagate and get larger, stem cells reproduce and die younger.

Moreover, the model includes some considerations about hallmarks listed in [4] with the acquired capabilities of cancer:

1. Self-sufficiency in growth signals
2. Insensitivity to antigrowth signals
3. Evading apoptosis
4. Enabling replicative immortality
5. Sustained angiogenesis
6. Cells invasion and metastasis

In order to represent this behavior, the model includes parameters for modeling the number of initial stem cells, the growth factor in order to model the self-sufficiency of growth signals and insensitivity to anti-growth signals, three values to represent the apoptosis grade (normal, low, very low) and a replication factor in order to model the replicative immortality and angiogenesis capacity of the tumor.

The simulated model represents, with different parameters, the original stem cells, the transitory cells, and the metastasis process. The visualization of the tumor grows exponentially and the metastasis process is showed as new tumors are allocated outside the original location, evolving in an independent location. Figure 1 shows the simulation evolution of 6 original stem cells (center) with a random position metastasis in new independent tumors with growth-factor=1.25, replication-factor=high and apoptosis=normal. In this case, the evolution from the 6 stem cells reaches, in steady state, a range of 23,000 to 29,000 cells (agents).

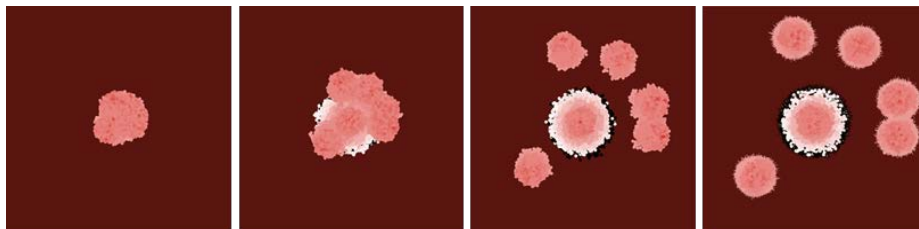


Fig. 1. Stem cells evolution and metastasis visualization (from 6 original stem cells to 23,000-29,000 in a steady state).

Figure 2 shows the steady state of a tumor metastasis visualization with 6 stem cells and the growth-factor=1.75, replication-factor=high, and apoptosis=low. As can be seen, the growth of metastasis is more aggressive and through reducing apoptosis, there is a greater number of cells that do not die, amounting to near 200,000 cells (agents).

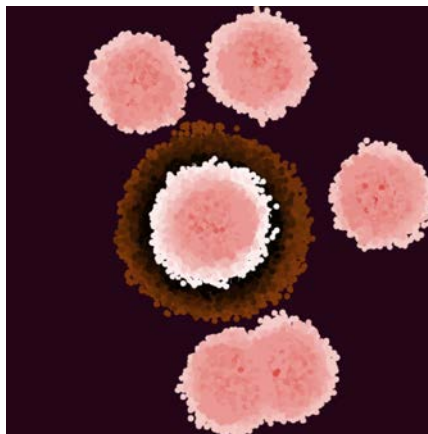


Fig. 2. 6 stem cells evolution and metastasis visualization with grow-factor=1.75, apoptosis=low and replication-factor=high (near 200,000 cells in steady state).

4 High Performance Approaches for Tumor ABM Model

In agent based model, each cell is represented as an agent with rules, parameters and interactions with the environment to reproduce the emergent growth behavior of the tumor and metastasis. In this model, all parameters, rules and interactions are parametrized and it is easy to adjust a rule or a specific constant for describing different tumor behaviors. The main problem of this type of simulation is the higher number of parameter combinations that is necessary to explore the data space of the model.

Netlogo [1] includes the Behavior Space tool that allows the exploration of the model data space using parametric executions in varying settings of the model and for recording the results of each model run. This process (*parameter sweeping*) permits us to explore the model's space of possible behaviors and determine which combinations of settings cause the behaviors of interest. The main problem of these executions is that the Behavioral Space only supports multithreading, so its performance is limited to the number of cores/threads at the local infrastructure.

The solution for this problem is to execute the parametric simulations using a HPC cluster in order to reduce the necessary time to explore a determinate model data space. This is an interesting idea but, considering HPC cluster complexity, it can be complicated for scientists to run these models in an easy way within an HPC cluster.

In order to provide an easy to use and user-friendly solution for non-technical users, we have developed a workflow, named HpcNetlogo [13] to deploy the simulation of ABM models on a HPC cluster. The execution of this workflow is unassisted and the initial user interaction is through an interface based on the web technology. In the local environment, the user configures his model, performs the

local tests, sets up the Behavior Space and, through a service, schedule his model, and indicates where and how he wants it to be run.

There is a Frontend service that analyzes the user-specified model and generates all configuration files required for the proposed simulation scenarios, creates the initialization files for the high-performance cluster queue system and then sends them to Backend. The backend executes the experiment and when it finishes, the user can see the result through the same web interface. HpcNetlogo is open source (GPLV3) and can be adapted for different cluster/queue environments (see more information in [13]). Figure 3 shows the diagram of the architecture and its web interface.

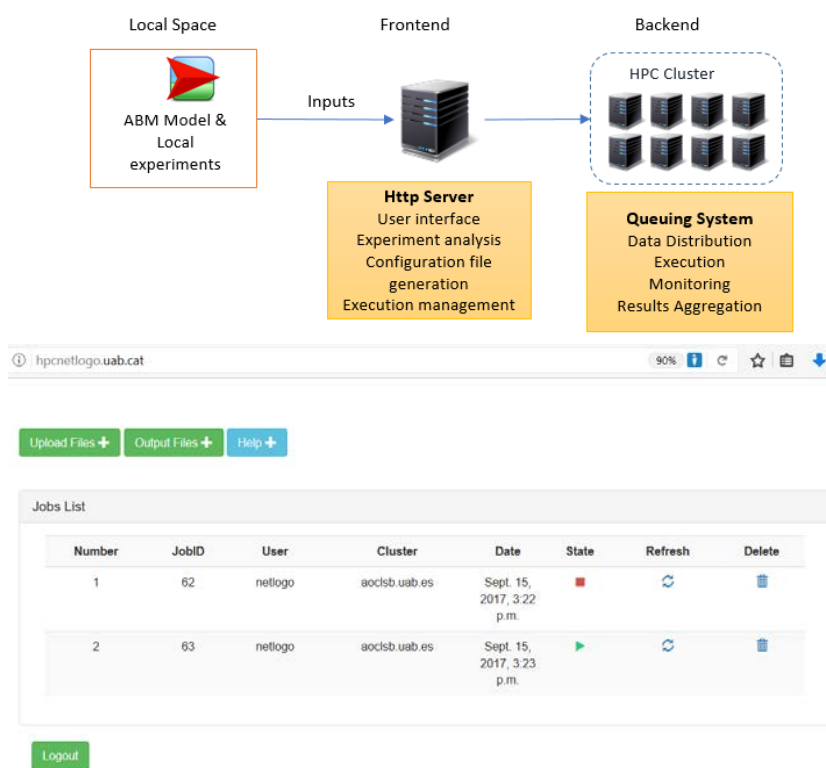


Fig. 3. HPCNetlogo Architecture and Web Interface

5 Results and Discussion

With the objective of analyzing the behavior of the developed model on an HPC cluster, a set of simulations with different parameters and values have been performed

to explore, from the point of view of performance, the behavior of the model (speedup, efficiency and scalability).

These experiments are the proof of concept considering HPC performance where the scalability, number of agents and performance of the model are analyzed but not the simulation result as growth factor, apoptosis, angiogenesis or other interesting values for an oncologist or researcher in tumor propagation (however these are available as simulation results).

The figure 4 shows the speed-up and efficiency for these executions. In this figure, the speed-up is close to lineal in the lower number of cores and gets reduced towards the higher number of cores. Although the results of speedup and efficiency (about 80%) seems good, but the main issue is the memory contention problems at the Java runtime (Netlogo) for higher core configurations and the I/O from network file system (NFS).

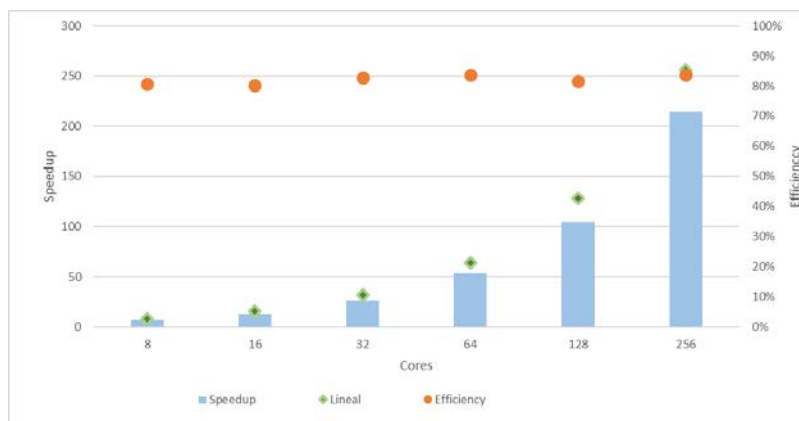


Fig. 4. Speedup and Efficiency for 8 to 256 Cores

Figure 5 shows how scalability effects of the model over the speedup and efficiency considering a grow-factor equals to 2.0 and the speedup in compare with the values of grow-factor equals to 1.25. This growth-factor, in this case study, implies that for the value of 1.25 there are about to 13,000 agents and for the value of 2, 00 there are about to 110,000 agents. As can be observed, the speedup and efficiency are affected by the grow-factor equals to 2.00 but it keeps on showing the values performance.

Figure 6 shows the impact of the shared resources at the Java runtime (NFS, memory) during the model execution for the larger number of agents: 30 initial stem cells with a growth-factor equals to 2.00 that arrives to 361,000 agents (cells). In this case, the figure show the distribution of the execution time in a box and whisker plot. As can be observed, the model is executed in a fixed number of time units (1,000) at the eight cores and runs about 2,234 seconds (average) but on 64 cores it will be run about 4,417 seconds (average). In order to solve this problem, it is necessary to change technology (java runtime) and evolve towards high performance distributed simulation environments as Care HPS [14].

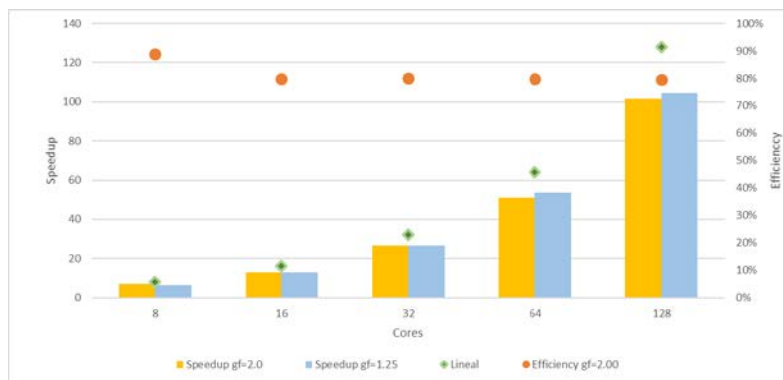


Fig. 5. Speedup and Efficiency for 13,000 and 110,000 Agents (cells)



Fig. 6. Box Diagram of Execution Time for 361,000 Agents

Figure 7 shows the scalability of the model for different number of stem cells (from 6 to 60) and to values of grow-factor (1.25 and 1.75). As can be observed the scalability of the model is lineal. In this experiment, the simulation shows that in the steady state (1.000 simulation steps) for grow-factor of 1.25 and 6 initial stem cells the tumor arrived 6.6k agents (cells) and for 60 initial stem cells the tumor arrived to 122k agents (cells). For growth-factor of 1.75 and 6 initial stem cells, the number of agents (cells) was 76k and for 60 initial stem cells the number of agents (cells) augments to 404k.

It is necessary to remark that employing HPC is extremely necessary because this type of analysis is time consuming as can be seen. At these proofs of concepts only the minimum number of repetitions to obtain stable statistics data which have been executed and only 1000 simulation steps have been performed. It is important to

consider that in other configurations/parameters may be necessary to increase the number of simulation steps to justify the necessity of employing HPC for this type of experimentation.

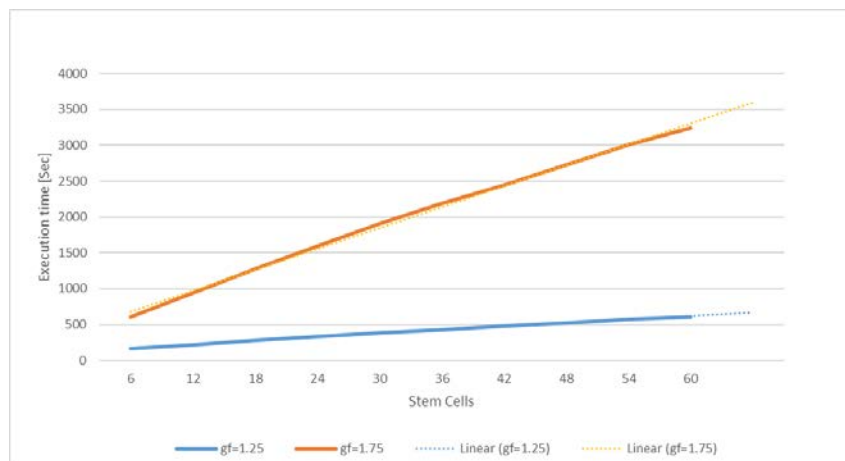


Fig. 7. Simulation runtime (average) for growth factors 1.25, 1.75 and from 6 to 60 stem cells

6 Conclusion

In this paper, we introduced a new tumor propagation model based on Wilensky model and considering the premises stated by Hallmarks of Cancer and its analysis using HPC cluster. For this experimentation, we used an easy-to-use HPC frontend, developed by the authors and named HpcNetlogo, that is an environment that allows the scientist to execute Netlogo model (in this case study) in an HPC cluster using a web interface and the Behavior Space tool included in Netlogo.

As it could be observed, the experiment shows good values for the speed-up and efficiency. However the drawback imposed by the Java Virtual Machine and access to shared resource is bigger when a high number of agents are used. These limitations could be addressed using distributed (non-parametric) simulation environments such as Care HPS [14] or Repast HPC [15] to reduce the impact of the shared resources access and to execute it with a large number of agents.

The future researches will be oriented toward:

- Detailed validation of the tumor propagation model: it is necessary to compare the results of the proposed model with the real data in order to validate the grow factor, replication rate and apoptosis factor in order to adjust this values to real ones.
- Implementation of tumor propagation model in Care HPS in order to analyze the performance values using a distributed simulation environment and its scalability for higher number of agents (cells).

- Time limitations prohibited more extensive tests to show performance & scalability value for different combinations of the main parameters. It is necessary to explore a bigger data space of the model with gradual increase in agent creation in order to analyze the impact of these in the simulation execution time.

Acknowledgments. This research has been supported by the MINECO Spain under contract TIN2014-53172-P.

7 References

1. Wilensky, U.: NetLogo Tumor model. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL. Project Integrated Simulation & Model Environment. National Science Foundation REC-9814682 and REC-0126227 (1998).
2. Dréau, Didier, et al.: An agent-based model of solid tumor progression. *Bioinformatics and Computational Biology*. LNCS, vol. 5462, pp. 187--198. Springer, Heidelberg (2009)
3. Knowles, J., Loizidou, M., Taylor, I.: Endothelin-1 and angiogenesis in cancer. *Current Vascular Pharmacology*, vol. 3, pp. 309--314. Bentham Science (2005)
4. Hanahan, D., Weinberg, R.: Hallmarks of Cancer: The Next Generation. *Cell*, vol. 4(5), pp. 646--674. Elsevier (2011)
5. Loeb, L.: Mutator phenotype may be required for multistep carcinogenesis. *Cancer Research*, vol 51(12), pp. 3073--3079. AACR (1991)
6. Wang, Z. et al.: Simulating Cancer Growth with Multiscale Agent-Based Modeling. *Seminars in cancer biology*, vol 30, pp. 70--8. Elsevier (2015)
7. Pérez-Rodríguez, G., et al.: High performance computing for three-dimensional Agent based molecular models. *Journal of Molecular Graphics and Modelling*, vol. 68 pp. 68--77. Elsevier (2016)
8. Lathia, J. D., et al.: Direct *In Vivo* evidence for tumor propagation by glioblastoma cancer stem cells. *PLoS ONE*, vol 6(9), e24807. PLOS (2011)
9. Kohandel, M., Kardar, M., Milosevic, M., Sivaloganathan, S.: Dynamics of tumor growth and combination of anti-angiogenic and cytotoxic therapies. *Physics in Medicine & Biology*, vol 52(13). IOP Publishing (2007)
10. De Pillis, L.G., Gua, W., Radunskaya A.E.: Mixed immunotherapy and chemotherapy of tumors: modeling, applications and biological interpretations. *Journal of Theoretical Biology*, vol. 238(4), pp. 841--862. Elsevier (2006)
11. Mallet, D., De Pillis, L.G.: A cellular automata model of tumor-immune system interactions. *Journal of Theoretical Biology*, vol. 239(3), pp. 334--350. Elsevier (2006)
12. Chaplain, M., McDougall, S., Anderson A: Mathematical modeling of tumor-induced angiogenesis. *Annual Review of Biomedical Engineering*, vol 8, pp. 233--257. Annual Reviews (2006)
13. Tashakor, G., Suppi, R.: HpcNetlogo. Frontend for the concurrent execution of Netlogo experiments using SGE HPC cluster. <https://github.com/hpcnetlogo/hpcnetlogo>. (2017)
14. Borges, F., Gutierrez-Milla, A., Luque, E., Suppi, R.: Care HPS: A high performance simulation tool for parallel and distributed agent-based modeling. *Future Generation Computer Systems*, vol. 68, pp. 59--73. Elsevier (2017)
15. Collier, N., North, M.: Parallel agent-based simulation with Repast for High Performance Computing. *Simulation*, vol 89(10). pp. 1215--1235. Sage Journals (2012)

Performance Analysis of ABM Distributed Simulation for Real Crowd Evacuation Scenarios

Mohammed J. Alghazzawi¹, Ghazal Tashakor¹, Francisco Borges², and Remo Suppi¹

¹ Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Barcelona, Spain.
Malghazzawi,GTashakor@caos.uab.cat, remo.suppi@uab.cat

² Federal Institute of Bahia, Santo Amaro, Bahia, Brazil
franciscoborges@ifba.edu.br

Abstract. Managing crowds is a key problem in a world with a growing population. Being able to predict and manage possible disasters directly affects the safety of crowd events. This kind of problem can be modeled using Agent-Based Model techniques and consequently simulated in order to study evacuation strategies. Our aim from this paper is to prove that this model albeit simple can be expanded and adapted for experts to test various scenarios and validate the outcome of their design. Preliminary experiments are carried out using different initial locations for the agents inside Fira of Barcelona building, whose results are presented, validated and discussed. We shown that crowd evacuation problem has bottlenecks in reality, and the initial location for all agents can increase or decrease the bottlenecks. Finally, we draw some conclusions and point out ways in which this work can be further extended.

Keywords: Crowd Evacuation, Agent-Based Model, Distributed Simulation

1 Introduction

The analysis of building evacuation has recently received increasing attention, as people are keen to assess the safety of occupants. Reports on past huge disasters indicate that human behaviors characterize evacuation during emergencies.

Cases as the Hillsborough Stadium disaster (Sheffield, England in 1989) causing 96 deaths, the Love Parade disaster (Duisburg, Germany in 2010) causing 21 deaths or the Kiss disaster (Santa Maria, Brazil in 2013) causing 239 deaths shows us how bad decisions and planning can lead to human harm. Evacuation management and planning then become a crucial issue. The disasters have occurred in numerous different countries, each of the disasters has a major contribution factor, which is given as either crowd behavior or building design. Most of the disasters have occurred due to building design, a factor that can be controlled by engineers and decision makers.

Exercises such as fire drills allow us to know the evacuation times, but not the conditions, because people have stochastic behavior that determines the variability of evacuation. It is not possible to perform drills to cover all scenarios. In these cases simulations becomes necessary. We will be able to study the evacuation strategies for a specific scenario and with this information take decisions in evacuation cases. This way helps to have knowledge to carry out what is most appropriate evacuation plan, the

safest, and fastest evacuation possible. Simulators allow us to have knowledge to take informed decisions that can predict the behavior. Decision Support Systems (DSS) have gained importance in the last years, and the usage of simulators allows predicting the behavior of a system according to a model and helping experts to take decisions.

In crowd evacuations there are random patterns among the public, therefore we need statistically reliable results in order to predict the outcome of each scenario. For that, it is important to use high performance computational solutions and tools. Care HPS [1] is an Agent-Based Model tool that helps to execute experiments in a parallel and distributed architecture. This tool allows us to manage all the components of the model, such as the environment with obstacles, and to execute our simulations efficiently with thousands of agents and thousands of simulation steps.

Therefore, in this paper, we use as a case study Fira of Barcelona model based on the Helbing model [20]. Fira model, implemented in C++, and simulated using Care HPS [2]. We analyzed the behavior of the model in different scenarios and different initial locations for the agents. This gives us more statistically reliable results.

The rest of the paper is organized as follows. We present related works where we show previous previous contribution in this field 2. Then we dedicated section 3 to show the challenges of crowd evacuation simulation and to present our Fira crowd evacuation model, and Care HPS tool that we used to implement and simulate the Fira of Barcelona model. Then, the model experiments are presented and discussed in section 4; and finally, some conclusions are drawn in section 5.

2 Related Works

In the last years there has been a growing effort in solving the evacuation safety problem. The research areas interested on the problem vary from engineering, computer science, psychology, architecture or sociology, and each one has been tackling different aspects of it. Inside the evacuation problem lay several sub problems such as: the person model, the psychological components, free navigation through the space, path planning to reach the exit, simulation techniques, and performance issues of the simulator. The aim of all these areas and components have common goals: provide a better understanding of the model and have an advanced way to implement them with simulators. For this common purpose most of the research considers the same starting premise: evacuate a certain number of agents inside a specific space.

An analytic study of crowd dynamics through exits may provide useful information for crowd control purposes. Proper understanding of the evacuation dynamics will allow, for example, improvements of pedestrian facilities designs. In particular, the dynamics of evacuation through a narrow door during an emergency is a complex problem that is not yet well understood. The possible causes for evacuation may include building fires, military or terrorist attacks, natural disasters such as earthquakes, etc. In the light of tightened homeland security, research on evacuation modeling has been gaining impetus and attracting the attention of researchers from various fields.

There are three main reasons for developing computer simulations of crowd behaviors: firstly, to test scientific theories and hypotheses; secondly, to test design strategies; and finally, to recreate the phenomena about which to theorize [3]. Computer models

for emergency and evacuation situations have been developed and much research on panics of empirical nature and carried out by researchers from the social science [4], [5], [6], [7], and [8].

A critical review offered selected simulation models of evacuation [9]. Also, authors have identified social science approaches that could improve contemporary simulation models. They argue that social sciences could provide important new directions for simulation models of emergency evacuations. [4], and [5] developed a continuous pedestrian model based on plausible interactions, and pointed out that pedestrian dynamics shows various collective phenomena, such as arching, clogging, and herding [7]. According to their findings, every simulation model should reproduce such behaviors in order to be rather realistic. They believe that the above models can serve as an example linking collective behavior of a mass psychology phenomenon (from the socio-psychological perspective) to the view of an emergent collective pattern of motion (from the perspective of physics). They suggest that the optimal behavior in escape situations is a suitable mixture of individualistic and herding behavior. [10] present a simulation model for emergency planning and crowd management purposes.

Other simulation studies, combined with optimization algorithms, aim at improving the evacuation efficiency. This can be achieved in terms of evacuation times, assessment and analysis of evacuation plans, as well as routes/path optimization. [11] has deeply studied the mathematical and analytically models, comparing and analyzing them on a systematic basis. In [12], the analysis of evacuation processes on-board is taken into consideration. The authors of the study apply Cellular Automata to reproduce crowd motion for the detection of possible bottleneck situations during the evacuation process. [13] address the building evacuation optimization problem. [14] developed a discrete-event computer simulation model for assessing evacuation programs and provide a comprehensive idea of evacuation plans for hospital buildings in the event of a possible bomb threat. Crowd pedestrian behavior was solved in areas of AI and computer graphics rendering. The goal is to introduce in agents a realistic pedestrian behavior, drive them to the goal and render big populations of individuals [15]. Rendering indexes are important metrics in these researches.

In the area of HPC to solve crowd problems, GPUs have been used to implement the crowd dynamics algorithm and rendering. Also crowd turbulence phenomena integrated on GPU cards [16], The implementation is a hybrid model between Verlet integration method and Agent Based Model. MPI was used in city evacuations distributing the space and agents [18]. These tools are normally tested with hundreds of individuals. Many crowd evacuation models using agent based model simulated people behavior by deploying cloud services for high performance simulation [19], the input is just a PNG image and the output are statistical results of the simulation executed on the cloud.

In this literature study, we note the need for further contributions in parallel and distributed solutions to carry out models with thousands of people. Also, tools can easily change the environment characteristics in order to simulate better layouts of space for fastest evacuations.

3 Crowd Evacuation Model

The present work is focused on the crowd evacuation modeling problem, crowd simulators, and their performance. Crowd models aim to understand the behavior for big populations of people in a closed space.

In modeling the problem of representing the reality can be tackled from different perspectives. Each view will have its own pros and cons based in terms such as: complexity, performance, simplifications, and aim. Modeling has allowed the humanity to advance its knowledge and improve its day living. In the particular case of evacuations and crowd dynamics there has been an increasing interest in reproduction of evacuations through several approaches during the last years, such as discrete time or continuous time models. These approaches differ on the focus of how to extract the natural characteristics of the reality and how to implement them. Each approach has pros and cons, and choosing one or another is always a trade-off.

Simulators implement the ideas designed by models, and simulation allows us to interact with the model and generate results without interacting with the real system. Our model and simulator have mainly the purpose of provide knowledge of potential problems, also analyze theories and simulate them. Moreover, one of the main final goals is to provide a decision support system where the complexity of the model and the performance is hidden to the final user by integrating the crowd evacuation model with Care HPS tool. Whereas we work with thousands of agents interacting between each other and with the environment by reproducing different scenarios. The simulation will have high computational needs that has been resolved by integrating Care HPS and improve it to be adapted with this model's needs.

Our present model is a discrete time model able to handle thousands of agents, whereas the model recovers ideas from Cellular Automata to be computationally efficient. This model reproduces crowd phenomena in Sala del Manga, which is the building 2 of Fira de Barcelona. And it can hold almost a hundred thousands of people. This environment has different zones and obstacles, which make the evacuation more complex.

Some modification made on the map for simplification purpose, while it includes chairs, important exits and the bar exit. All agents in this model has a behavior and rules to control its movements. We used an algorithm that allows agents to navigate freely avoiding obstacles and walls. The model is divided into two sub-models, agent and environmental intelligence. The agent move between cells approaching to the exit, while there can be only one agent per cell, and when agents achieve the exit, they are released.

The model is described by a set of attributes, which are (A1: Stopped, A2: Moving, A3: Evacuated, N: Approaching the exit, E: Exit reached, F: Free patch), and transitions between states are determined by the attributes of the model. The algorithm used for the environment sub-model is the potential field algorithm, which provides the shortest path and support obstacle avoidance feature. This algorithm is a recursive algorithm that visits the cells in breadth search giving higher values every time.

Figure 1 shows our environment's map with obstacles and exits. This image shows color representation of the potential field. Ignoring the dark blue color, which are ob-

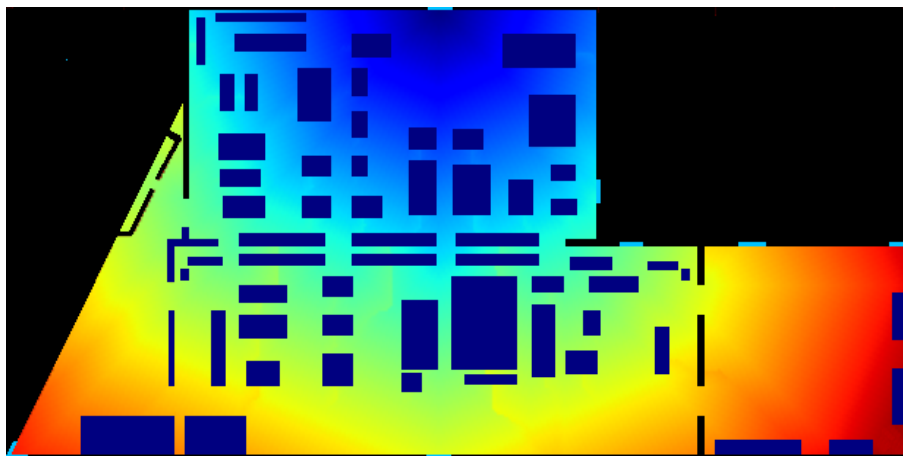


Fig. 1. Potential Field Algorithm applied on Fira de Barcelona map.

stacles. The light blue color indicates that the exit is getting closer and the red color indicates that the exit is the farthest. Notice that the obstacles and walls have the darkest blue and black because the agents will avoid these spaces as these are not approaching the exit. While the potential field is dependent on the exit, there are as many potential fields as exits. Every agent needs to have one exit assigned, therefore, one potential field. In this map we considered that the target exit for the agent is the top middle exit in the map, for this reason the blue color shown around this concern exit.

The data structure is a list of matrices where every matrix are the potential field values that are precalculated at the begging and stored in main memory.

There are several potential field algorithms and we chose a mix of the Manhattan and Chessboard potential field calculation [17]. Every individual will have knowledge of the potential field of their exit and will use it to take decisions between steps.

The (algorithm 1) implements the potential field algorithm. While the implementation is recursive, using the implicit stack of recursion to explore the nodes in breadth way. There is a global structure accessed by all the function's calls that tracks the visited nodes and the values are updated. This happen with the current value which is updated for every level of the recursion until it fills all the structure. Variables x and y are temporal variables referring to positions in the map which are being visited.

We also improved and adapted Care HPS to fit our models needs, where a statistical significant amount of simulations are executed to evacuate all agents through the available exits. The model and Care HPS tool implemented in C++ programming language. In order to test the model with Care HPS, we distributed the simulations using MPI, while the simulator were executed for several configurations with 2,8,32 and 64 OpenMP threads. Therefor we show the efficiency of these configurations in each scenario.

Algorithm 1 Potential field algorithm

```

1: procedure POTENTIALFIELD
2:    $TotalUnvisited \leftarrow 1$ 
3:    $UnvisitedPush(exit)$ 
4:   while  $TotalUnvisited \neq 0$  do
5:     for all  $x \in neighbours(r)$  do
6:       if  $x \notin unvisited$  then
7:          $UnvisitedPush(x)$ 
8:          $TotalUnvisited \leftarrow TotalUnvisited + 1$ 
9:       end if
10:    end for
11:    if  $y \in unvisited$  then
12:       $calculateFlood(y)$ 
13:    end if
14:  end while
15: end procedure

```

4 Experiments

In this section, we will present the experiments conducted over Fira of Barcelona model. Fira is a public space in Barcelona city (Spain) which holds many events that receive a considerable number of people. We will simulate 30000 of agents with the environment using same objects' position found on real space.

We made three variant scenarios (figure 2), each scenario represents different initial location for all agents, the first scenario spreads all agents randomly inside the whole environment, while the second scenario locates all agents in the middle of the environment, and the last scenario locates agents in all the corners of the environment.



Fig. 2. Agents locations for middle and corners scenarios

(Figure 2) shows the main initial location in the second and the third scenarios.

All scenarios were carried out for 30000 agents, and the agents location changed based on the scenario. These experiments were executed using one MPI process and n threads were created. We used the hybrid strip partitioning, this partitioning is one of the features available in Care HPS, where we use MPI and OpenMP threads to simulate agents behaviors. For these experiments, the partitioning was configured using one partition in MPI process and create n constant number of threads where the experiments were executed. The best execution time found in the first scenario (Figure 3) was with eight threads, and in the second scenario was with eight threads as well, but in the third scenario was with 32 threads.

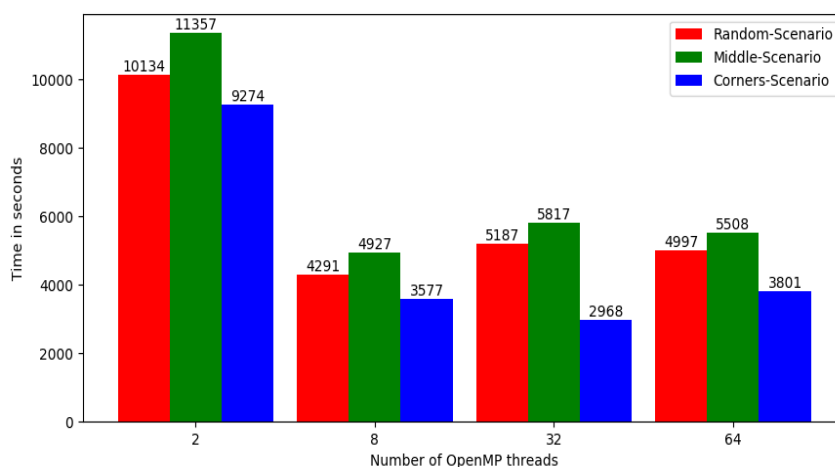


Fig. 3. Compare the scalability for all scenarios with 30000 agents.

As we note in (figure 3), the total execution time increased in the first and the second scenario while it decreased in the third scenario, and the reason behind this is that the model has a limited number of exits and all agents look up and arrive to these exits at the same time, which will create a bottleneck, Therefore, increasing the number of threads does not decrease the execution time, because of the access to exits is sequential and computing of the stopped agents is not too high. But in the third scenario as all agents were located in different corners, and each corner has an exit nearby, then the possibility to have more load on one exit is less, and this will let agents to use all exits efficiently and leave the environment smoothly through those exits.

After analyzing the data of these experiments, we show the speedup in (figure 4) to compare all scenarios from the speedup perspective. In the first scenario we obtained the best speedup with 8 threads, and in the second scenario was with 8 threads as well, but the best speedup we obtained in the third scenario was with 8 and 32 threads.

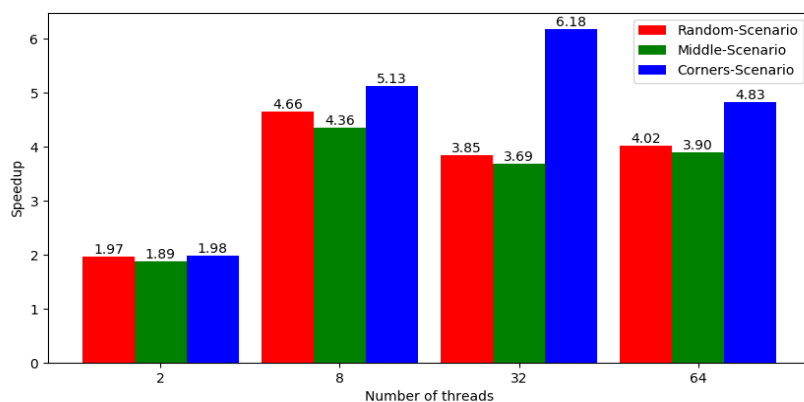


Fig. 4. Speedup comparison for all scenarios.

As we can see from the speedup diagram, the model could not take the advantage of increasing the number of threads because of the reason noted before.

In (figure 5) we show and compare the efficiency of all scenarios. It was clear that the best efficiency recorded in the third scenario where we located all agents in five corners and those corners has equal number of agents, (total number of agents divided by number of corners in the environment).

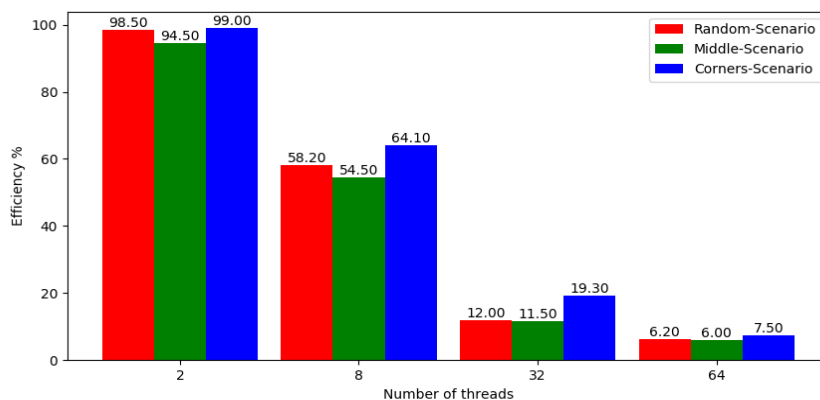


Fig. 5. Efficiency comparison for all scenarios.

From these experiments, we figured out a problem of memory contention due of the concurrency when we increase the threads number. This is a consequence of the bottle-

neck phenomena. However, from the HPC perspective, we can see that there are very good speedup and efficiency values for the configuration and scenarios we analyzed, for example when 8 threads used.

5 Conclusions

In this paper we described several experiments using different scenarios to evaluate their performance. The parallelism and performance were evaluated on different architectures and discussed. The discrete model was implemented and distributed in a cluster, dividing the total number of simulations. We also tested the model with different scenarios, whereas each scenario initiate the agents in variant locations, by locating all of them in the center of the environment, spread them randomly or spread them only on the corners. We demonstrated how the initial location for agents can affect the evacuation time and increase or decrease the crowd evacuation efficiency.

Actually this kind of scenarios can help decision makers to decide according to the time required to evacuate a certain number of people by reconfiguring the environment and change the position of obstacles and exits in the reality.

Through the experimentation, we found the best scenario is to divide all agents and locate them in the corners, in order to decrease the load on a specific exit. Also, we figured out that crowd evacuation problem has bottlenecks in the reality that will not be resolved by only increasing the number of threads, but also it needs a load balance on each exit. as we found the best case is to have them located in different areas and each area has an exit around. It occurs because of agents dispute the few exits, same way that occurs in the reality.

Acknowledgments. This publication is based upon work supported under contract TIN2014-53172-P, funded by the MINECO Spain. We thank Albert Gutierrez-Milla for his contribution in this work. All trademarks, material, images and information used in this work are propriety of their owners and here is used with academic purposes.

References

1. F. Borges and A. Gutierrez-Milla and R. Suppi and E. Luque: Strip Partitioning for Ant Colony Parallel and Distributed Discrete-event Simulation. *Procedia Computer Science*. 51, 483–492 (2015)
2. F. Borges and A. Gutierrez-Milla and E. Luque and R. Suppi: Care HPS: A high performance simulation tool for parallel and distributed agent-based modeling. *Future Generation Computer Systems*. 68, 59–73 (2017)
3. X. Pan and C. S. Han and K. Dauber and K. H. Law: A multi-agent based framework for the simulation of human and social behaviors during emergency evacuations. *AI and society*. 22, 113–132 (2007)
4. D. Helbing and I. Farkas and P. Molnar and T. Vicsek: Simulating of pedestrian crowds in normal and evacuation situations. *Pedestrian and Evacuation Dynamics*. 21–58 (2002)
5. D. Helbing and P. Molnar and I. Farkas and K. Bolay: Self organizing pedestrian movement. *Environment and Planning B: Planning and Design*. 28, 361–383 (2001)

6. C. W. Reynolds: Flocks, Herds, and Schools: A Distributed Behavioral Model. Proceedings of the 14th annual conference on Computer graphics and interactive techniques. ACM SIGGRAPH Computer Graphics. 25–34 (1987)
7. A. L. Coelho: Modelao de Evacuao de Edifcios Sujeitos Aco de um Incndio (in Portuguese). LNEC, Lisbon. Ph.D. Thesis. Lisbon, Portugal (1997)
8. E. Cordeiro: A Influência do Comportamento das Pessoas e suas Limitaes na Evacuao dos Edifcios (in Portuguese). . LNEC, Lisbon. Ph.D. Thesis. Lisbon, Portugal (2009)
9. G. Santos and B.E. Aguirre: A Critical Review of Emergency Evacuation Simulation Models. NIST Workshop on Building Occupant Movement during Fire Emergencies. National Institute of Standards and Technology. Gaithersburg, Maryland. 25–50 (2004)
10. D.J. Kaup and T.L. Clarke and L. Malone and R. Olosn: Crowd Dynamics Simulation Research. Proceedings of the Summer Computer Simulation Conference. Society for Modeling and Simulation International. Calgary, Canada. 365-370 (2006)
11. A. Coelho: Incndios em Edifcios (in Portuguese). Edies Orion. Lisbon, Portugal (2010)
12. H. Klpfel and T. Meyer-Koning and J. Wahle: Microscopic Simulation of Evacuation Processes on Passenger Ships. Proceedings of the Fourth International Conference on Cellular Automata for Research and Industry: Theoretical and Practical Issues on Cellular Automata. Springer-Verlag. London, UK. 63–71 (2000)
13. A. Filippopolitis and E. Gelenbe and D. Gianni and L. Hey and G. Loukas and S. Timotheou: Distributed Agent-based Building Evacuation Simulator. Proceedings of the Summer Computer Simulation Conference. The Society for modeling and Simulation International. Edinburgh, Scotland. 46–53 (2008)
14. S. Su and S.T. Tsai and C. L. Shih and R.J. Kuo and H.C. Wang and J.C. Chen: Use of Computer Simulation in the Evacuation System for Hospitals. Proceedings of the Summer Computer Simulation Conference. The Society for modeling and Simulation International. Edinburgh, Scotland. 205–212 (2008)
15. A. Treuille and S. Cooper and Z. Popovi: Continuum crowds. ACM Transactions on Graphic. 3, 1160–1168 (2006)
16. A. Gutierrez-Milla and F. Borges and R. Suppi and E. Luque: Crowd Turbulence with ABM and verlet integration on GPU cards. Proceedings of the International Conference on Computational Science. ELSEVIER. San Diego, California. 80, 2428–2432 (2016)
17. T. Kretz and C. Bnisch and P. Vortisch: Comparison of various methods for the calculation of the distance potential field. Pedestrian and evacuation dynamics 2008. Springer. 225–246 (2010)
18. H. Kobayashi and Y. Ishimoto and M. Fujioka and K. Ishibashi: A multi-agent evacuation simulator to design safe cities for high quality of life with computer clustering. SICE Annual Conference. IEEE Control Systems Society. Kagawa, Japan (2007)
19. A. Gutierrez-Milla and F. Borges and R. Suppi and E. Luque: Crowd evacuations SaaS: an ABM approach. Proceedings of International Conference on Computational Science. ELSEVIER. Reykjavk, Iceland. 51, 473–482 (2015)
20. D. Helbing and I. Farkas and T. Vicsek: Simulating dynamical features of escape panic. Nature. 6803, 487–490 (2000)

Procesamiento Distribuido y Paralelo de Bajo Costo Basado en Cloud&Movil

David Petrocelli¹, Armando De Giusti^{2,3}, Marcelo Naiouf²

¹ Accesibilidad de la información en los sitios web Argentinos (AISWA)
Universidad Nacional de Luján – UNLu

Ruta 5 y 7 – Luján – Buenos Aires – Argentina

² Instituto de Investigación en Informática LIDI (III-LIDI), Facultad de Informática,
Universidad Nacional de La Plata – CIC-PBA
50 y 120 – La Plata – Buenos Aires - Argentina

³ CONICET – Consejo Nacional de Investigaciones Científicas y Técnicas
dpetrocelli@unlu.edu.ar, {degiusti,mnaiouf}@lidi.info.unlu.edu.ar

Abstract. Actualmente el procesamiento intensivo se realiza a través de estructuras HPC híbridas (Grid, Cluster, Cloud) utilizando procesadores de arquitectura x86 y GPUs Nvidia, AMD o Intel, incurriendo en altísimos costos económicos y energéticos. Sin embargo, gracias a la evolución constante del hardware y con el advenimiento de los dispositivos móviles/microcomputadores con CPUs/GPUs ARM acompañado de la masividad de los mismos es posible pensar en una solución de bajo costo y consumo energético para solventar este tipo de problemas. Estos dispositivos incrementan su capacidad, eficiencia, estabilidad y potencia a diario, mientras ganan mercado, conservando un bajo costo, tamaño y consumo energético. A su vez, presentan lapsos de ociosidad, lo que representa una gran capacidad de recursos desaprovechados. Por tal motivo, el objetivo de este trabajo es presentar un prototipo de arquitectura distribuida dinámica, escalable y redundante geográficamente para explotar esta disponibilidad y realizar procesamiento intensivo aprovechando recursos y reduciendo costos.

Keywords: Computación en la nube, Computación móvil, Android, ARM, x86, GPU, Sistemas distribuidos, eficiencia energética.

1 Introducción

En el desarrollo inicial de los procesadores x86 (Intel/AMD) y sus correspondientes GPUs (Intel/AMD/Nvidia) las preocupaciones ambientales fueron débiles, y la atención al consumo energético fueron desvalorizadas. Si analizamos el contexto actual, en los centros de datos y procesamiento de cualquier organización el consumo eléctrico se está convirtiendo cada vez más en uno de los recursos más afectados y que contribuye al costo total de funcionamiento de una sala de procesamiento. Inclusive, dependiendo de la región y del uso de los recursos, el costo total de la energía puede exceder el costo de la adquisición de infraestructura de hardware. Entonces hoy, la energía y la eficiencia son ya una preocupación primordial para el diseño de cualquier sistema informático y se reconoce unánimemente que para alcanzar los retos

computacionales no sólo implica pensar en los objetivos matemáticos, sino también tener en cuenta el consumo energético, el desgaste al ambiente y los costos asociados.

Por tal motivo, desde otra rama de la computación, se desarrolló una línea de procesadores y chips donde por el contrario, el consumo energético es un factor clave del diseño y construcción de estos productos ya que los mismos están orientados a un mercado de dispositivos móviles/microdispositivos los cuales pueden estar alimentados por baterías con energía limitada. Si bien el rendimiento de estos procesadores y sus GPUs integradas es, naturalmente, menor y rezagada en comparación a los procesadores de propósito general, es interesante analizar su potencial y explotar sus capacidades. Al mismo tiempo, además de la potencia de hardware disponible, existe una fuerte masividad de estos dispositivos con grandes períodos de ociosidad mientras se encuentran sin tareas asignadas o recargando su batería. Este tiempo en desuso, en el pasado, fue aprovechado por arquitecturas colaborativas basadas en arquitecturas x86, por ejemplo, en proyectos de SETI@Home [1] para utilizar esa potencia de cálculo y construir un gran centro de datos para analizar señales, entre otras, que requerían gran poder computacional. En la actualidad, y siguiendo el uso de arquitecturas móviles, algunos autores [2,3,4] llevaron, aprovechando características similares, este modelo colaborativo a los dispositivos móviles.

Habiendo analizado la metodología, estructuras y resultados obtenidos de estos trabajos, el objetivo de ésta investigación es presentar ajustes y mejoras respecto a los estudios anteriores.

En primer lugar, la arquitectura diseñada se basa en un prototipo puramente distribuido, flexible y tolerante a fallos para dar soporte a las necesidades de los usuarios a través de un modelo para la gestión de la entrada y salida de dispositivos así como también de las tareas de división, ejecución de trabajos por nodo y recomposición basado en un modelo de recursos en la nube que se ajusta automáticamente en tamaño y capacidad dependiendo de la carga de trabajo, apartado no estudiado anteriormente. En segundo lugar, y siendo una de las motivaciones principales de este trabajo, el escenario de prueba elegido ataca directamente uno de los puntos que se considera de más vulnerabilidad para los procesadores ARM: las tareas de compresión y transcodificación de video, aspecto que fue solo mencionado en investigaciones anteriores. Además, se realiza un análisis comparativo de tiempo de procesamiento y consumo eléctrico implicado entre arquitecturas x86 y ARM.

El resto de este artículo se organiza de la siguiente manera. La sección 2 ofrece una visión general del estado del arte de los procesadores ARM y de los dispositivos móviles. La sección 3, presenta un análisis de factibilidad que justifica la realización de pruebas de compresión de video sobre arquitecturas ARM. Derivado del apartado anterior, la sección 4 define la arquitectura del modelo distribuido planteado. En el capítulo final se presentan las conclusiones y los trabajos futuros.

2 Estado del arte: ARM y los dispositivos móviles

La arquitectura de procesadores más utilizada en términos de dispositivos móviles corresponde a la empresa ARM (por sus siglas en inglés “Advanced RISC Machine”). Sus chips se basan en la arquitectura RISC (por sus siglas en inglés “Reduced

Instruction Set Computing”), la cual se compone por un conjunto de instrucciones reducidas. Una arquitectura RISC típica consta de un único archivo de registro uniforme, arquitectura de carga y almacenamiento y dispone de un modo de direccionamiento simple e instrucciones de longitud fija. Debido a estas características, la arquitectura consigue un alto rendimiento respetando un bajo consumo de energía [5] y un tamaño reducido de chip [6]. Por el contrario, los procesadores de arquitectura x86 tienen el problema de que poseen TDPs (por sus siglas en inglés “Thermal Design Power”) muy por encima de los que se manejan en los desarrollos de fabricantes basados en ARM como Qualcomm, Samsung o MediaTek. Según informes Qualcomm, el TDP para un smartphone está por debajo de los 2,5 W, mientras que para una tablet esa cifra debe ser inferior a los 5W [7]. Recién éstos últimos años, Intel ha podido crear procesadores de bajo consumo con la gama de los Intel Core M [8], con un TDP promedio de 4,5 W pero aún a rendimientos menores [9] o con algunos Intel Atom de bajo consumo, rondando en un TDP de 5 W [10].

Tomando el último producto lanzado por Samsung en 2017, el Samsung Galaxy 8, puede destacarse que se equipa con un procesador ARM de 64 bits, el Snapdragon 835, el cual ha alcanzado 8 núcleos a una frecuencia de 2.45 GHz cada uno con un tamaño de 10 nm, su GPU, la Adreno 540, da soporte a cálculos de punto flotante (en simple y doble precisión) y dispone de 256 cores con un promedio de 645 Mhz por cada uno de ellos, con soporte a OpenCL 3 [11,12] a una tasa muy baja de consumo energético (TDP < 3W) [13].

Por otra parte, además del aumento de la capacidad de hardware, los dispositivos móviles cuentan con otro aspecto interesante que diversas instituciones han investigado: la gran masividad y penetración en el mercado. Según la Unión Internacional de Telecomunicaciones (ITU), desde finales de 2016 existen aproximadamente 7.300 millones de usuarios móviles [15]. A su vez, en el mismo artículo, predicen que el número de dispositivos móviles per cápita se espera que llegue a 1,5 o 2 en 2020 [14]. Por su parte, Adam Lella [16] anunció estadísticas que se aproximan al 80% de adopción de teléfonos inteligentes en las personas a lo largo del mundo desde la misma fecha. Otras fuentes como Wikipedia [17] y WeAreSocial [18] estiman una penetración de estos dispositivos entre el 60 y 70 % a lo largo del planeta. Según IDC [19] se estima para 2017, se establecerán 1534 millones de ventas y entregas de dispositivos móviles. A través de los reportes de TechMalak y Statista [20], se estima que habrá un total de 2.3 billones de dispositivos inteligentes para finales de 2017

3 Análisis de Factibilidad: Compresión de video sobre arquitecturas ARM

Habiendo realizado un análisis del estado actual de los dispositivos móviles, la capacidad y evolución del hardware de los procesadores ARM, en esta sección se presenta como este trabajo contribuye a manejar los recursos disponibles para realizar tareas de compresión de video, teniendo en cuenta un bajo costo y consumo energético. Por lo tanto, se examina (a) cantidad y calidad de videos fuentes acordes para realizar la tarea de compresión, tamaño de fragmentos de video y parámetros básicos para la

compresión del video y (b) las exigencias individuales para realizar dicha tarea en segmentos acotados de las fuentes comparando dispositivos ARM y x86.

(a) Para generar una prueba que produzca una saturación en las arquitecturas de los procesadores, se analizaron trabajos y configuraciones previas [21, 22]. Por tal motivo, para este trabajo, se recopilaron 4 fuentes de video con propiedades características de ambientes de pruebas [23,24,25] las cuales contienen una amplia gama de características para realizar las pruebas de factibilidad. El resumen de las propiedades más importantes de cada video se detalla en la figura a continuación:

Propiedades de Videos elegidos para realizar las pruebas									
Nombre	Tamaño	Proporciones	Contenedor	Codec	Perfil	Nivel de Compresión	Tasa de bits	FPS	BPP
512.mp4	36 MB	416x240	MP4	AVC x264	BaseLine	@L1.3	512 kbps	24	0.21
2500.avi	210 MB	854x480	AVI	Lavc	Simple	@L1	2500 kbps	24	0.25
4000.mp4	339 MB	1920x1080	MP4	AVC x264	High	@L4.2	4000 kbps	60	0.03
8000.mp4	642 MB	3840x2160	MP4	AVC x264	High	@L5.1	8000 kbps	60	0.01

Fig. 1. Características principales de videos utilizados para las pruebas de compresión

Por su parte, se trabaja con 2 tamaños de fragmentos, medido en segundos, para el procesamiento en un nodo trabajador: 3 y 10 segundos, valores que se encuentran dentro del tamaño recomendado para el servicio de streaming de video [26, 27]. Respecto del proceso de transcodificación y re-empaquetamiento a realizar, se describen dos tareas a continuación. En ambos casos se empaqueta a archivos MP4 y se re-comprime con el códec AVC H.264, siendo x264 la librería de código abierto utilizada por la herramienta FFMpeg, siguiendo para la misma las referencias recomendadas [28].

El resumen de las propiedades de los perfiles de compresión se presenta a continuación:

Tabla de Perfiles de compresión de Video											
Perfil	Códec Video	Int. KeyFrames	Kbps Video	Perfil de Comp.	Niv. de Comp.	FPS	Preset	Códec Audio	Kbps Audio	Muestreo	Canales
Perfil 1	libx264	50	220	baseline	@L3.0	15	Rápido	aac	64	44100	2
Perfil 2	libx264	250	8000	high	@L5.1	30	Lento	aac	256	48000	6

Fig. 2. Características principales de los videos utilizados para las pruebas de compresión

(b) Para evaluar las exigencias individuales implicadas en las tareas de compresión se utilizó un conjunto de recursos x86 y ARM. La selección de dicho equipamiento busca presentar escenarios que permitan contrastar rendimiento y consumo energético, con el objetivo de justificar el modelo de este trabajo. Los dispositivos y características más relevantes se detallan a continuación:

Tabla de recursos utilizados para experimentación					
Nombre dispositivo	Tipo Procesador	Procesador	Características	TDP Max	Memoria
Samsung Galaxy J7 (2017)	ARM	Exynos 7870 (Cortex-A53)	8x1.6 Ghz	2.5 W	3 GB
Samsung Galaxy S7	ARM	Exynos 8890 (Cortex-A53)	4x2.3 GHz + 4x1.6 GHz	2.5 W	4 GB
BGH Positivo SchoolMate 15 TV	x86	Intel Celeron N2806	2x2 Ghz	4.5 W	4 GB
BGH Positivo A-497	x86	Intel i7 2630QM	4x2.00 GHz	45 W	4 GB
Pc de Escritorio Genérica	x86	Intel i7 6770k	8x4.0 Ghz	91 W	16 GB

Fig. 3. Características principales de videos utilizados para las pruebas de compresión

Con objeto de realizar las pruebas se instala FFMpeg (versión 3.3.2) en cada uno de los recursos ARM (Android) y x86 (Windows) y se toma el tiempo en milisegundos para cada tarea. Luego de ejecutar los procesos definidos en el punto (a) sobre los fragmentos de 3 y 10 de cada fuente de video, la tabla resultante es la siguiente:

Tabla de tiempos (en milisegundos) de las tareas de compresión especificadas						
Video	Perfil	Intel Celeron N2806 (4,5 W)	Intel i7 2630QM (45w)	Intel i7 6700k (91 w)	Samsung Galaxy J7 (2,5 W)	Samsung Galaxy S7 (2,5 W)
512_3segs	Perfil Base	1030	270	149	1426	1102
512_10segs	Perfil Base	4664	714	336	2521	1740
512_3segs	Perfil High	3771	739	347	3192	2388
512_10segs	Perfil High	14927	2747	1293	6885	3852
2500_3segs	Perfil Base	3854	660	330	4451	2273
2500_10segs	Perfil Base	8314	1530	733	10887	4398
2500_3segs	Perfil High	13861	2216	1145	7954	3826
2500_10segs	Perfil High	56180	8990	3562	20219	10729
4000_3segs	Perfil Base	8653	1330	544	5203	2854
4000_10segs	Perfil Base	26577	4494	1448	14164	7400
4000_3segs	Perfil High	15869	2990	1115	10970	5440
4000_10segs	Perfil High	142327	21059	7930	49634	27170
8000_3segs	Perfil Base	28992	11515	1673	9429	6644
8000_10segs	Perfil Base	104252	15630	4799	31384	18892
8000_3segs	Perfil High	48117	17120	3260	32988	22287
8000_10segs	Perfil High	424868	105970	21802	242997	141483

Fig. 4. Características principales de videos utilizados para las pruebas de compresión

A partir de dicha experimentación, la evaluación (b) es la siguiente:

- Los procesadores Intel i7 de arquitectura x86 presentan los mejores resultados de compresión, pero incurriendo en muy alto consumo energético, tomando como base la información de sus TDPs.

- Tomando el i7 6700k y el Samsung Galaxy S7, como mejores candidatos de cada arquitectura, para las diferentes pruebas se deduce que:

-- El i7 realiza el proceso de compresión, en promedio, 5 veces más rápido que el chip ARM. Sin embargo, el consumo del es 36 veces mayor.

-- Las tareas de compresión sobre fragmentos de 10 segundos presentan tiempos más cortos respecto de los de 3 segundos (4,5 contra 5,7 veces más lento) en proporción.

-- Las tareas de compresión sobre el perfil High muestra una relación más positiva respecto de la Baseline (4,73 contra 5,46 veces más lento)

- Si se compara el Intel Celeron N2806 (x86) y Exynos 7870 (ARM), ambos de poco consumo energético, el procesador ARM realiza la tarea asociada 3 veces más rápido.

4 Arquitectura distribuida para HPC sobre ARM

4.1 Introducción

A partir de los resultados obtenidos se desprende y resulta factible la propuesta de este trabajo: La construcción de una arquitectura distribuida sobre dispositivos ARM para realizar tareas de cálculo intensivo utilizando la capacidad de cálculo disponible, el tiempo ocioso y la masividad de los mismos.

Cómo punto de partida, es importante destacar que la arquitectura aquí planteada se basa en un modelo híbrido compuesto por (a) recursos Cloud, y (b) por una red P2P de dispositivos móviles.

Los recursos Cloud (a) se utilizan para construir el sistema de administración, gestión, clasificación y control de dispositivos P2P, siendo además el punto redundante de acceso al sistema, gestión y balanceo de carga, así como también fragmentación y recomposición de segmentos de recursos. Por su parte, la red P2P (b) se utiliza, como cliente, para realizar las solicitudes de trabajo de procesamiento intensivo y para procesar de manera distribuida las solicitudes cuando el nodo actúa como servidor.

Para continuar con las pruebas de factibilidad realizadas, el modelo plantea disponer en cada nodo la herramienta FFMpeg.

4.2 Arquitectura del prototipo distribuido

El prototipo aquí presentado se compone de tres grandes partes: (a) Punto de Acceso al sistema; (b) Nodos de Administración y (c) Nodos Móviles

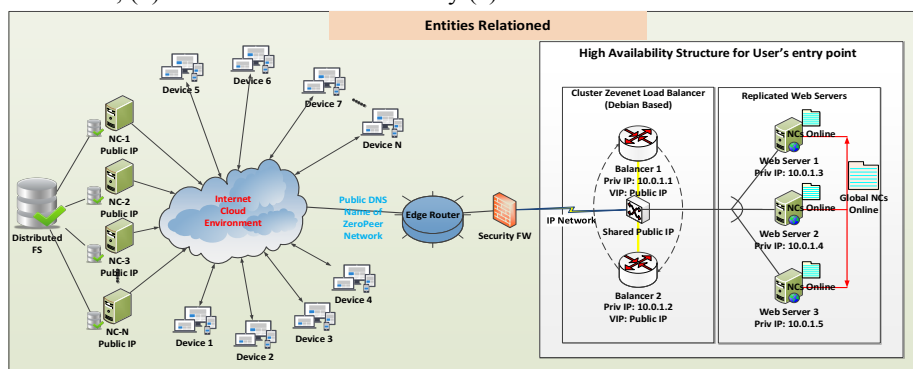


Fig. 5. Entidades básicas que se relacionan en el sistema

- Punto de Acceso al sistema (Nodos Cloud): La estructura de acceso al sistema evita tener puntos únicos de errores y se encarga de mantener replicado en al menos 2 nodos los datos necesarios para mantener el sistema disponible. Para ello, a través de la plataforma Zevenet (gratuita), se construye un balanceador en cluster que se encarga de manejar dos equipos para una única dirección IP virtual o dirección DNS. A su vez, el balanceador reparte la carga entre ambos nodos y mantiene el servicio disponible ante posibles fallas. También, detrás de esta capa de balanceo, se encuentra un conjunto de servidores Web replicados (2 o más) los cuales registran y administran cuáles nodos se encuentran disponibles para atender a los clientes. A su vez, estos servidores Web son los que los usuarios consultan, a través de la dirección DNS, para ingresar al sistema y obtener el listado actualizado de los nodos de administración que se encuentran actualmente activos.

- Nodos de Administración (Nodos Cloud): Los recursos Cloud se organizan para construir una red de administración flexible. Para ello, definen un conjunto de características:

En primer lugar, tienen la función de administrar su propia estructura flexible. Esto significa que la red de administración conoce el estado cada nodo y de la red en general, de manera de poder decidir si es necesario aumentar o reducir la cantidad de nodos que actualmente están atendiendo a los clientes, siempre respetando un mínimo de 3 nodos para garantizar tolerancia a fallos. En este primer modelo se plantea tener en cuenta, a nivel de conjunto, el análisis de las siguientes métricas: (a) Porcentaje de uso del procesador y memoria. (b) Cantidad de nodos clientes atendidos. (c) Cantidad de peticiones encoladas. En caso de disponer de valores de (a) superiores al del 80 %, de (b) por encima de 15 (por nodo) o (c) mayores a 10, se gestiona la creación de un nuevo nodo de administración. De igual manera, si la cantidad de (a) es menor al 40%, y la cantidad de nodos es superior al mínimo necesario (3), se apaga un nodo. Si algún recurso en la nube está apagado por más de una semana, se elimina con objeto de no incurrir en costos de reserva de almacenamiento.

En segundo lugar, estos nodos son el punto de acceso al desarrollo de actividades del sistema, motivo por el cual todos los dispositivos móviles, luego de haber obtenido la lista de los recursos Cloud, se conectan para solicitar tareas de cómputo intensivo.

En tercer lugar, relacionado con el punto anterior, la red de administración se encarga de (a) la gestión y (b) la clasificación de los dispositivos móviles que se conectan a la red. Respeto del punto (a), su tarea es administrar la entrada y salida de los recursos con objeto de mantener una lista de recursos de procesamiento disponible; relacionado con el apartado (b), se encarga de clasificar los dispositivos que se conectan según métricas estáticas y métricas dinámicas. Las primeras se componen de: Nombre del dispositivo, Sistema Operativo (familia y versión), información de hardware (CPU, memoria total, GPU) y posición geográfica. Con esas métricas se realiza una clasificación inicial en uno de tres grupos: 1. Gran capacidad de Hardware; 2. Capacidad promedio de Hardware; 3. Baja capacidad de hardware. Por su parte, los controles dinámicos se componen por un ajuste de esa evaluación inicial, donde una tarea se ejecuta cada 10 segundos y permite reclasificar el orden que obtienen dentro de cada grupo según varíen dinámicamente a lo largo del tiempo, siendo los mismos los siguientes: % de uso de CPU, memoria, GPU y ancho de banda de bajada y subida disponible.

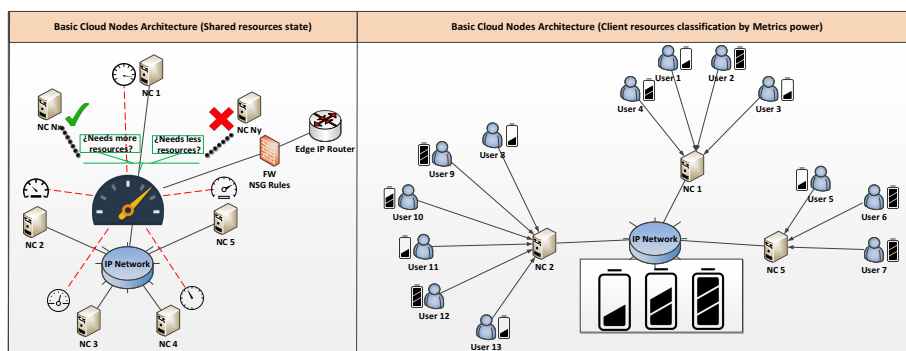


Fig. 6. A la izquierda representación de administración de recursos Cloud en base a carga del sistema; a la derecha ejemplificación de clasificación de dispositivos.

En cuarto lugar, a partir de las peticiones de los usuarios finales, el nodo de administración que lo atiende se encarga de gestionar la solicitud y derivar los fragmentos de tareas a procesar. Cuando la tarea se completa la devuelve al usuario final. Siguiendo la experimentación realizada en el análisis de factibilidad, cada nodo debe disponer la herramienta FFMpeg para realizar las tareas asociadas. Es importante destacar que el nodo Cloud dispone de la capacidad para gestionar los errores que pueden originarse en los nodos trabajadores. Para ello, mantiene para cada tarea una lista con la siguiente información {IDUsuario; IDTarea; IDSubtarea; EstadoTarea}, donde los estados pueden ser: {en cola, procesando, terminada, #errores}. Si una tarea por algún problema con el nodo trabajador llega corrupta, el dispositivo informa un error, o se vence el tiempo de espera, se marca el error y se valida si la cantidad de intentos fallidos es menor a 3. Si es mayor, se aborta la tarea, si es menor pasa a ser encolada nuevamente y se envía a un nuevo usuario para ser procesada.

- Nodos Móviles (Nodos P2P): Los dispositivos móviles se organizan como una red P2P para aportar su poder de cálculo en los tiempos que se encuentran ociosos. Para que un dispositivo sea considerado disponible deben darse un conjunto de condiciones: (a) se debe encontrar conectado a la red eléctrica con un tiempo mayor a 10 minutos y debe disponer de una carga de batería superior al 60%; (b) el equipo se debe encontrar con un uso de recursos menor al 10 % y (c) debe estar conectado a una red WiFi. Estos tres puntos buscan garantizar que, por un lado, no se afecte el uso de los dispositivos mientras los usuarios lo están utilizando, que no le produzca gasto al usuario y por otro, que estén conectados a la red desde un tiempo suficiente con objeto de no incluir demasiados equipos que entren y salgan constantemente a la red, evitando demasiadas tareas de gestión y problemas de reasignación de trabajos.

Como estos dispositivos construyen una red P2P, pueden cumplir dos roles: (a) Clientes o (b) Servidores de la red. Siguiendo la experimentación realizada, en ambos casos se debe incluir la herramienta FFMpeg para realizar las tareas asociadas.

En el caso que un nodo actúe como cliente, una vez conectado a la red de administradores solicita una tarea, en este caso de compresión, para la cual define un conjunto de perfiles y parámetros de compresión que desea obtener. Para cada perfil describe: códec de video y audio, cantidad de bits de audio y video, perfil de compresión, dimensión de pantalla, aspecto de pantalla, cuadros por segundo, entre otros.

Finalmente, dependiendo de las capacidades del dispositivo móvil emisor pueden darse dos casos: El primero es que disponga de la suficiente potencia para fragmentar el video e ir enviando a la red de administración pequeños segmentos de tamaños variable, siempre entre 2 y 10 segundos, para realizar la tarea de compresión. A su vez esto implica que en la medida que la red comprima los fragmentos, los reciba el nodo de administración, éste último debe devolverlos directamente al cliente final para que lo reconstruya. El segundo caso, es que sea un dispositivo de baja potencia y obligatoriamente deba subir el video completo a la red para realizar el proceso. En este caso deriva no solo la tarea de compresión al sistema sino que también los procesos de fragmentación y unificación a la red de administración.

En el caso que el nodo actúe como servidor, es tarea del mismo llevar a cabo los procesos de cómputo intensivo a partir de los datos y parámetros que se reciben por parte de la red de administración.

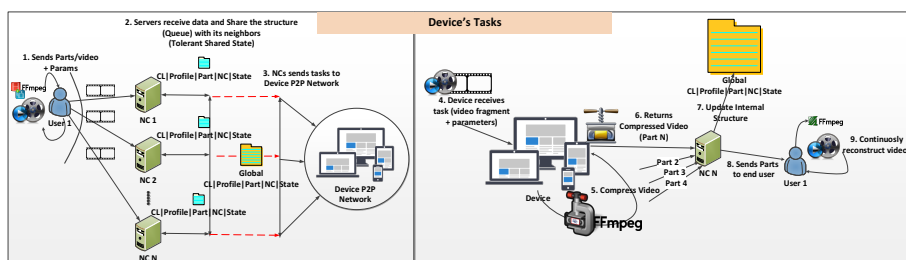


Fig. 7. A la izquierda representación del funcionamiento de un dispositivo móvil como cliente, a la derecha como servidor.

En este caso, a partir de los fragmentos de video y de los parámetros especificados, utilizará los mismos para ejecutar la tarea con la herramienta FFMpeg y se realice así el proceso de compresión. Una vez finalizada la tarea, se devuelven los resultados a la

red de administración. En caso que no pueda realizar la tarea, falle el proceso de compresión, o se desconecte de la red, debe avisar al nodo de gestión para que éste reasigne la tarea.

5 Conclusiones

Al inicio de este artículo, se investigaron y analizaron las características, evolución y masividad de los recursos ARM y se comprobó la factibilidad de realizar procesamiento intensivo y cómo es que estos procesadores presentan ahorros significativos respecto del consumo energético para dichas tareas. Finalmente, se propuso un modelo de arquitectura distribuida con dispositivos móviles para realizar tareas de cálculo intensivo de manera flexible y escalable explotando la capacidad de hardware ya disponible y ociosa. La propuesta se basa en varias observaciones, entre ellas las siguientes: a) la cantidad masiva de estos dispositivos; b) La ociosidad de estos recursos; y c) Los beneficios significativos que dicha infraestructura podría generar en términos de costos y procesamiento para diversos sectores de la comunidad. Finalmente, la misma se considera lo suficientemente robusta para afrontar diversas evaluaciones y afrontar los retos de este tipo de aplicaciones.

6 Trabajos Futuros

A partir de los aspectos investigados en este trabajo, se desencadenan un conjunto de actividades a realizar en etapas posteriores y que complementan el trabajo aquí desarrollado:

- Construir el prototipo del sistema, con el objetivo de poder evaluar la estructura de manera de obtener resultados concretos.
- Incluir dispositivos de diversas características y evaluar cómo responde la escalabilidad del sistema en la medida que existe mayor cantidad de recursos y cantidad de tareas a procesar.
- Incluir las capacidades de Hardware Accelerator de FFMpeg para evaluar rendimientos no solo con CPU sino también con las capacidades GPUs con OpenCL.
- Implementar otros servicio de algoritmia paralela que permitan realizar cálculos sobre las tarjetas gráficas. En ambos casos comprobar los resultados obtenidos entre arquitecturas x86 y ARM.

7 Referencias

1. David P. Anderson.: BOINC: A System for Public-Resource Computing and Storage. GRID '04 Proceedings of the 5th IEEE/ACM International Workshop on Grid Computing. Pages 4-10, (2004)
2. Foster, I., Kesselman, C.: Computing while charging: building a distributed computing infrastructure using smartphones. CoNEXT '12 Proceedings of the 8th international conference on Emerging networking experiments and technologies pages 193-204, Nice, France. (2012)
3. Sriraman K R, Lead Architect.: Grid Computing on Mobile Devices - A Point of View. IEEE Xplore Conference: Grid Computing. Proceedings. Fifth IEEE/ACM International Workshop on (2004)

4. A New Purpose for Old Smartphones: Cluster Computing, <https://www.hpcwire.com/2015/01/27/new-purpose-old-smartphones-cluster-computing/>
5. ARM architecture, <http://infocenter.arm.com/help/index.jsp?topic=/com.arm.doc.subset.architecture.reference/index.html>
6. ARM's Cortex M: Even Smaller and Lower Power CPU Cores, <http://www.anandtech.com/show/8400/arms-cortex-m-even-smaller-and-lower-power-cpu-cores>
7. Qualcomm aims at 2.5 to 3W TDP for phones, <http://fudzilla.com/31532-qualcomm-aims-at-25-to-3w-tdp-for-phones>
8. Core M: paso atrás en potencia, paso adelante para el formato híbrido, <https://www.xataka.com/componentes/core-m-paso-atras-en-potencia-paso-adelante-para-el-formato-hibrido>
9. Versus: Intel Core M 5y10 vs Qualcomm Snapdragon 835, <https://versus.com/en/intel-core-m-5y10-vs-qualcomm-snapdragon-835>
10. Intel Atom z8000 Processor Series, <https://www.intel.com/content/dam/www/public/us/en/documents/datasheets/atom-z8000-datasheet-vol-1.pdf>
11. Qualcomm Details Snapdragon 835: Kryo 280 CPU, Adreno 540 GPU, X16 LTE, <http://www.anandtech.com/show/10948/qualcomm-snapdragon-835-kryo-280-adreno-540>
12. Samsung Galaxy S8 Review, <https://www.digitaltrends.com/cell-phone-reviews/samsung-galaxy-s8-review/>
13. Aaron Carroll, Gernot Heiser.: An analysis of power consumption in a smartphone USENIXATC'10 Proceedings of the USENIX conference on USENIX annual technical conference. Pages 21-21. Boston, MA — June 23 – 25. (2010)
14. Visual Networking Index, <http://www.cisco.com/c/en/us/solutions/service-provider/visual-networking-index-vni/vni-infographic.html>
15. Statistics 2017. ITU. <http://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx>
16. U.S. Smartphone Penetration Surpassed 80 Percent in 2016, <https://www.comscore.com/Insights/Blog/US-Smartphone-Penetration-Surpassed-80-Percent-in-2016>
17. List of countries by smartphone penetration, https://en.wikipedia.org/wiki/List_of_countries_by_smartphone_penetration
18. DIGITAL IN 2017: GLOBAL OVERVIEW, <https://wearesocial.com/special-reports/digital-in-2017-global-overview>
19. Smartphone Volumes Expected to Rebound in 2017, <http://www.idc.com/getdoc.jsp?containerId=prUS42334717>
20. Number of smartphone users worldwide from 2014 to 2020 (in billions), <https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/>
21. Encoding. Lee, Byoung-Dai.: Empirical Analysis of Video Partitioning Methods for. International Journal of Multimedia and Ubiquitous Engineering. 10. 81-90. (2015).
22. Adriana Garcia, Hari Kalva, Borko Furht.: A study of transcoding on cloud environments for video content delivery. MCMC '10 Proceedings of the 2010 ACM multimedia workshop on Mobile cloud media computing Pages 13-18. Firenze, Italy — October 29 – 29. (2010)
23. Xiph.org Video Test Media [derf's collection], <https://media.xiph.org/video/derf/>
24. ITEC – Dynamic Adaptive Streaming over HTTP – Dataset, http://www-itec.uni-klu.ac.at/dash/?page_id=207
25. Download Guide: Big Buck Bunny, Sunflower versión, <http://bbb3d.renderfarming.net/download.html>
26. Optimal Adaptive Streaming Formats MPEG-DASH & HLS Segment Length. <https://bitmovin.com/mpeg-dash-hls-segment-length/>
27. Choosing the Optimal Segment Duration, <https://streaminglearningcenter.com/2016/08/04/choosing-the-optimal-segment-duration/>
28. Linux Encoding - x264 FFmpeg Options Guide, <https://sites.google.com/site/linuxencoding/x264-ffmpeg-mapping>

Using an Improved Data Structure in Hybrid Memory for Agent-Based Simulation

Silvana Gallo^{1,4}, Francisco Borges², Laura De Giusti¹, Marcelo Naiouf¹, and Remo Suppi³

¹ Instituto de Investigación en Informática III-LIDI, Facultad de Informática, Universidad Nacional de La Plata, La Plata, Buenos Aires, Argentina

² Federal Institute of Bahia, Santo Amaro, Bahia, Brazil

³ Universitat Autònoma de Barcelona, Bellaterra, Barcelona, España

⁴ CONICET, Facultad de Informática, UNLP, La Plata, Buenos Aires, Argentina
{sgallo,ldgiusti,mnaiouf}@lidi.info.unlp.edu.ar
franciscoborges@ifba.edu.br,remo.suppi@uab.cat

Abstract. Data structure is an important issue to get good performance in parallel and distributed applications. These data structures have to be designed with the memory paradigm in mind where the data structure will be used in order to explore the architecture in a better way and subsequently obtain the best Speedup. Current parallel programming languages enable us to easily transform a parallel solution developed for a distributed paradigm to a hybrid solution just by adding pragma codes. At first approach, this is an interesting solution because it does not require several code modifications. Nevertheless, this interchange can cause a slowdown if an appropriate and deep adaptation is not carried out in the code. In this paper, we present our experience when we migrated a data structure developed for a distributed paradigm to a hybrid paradigm. This data structure was implemented in our Fish Schooling Agent-Based simulator where it might be useful either as a distributed paradigm or a hybrid paradigm. The results show the importance of customizing the data structure for the appropriate infrastructure and parallel programming paradigm. We believe that the data structure should have a flexible and dynamic behavior in accordance with the paradigm used.

Keywords: Parallel and distributed simulation, High performance simulation, Hybrid MPI+OpenMP programming, Individual-oriented model

1 Introduction

Parallel and distributed architectures have been providing more power computing for the scientific community. They enable researchers to develop more complex and realistic models. However, the computing resources available can be underused if the software does not take advantage of these architectures. Data structures are an important issue that must be considered in order to

obtain maximum speedup in the parallel and distributed solutions. Data structures should fit the modeled problem on the parallel and distributed architecture and with the programming paradigm. Otherwise, the solution will not reach an acceptable level of performance.

In fact, the parallel and distributed architecture and the programming paradigm are strongly related. The concepts of the memory paradigm: Distributed Memory (DM) and Shared Memory (SM) are the key to understand how to explore the architecture in a better way. DM paradigm uses an addressing memory for each process. In this paradigm, each process has only access to the variables that are inside of its addressing memory. Hence, there are no parallel and concurrent writing in the same variables and data structures. On the other hand, the SM paradigm shares the same addressing memory space among threads, where each thread can concurrently access the same variables and data structures. This paradigm requires that lock mechanisms be applied in order to maintain data coherence. Generally, lock mechanism implies a sequential access to variable and data structures and can become a bottleneck in a parallel execution.

DM is a very appropriate solution for scalability [5] because it enables us to easily add other processing nodes in order to execute a solution. The communication among the processes occurs by message passing. Communication among the processes is the main cause of the inefficiency of DM solutions. The SM allows the solution to take advantage of the locality of the data using the cache hierarchy. In addition, SM decreases the communication volume in the network because the communication occurs by memory copy. Therefore, the Hybrid Memory (HM) paradigm emerges as an approach that can use the main advantages of both DM and SM paradigms. Accordingly, these were the main motivations in using the data structure proposed by [14] in an HM paradigm. The volume communication in MPI is too high; therefore, we would decrease the communication volume among the agents because they are in the same core and an MPI communication routine is not required. In addition, we can take advantage of data locality. The data structure proposed by [14] is a data structure used to partition the data over the parallel and distributed architecture. This data structure is implemented in our Fish Schooling Agent-Based simulator. Our group has been researching High Performance Parallel and Distributed Simulation over the last few years [2, 14]. That is why we were interested in analyzing the data structure behavior in the hybrid memory.

In this paper, we present our experience in data structure migration proposed by [14] to the HM paradigm. We report the solutions and problem in two phases: in the first one, we just add an OpenMP pragma without changes to the data structure [14]. In the second phase, we show the changes and adaptation of data structure and we present, with software point view, the resulting data structure [2]. We have used a Fish Schooling Agent-Based Model simulation [14] in order to test the proposed data structure. Thence, some questions emerge and are answered throughout the paper: could we use a data structure, initially developed for DM, in an HM paradigm? What changes in data structure would

be necessary in order to maintain or improve the previous performance? What are the challenges?

The remainder of the paper is organized as follows: the related works are discussed in Section 2. After that, the hybrid data structure is presented in Section 3. Then, we present the experimentations in Section 4; and, lastly, our conclusions and final considerations are shown in Section 5.

2 Related Work

Data structures drive the computational complexity, and memory is an important issue in HPC. The data structures used depend, directly on the behavior of the model. In our case, we have a fixed amount of nodes and the behavior of the agents inside the node is dynamic, so we use a dynamic list. In previous works, we can find both Distributed memory and Shared memory approaches in order to improve the performance of the solution. In the literature, we have not found an agreement on the use of Hybrid Memory paradigm for parallel and distributed applications. Basically, the works present different applications and results about this point. What we observe is that the gain of performance depends on a combination of several factors, for example, architecture [7, 9, 12], network interconnection [6], and application [4, 7]. Adhianto and Chapman [1] provide an additional list of references on this topic.

In addition, agent based models have been solved by software engineers used widely in software engineering. This is a common problem; therefore, several tools have been created in order to encapsulate the programming complexity. In the case of non-distributed computing, we can find tools such as Netlogo [15], Anylogic [3]. They use dynamic structures such as List of Lists in the case of grid based maps or hash in the case of agent definitions, but all the structures are used without distributing. Some parts of Netlogo [15] are programmed using Scala, which uses Actor programming. Repast HPC [8] is an HPC tool programmed in C and MPI library. Agents communicate their information sharing by using messages. It is programmed using distributed memory with a language description based in environmental agents, networks, coordinating and mobile agents, in a Logo-style. Pandora [13] is an open framework for discrete event simulation and ABM HPC tool that uses shared memory in an OpenMP level and distributed memory in an MPI level. Flame [11] is another framework for agent based modeling. It also has a GPU version that has the same modeling approach but a different computational model and architecture. In the case of the main, Flame uses MPI and it does not use OpenMP, but Pthreads are used for asynchronous communications.

3 Data structure

We organize this section in two parts. First, we review the list of cluster data structure [14], as this is the base for the data structure proposed in [2]. Then, we talk about our experience with using this data structure [14] at hybrid simulator

without appropriate adaptations. In the second part of this section, we present the list of clusters adapted for this solution.

3.1 List of Cluster Data structure

The data structure proposed by [14], see Figure 1, is a dynamic list composed by nodes. Each node has some information and link(s) to other node(s). Each node is called a cluster, therefore, the data structure is called a *list of clusters*.

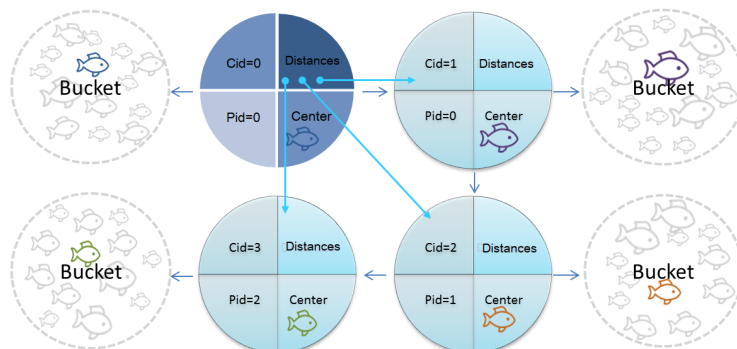


Fig. 1: The figure shows the data structure proposed by [14] where all communication is using MPI.

Each cluster consists of several data such as: a) Center field that represents a centroid, which is the most representative element of the cluster; b) the processor identifier (PID) indicates in which process each cluster is stored; c) a bucket in which agents belonging to the cluster are stored; d) the cluster identifier (CID), which indicates the cluster position in the list; and e) the distances to other clusters (see Figure 1). This data structure allows for defining areas in which agents can interact only with agents belonging to adjacent areas (see Figure 2a), reducing the computing involved in the neighbour selection process.

In the DM paradigm, the clusters are distributed between processors by MPI processes; therefore, each MPI process can have many allocated clusters (see Figure 2a). Some adjacent clusters are allocated in the same processors according to the PID field. Therefore, each MPI process has a copy of its list of clusters. During the simulation, the agent moves in the environment, thus its centroid can change and it begins to suffer from the influence of other centroids. Therefore, a migration process of agents among the MPI processes occurs. In this moment, the MPI processes send and receive these agents and update their local list of clusters.

The motivation to proof a hybrid implementation for a list of cluster data structure was that we had observed that many agents were in same processor; therefore, the OpenMP strategy in this situation would be interesting as a performance improvement. Thus we developed a hybrid version using the list of clusters

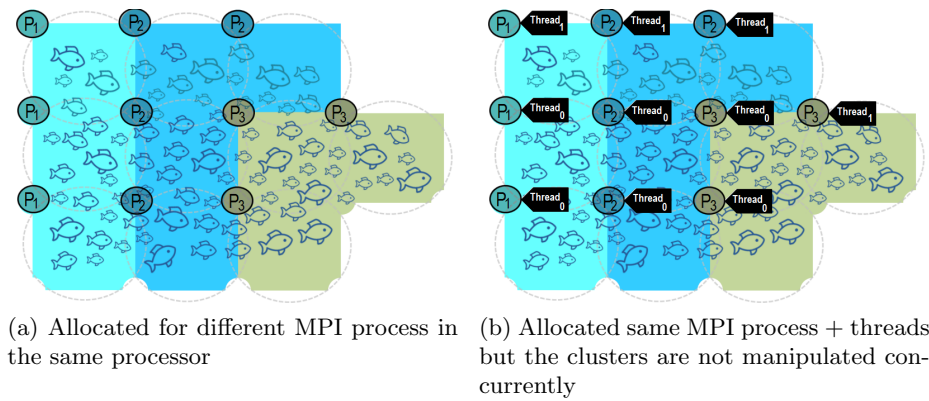


Fig. 2: Cluster adjacent areas.

and adding OpenMP pragmas without changes to the data structure. Figure 3 shows the execution time with list of cluster being used as a Distributed Memory paradigm (only MPI) and a Hybrid Memory paradigm (MPI and OpenMP).

As we can note, this Hybrid solution has a slowdown in comparison with the DM solution. The list of clusters works very well in the pure MPI version because each processor has a local copy of this data structure. Therefore, there is no concurrent manipulation of data structure in each MPI process. On the other hand, this hybrid version has many threads of the same processes inserting and updating concurrently the same copy of data structure through OpenMP threads, see Figure 2b.

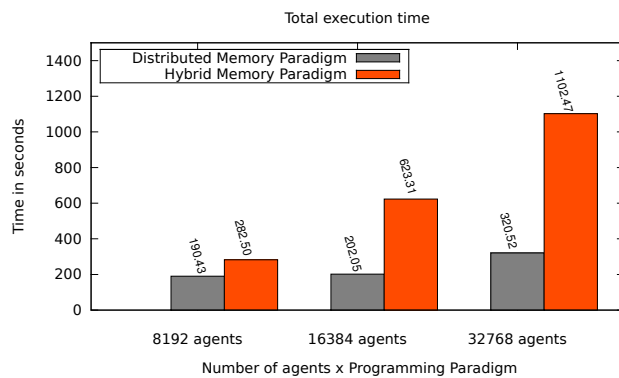


Fig. 3: Total execution time with Distributed Memory and Hybrid Memory paradigm

Therefore, the list of clusters has a performance problem when it is used in a hybrid solution, where many threads can write new information simultaneously in the cluster. This data structure becomes a bottleneck because a sequential access at data structure is required in order to maintain the data coherence. To avoid inconsistency in the simulation results, it is necessary to use OpenMP critical regions. In our DM simulator, the list of cluster is represented by a vector of list of cluster. Each MPI process, manipulates just the agents that are in it. The insert method includes a new agent in the bucket. This has the same behavior when we need to include new clusters inside of the list. The problem with this solution is that many threads execute this operation concurrently through the shared vector of lists. Therefore, the critical section is required. Critical section enables just one thread to have access to this code region at the same time. Thus only one active thread can access the data referenced by the code. As a result, this produces a sequential behavior and possible slowdown of the parallel code.

We can see in Figure 3, as the number of agents increases, the slowdown increases as well. This occurs because the number of updates in the agents is increased; therefore, more threads enter the critical section code. In this hybrid solution, each MPI process receives a set of agents by MPI message. Then, the process through OpenMP loop parallelization distributes the agents to other cores in the same node that execute the computing over the shared data structure. These operations have to be protected inside of the critical section because many threads are inserting and updating the individuals of this data structure at the same time. The slowdown happens because these operations are invoked constantly.

In order to solve the slowdown problem in the hybrid version, we propose an adaptation of the data structure that supports shared memory paradigm. This data structure can be used with either the SM or Hybrid paradigm.

3.2 Hybrid Memory Data structure

Figure 4 shows the data structure proposed. The main idea is to distribute the clusters among the OpenMP threads. In this way, there are no concurrent accesses to the same cluster when insert and update operations occur because each thread is able to manipulate the clusters that were assigned to it (see Figure 2b). Thus, we can take off many critical sections of the code. Since, in this solution, a thread can only access its clusters, there will be no inconsistencies.

The data structure, see Figure 4, is composed by a Metalist where n positions are dynamically created according to the MPI rank size. Suppose there are ten MPI ranks. Then 10 positions will be created. Each position of the Metalist represents an MPI Rank. Therefore, MPI Rank 0 manipulates the clusters of the position 0; MPI Rank 1 manipulates the clusters of the position 1, and so on until the last MPI Rank. Inside of each position, the proposed data structure has a vector of list of cluster. This vector is created dynamically and has the same number of OpenMP threads that were created in the execution time. Therefore, each MPI process distributes the clusters among its threads. Thus each thread will manipulate only the clusters assigned to it.

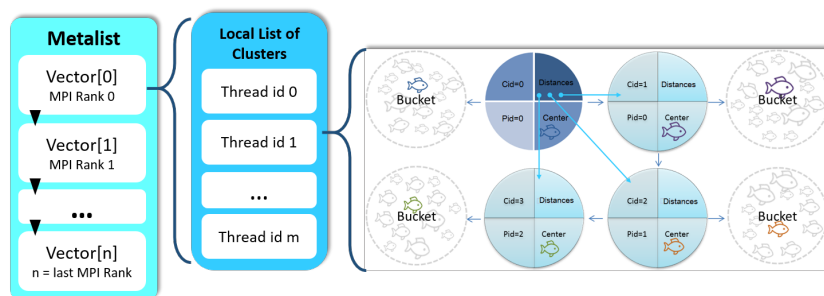


Fig. 4: This figure shows the adaptation of the partitioning data structure to the hybrid simulator version.

The main difference of the proposed data structure is that the pointer to the list of clusters is not shared. This data structure (Figure 4) tries to remove the drawback of list of cluster when applied to the Hybrid solution, see Algorithm 1 for a code example. Each thread receives a pointer to the list of clusters that will be computed. Now, the access to the structure is no longer at only one access point. Thus, the quantity of critical section code is reduced further still.

Algorithm 1 Cluster insertion with proposed data structure and without critical section.

```

1: vector<list_of_clusters* > vector_lc;           ▷ create a vector of list of cluster
2: ...
3: for (long pid=0;pid<this->nprocs;pid++) do      ▷ for each process...
4:   for (unsigned int i=0;i<this->num_thread;i++) do  ▷ for each thread...
5:     vector_lc.push_back(new list_of_clusters());
6:   end for
7:   metalist_lc.push_back(make_pair(pid,vector_lc));
8: end for
9: ...
10: metalist_lc[pid]->lc_thread[i]->insert(c);     ▷ The variable "pid" is MPI Rank
11:                                                 ▷ process, "i" identifies the thread that
12:                                                 ▷ has to manipulate these data and the
13:                                                 ▷ variable "c" is a new cluster

```

4 Experimentation

The environment configuration used in this experimentation has the following characteristics: AMD Opteron 6200 1.6 GHz, L2 (2MB), L3 (6MB), 8 nodes with 64 cores each distributed in 4 sockets with three cache levels, 64 GB RAM

per node, Interconnection Gigabit Ethernet. The simulator was developed by using C++ (gcc 4.3.2), STL (C++ standard template library), MPI namespace (openmpi 1.4.3). The experimentations were carried out for 131072, 262144 and 524288 agents in 32, 64, 128, 256 and 512 cores. We used one MPI Process per core for the DM paradigm and eight OpenMP threads per MPI Process for the Hybrid paradigm. In addition, the number of MPI processes is calculated as the total number of used cores divided by eight.

In the experimentation, see Figure 5, we compare the total execution time of simulation obtained by using the list of cluster and the hybrid data structure proposed. As we can observe, there is no significant difference between both data structure until 128 cores. But there is a significant difference of total execution time when the simulation is performed with 256 and 512 cores. This occurs basically because the communication volume is higher in DM than in the hybrid approach. As presented before, in Section 3, the MPI processes have to send and receive agents that migrate along the simulation. This migration procedure is implemented with MPI.Broadcast and MPI.Wait methods. Therefore, the impact of this type of message is more visible when the number of nodes is increased.

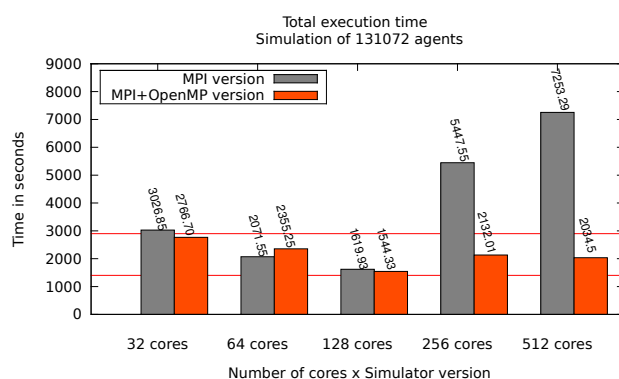


Fig. 5: Total execution time of Fish School simulation.

Also, we can note that the infrastructure with 128 cores is enough to execute 131072 agents. This means that 256 and 512 cores would only be necessary if the problem were bigger. In addition, we can observe that the execution time of the hybrid solution remains within a well-defined range, while the MPI curve has a significant increment. Therefore, if we extrapolate both curves for 1024 and 2048, we will get the hybrid solution with feasible execution times in contrast with the MPI solution. As we can see in the next experimentation, see Figure 6, we maintain the number of cores constant (512 cores) and increase the size of the agents dataset.

The MPI solution continues to show a high growth in total execution time. However, the hybrid solution does not follow the same exponential behavior as it is observed in Figure 6.

The data structure must be created thinking in the paradigm target in order to take advantage of it. Using a data structure developed for a specific paradigm requires adaptation if it is to be used in another one. Otherwise, slowdown at execution time can occur.



Fig. 6: Total execution time in 512 cores for different sizes of agent datasets.

5 Conclusion

In this paper, we detail a data structure for shared memory that supports a clustering algorithm. This data structure avoids the creation of critical sections and enables us to obtain more efficient parallel and distributed execution. We have tested this data structure using a Fish Schooling Agent-Based Model as a case of study.

The next stage of our research will be to validate this data structure with a bigger agent dataset in order to confirm the results obtained. In addition, it would be interesting to analyze the data structure with some memory and cache counters in order to evaluate the memory contention and cache misses.

As part of our main findings, we show that data structure must be conceived with the paradigm of programming in mind. The migration to another paradigm can cause slowdown to the solution without appropriate adjustment and changes. Also considering the current parallel and distributed architectures, it would be interesting for the data structure to have a flexible and dynamic behavior according to the paradigm used.

Acknowledgments

This research has been supported by the MINECO (MICINN) Spain under contracts TIN2011-24384 and TIN2014-53172-P. Some images come from [10].

References

1. Adhianto, L., Chapman, B.: Performance modeling of communication and computation in hybrid MPI and OpenMP applications. In: Proceedings of the 12th International Conference on Parallel and Distributed Systems, ICPADS'06. vol. 2, pp. 3–8 (2006)
2. Borges, F., Gutierrez-Milla, A., Suppi, R., Luque, E.: A Hybrid MPI+OpenMP Solution of the Distributed Cluster-based Fish Schooling Simulator. *Procedia Computer Science* 29(0), 2111–2120 (2014)
3. Borshev, A.: The Big Book of Simulation Modeling: Multimethod Modeling with AnyLogic 6. AnyLogic North America (2013)
4. Cappello, F., Etiemble, D.: MPI versus MPI+OpenMP on the IBM SP for the NAS Benchmarks. In: Supercomputing, ACM/IEEE Conference. p. 12 (2000)
5. Chan, M.K., Yang, L.: Comparative Analysis of OpenMP and MPI on Multi-core Architecture. In: Proceedings of the 44th Annual Simulation Symposium. pp. 18–25. ANSS '11, Society for Computer Simulation International (2011)
6. Chorley, M.J., Walker, D.W.: Performance analysis of a hybrid MPI/OpenMP application on multi-core clusters. *Journal of Computational Science* 1(3), 168–174 (2010)
7. Chow, E., Hysom, D.: Assessing Performance of Hybrid MPI/OpenMP Programs on SMP Clusters. Tech. rep., Lawrence Livermore National Laboratory (2001)
8. Collier, N., North, M.: Repast HPC: A platform for large-scale agent based modeling. Wiley (2011)
9. Hager, G., Jost, G., Rabenseifner, R.: Communication Characteristics and Hybrid MPI/OpenMP Parallel Programming on Clusters of Multi-core SMP Nodes. *Proceedings of Cray User Group Conference* 4(d), 54–55 (2009)
10. Icons8: Largest collection of icons (2016), <https://icons8.com>, icons/maps/draws under: CCBY-ND3.0
11. Kiran, M., Richmond, P., Holcombe, M., Chin, L.S., Worth, D., Greenough, C.: FLAME: Simulating Large Populations of Agents on Parallel Hardware Architectures. In: Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems. AAMAS '10, vol. 1, pp. 1633–1636. International Foundation for Autonomous Agents and Multiagent Systems (2010)
12. Rabenseifner, R., Hager, G., Jost, G.: Hybrid MPI/OpenMP Parallel Programming on Clusters of Multi-Core SMP Nodes. In: *Parallel, Distributed and Network-based Processing, 17th Euromicro International Conference on*. pp. 427–436 (2009)
13. Rubio-Campillo, X.: Pandora: A Versatile Agent-Based Modelling Platform for Social Simulation. In: *Conference Proceedings SIMUL 2014, The 6th International Conference on Advances in System Simulation*. pp. 29–34 (2014)
14. Solar, R., Suppi, R., Luque, E.: High performance distributed cluster-based individual-oriented fish school simulation. *Procedia Computer Science* 4, 76–85 (2011)
15. Wilensky, U.: NetLogo Ants model. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL. (1997), <http://ccl.northwestern.edu/netlogo/models/Ants>

XVI Workshop Tecnología Informática aplicada en Educación (WTIAE)

Coordinadores

Madoz, Cristina - Rueda, Sonia - Malberti, Alejandra - Russo,
Claudia

Comité Científico

Baldasarri, Sandra (España)
Cukierman, Uriel (Argentina)
Depetriz, Beatriz (Argentina)
Gorga, Gladys (Argentina)
Lanzarini, Laura (Argentina)
Madoz, Cristina (Argentina)

Malberti, Alejandra (Argentina)
Malbran, María (Argentina)
Manresa Yee, Cristina (España)
Rueda Sonia (Argentina)
Russo Claudia (Argentina)
Sanz Cecilia (Argentina)

A Concurrent Programming Language for Arduino and Educational Robotics

Ricardo Moran, Matías Teragni, Gonzalo Zabala

Centro de Altos Estudios en Tecnología Informática,
Facultad de Tecnología Informática,
Universidad Abierta Interamericana, Av. Montes de Oca 745, Ciudad Autónoma de Buenos
Aires, República Argentina
(+54 11) 4301-5323; 4301-5240; 4301-5248
{Ricardo.Moran, Matias.Teragni, Gonzalo.Zabala}@uai.edu.ar

Abstract. Arduino is currently one of the most popular platforms for educational robotics due to its low cost and large amount of available resources online. The Arduino software library provides an abstraction layer over the hardware details, making it possible for novices to build interesting projects. However, its lack of support for concurrency makes some educational robotics projects difficult. In this paper, we explore different approaches to solve this problem and we propose the implementation of a concurrent programming language supported by a virtual machine running on the Arduino board.

Keywords: education, robotics, programming language, virtual machine, Arduino, concurrency

1 Introduction

Arduino has become one of the most popular platforms for building electronic projects, especially among hobbyists, artists, designers, and people just starting with electronics. The Arduino software library and integrated development environment (IDE) provide an abstraction layer over the hardware details that makes it possible to build interesting projects without a complete understanding of more advanced microcontroller concepts such as interrupts, ports, registers, timers, and such. At the same time, this abstraction layer can be bypassed to access advanced features if the user needs them. These characteristics make the Arduino platform suitable for both beginners and experts.

However, there is one aspect in which the Arduino language lacks proper abstractions: concurrency. For all but the simplest projects, the `setup()` and `loop()` program structure [1] proposed by Arduino is not expressive enough. Even moderately complex problems require some sort of simultaneous task execution.

Furthermore, most educational robotics projects require the implementation of a device that performs two or more simultaneous tasks [2]. This poses a limitation on the type of educational projects that can be carried out, especially if the teaching subject is not robotics or programming itself.

2 Solution

We propose the implementation of a concurrent programming language [3] supported by a virtual machine running on the Arduino. We call this language UziScript and we expect it to become a suitable compilation target for visual programming environments such as Physical Etoys [4], Scratch for Arduino [5], and Ardublock [6], among others.

Given that the main purpose of this programming language is educational, it was designed based on the following principles:

- **Simplicity:** It should be easy to reason about the virtual machine and understand how it performs its job.
- **Abstraction:** the language should provide high-level functions that hide away some of the details regarding both beginner and advanced microcontroller concepts (such as timers, interruptions, concurrency, pin modes, and such). These concepts can later be introduced at a pace compatible with the needs of the student.
- **Monitoring:** It should be possible to monitor the state of the board while it is connected to the computer.
- **Autonomy:** The programs must be able to run without a computer connected to the board.
- **Debugging:** the toolchain must provide mechanisms for error handling and step by step code execution. Without debugging tools, the process of fixing bugs can be frustrating for an inexperienced user.

3 Related work

Even though the Arduino libraries don't provide any concurrency abstraction by default, third-party libraries are available that attempt to address this issue.

SCoop [7] is a library that implements cooperative multitasking. It provides the user a set of objects, functions and macros that essentially allow to specify multiple `loop()` functions with independent intervals. Each task has its own stack to hold the values of its computations, and instead of blocking it uses the reserved keyword "yield" to let another task use the processor. Some utility functions that simplify task creation are provided, as well as a `sleep()` function that works as a replacement for the Arduino `delay()` function.

ArduinoProcessScheduler [8] is another library that provides similar functionality. Unlike SCoop, this library forces the object-oriented paradigm into the user's sketch. Each task must be defined in its own class, which must inherit from `Process`, and the task's behavior must be implemented by overriding the `service()` method.

Other libraries worth mentioning are ArduinoThread [9] and everytime [10]. These two do not implement truly independent tasks, instead they provide facilities to schedule the execution of procedures at desired intervals. Each time a task is executed it will run to completion. The main difference between the two is that everytime

implementation is fully based on compile-time macros, which makes it impossible to truly dynamically create new processes.

A lot of other libraries exist but even though these libraries are useful, they all exhibit the same flaw: they require the user to understand the execution model provided by the library to use it effectively. If the programmer fails to understand the strengths and limitations of the library, they may be punished with potentially hard to debug errors.

Forcing the users to write code in a particular style is difficult (if not impossible) with just a library. In contrast, a virtual machine approach has the advantage of providing an isolation layer between the guest and host systems. The developer of the virtual machine is responsible for deciding which native abstractions to expose to the programmer and it has entire control over the way in which these abstractions are presented.

We can find several implementations of high-level languages and virtual machines for the Arduino platform. Most of them are based on preexistent general-purpose programming languages such as Java [11], Scheme [12] or Python [13]. A more interesting example is the Transterpreter project [14], a virtual machine designed to exploit concurrency on embedded systems. This virtual machine runs occam-pi programs on several platforms, one of which is Arduino [15]. Occam-pi is a variant of the occam programming language [16], especially designed to write concurrent programs based on the communicating sequential processes (CSP) process algebra [17]. Occam-pi has a rich set of runtime libraries that provide functions for interacting with Arduino features such as the serial port, PWM and TWI. Regarding performance, the execution of bytecodes has been reported be 100 to 1000 times slower than the execution of native code.

4 Description of the language

UziScript syntax¹ is based on C, which is familiar to most programmers including Arduino developers. The “task” keyword has been added to represent behavior that can be executed periodically at a configurable rate. For example, the following code will declare a task that will toggle the LED on pin 13 every second.

```
task blink() running 1/s { toggle(D13); }
```

UziScript does not require any type declarations, so to distinguish a function from a procedure two new keywords are introduced: “func” and “proc”.

```
func isOn(pin) { return read(pin) > 0.5; }

proc toggle(pin) {
  if isOn(pin) { turnOff(pin); }
  else { turnOn(pin); }
}
```

¹ For the complete grammar, see: <https://richom.github.com/Uzi/cacic2017/uzi.pegjs>

A program can have any number of tasks, and each task can be defined with a different interval as well as a different starting state, which can be either “running” or “stopped”. Additionally, special tasks that need to be run just once can be defined using the keyword “once” after the task name. This is especially useful to initialize variables and can be used as a substitute to the Arduino `setup()` function. The execution of each task at the correct time is performed automatically by the virtual machine but the user can invoke certain primitives to start, stop, pause, or resume a given task. Each task execution is independent, it has its own stack, and it shares memory with other tasks through specially defined global variables. This design allows users to write sequential programs in Arduino’s usual style and make them run concurrently without being concerned about the processor scheduling. Primitive instructions like `delay()` are provided to allow the user to block the executing task for a given amount of time without affecting the rest.

The “import” keyword allows to include external libraries, which can extend the primitive functionality provided by the language. For now, only a library to interact with DC motors is implemented. We are currently working on implementing libraries to provide process synchronization. Even though these are just experiments, we have working examples of recursive locks and synchronous channels².

5 Implementation

The UziScript firmware is implemented as a regular Arduino sketch that can be uploaded using the Arduino IDE. We initially target the Arduino UNO because it is the most popular model amongst beginners but we plan to support other models in the future. All the code (including the firmware, compilation tools, and libraries) is open source and it is available online³.

Internally, the firmware implements a stack-based high-level language virtual machine that uses a decode and dispatch bytecode interpreter [18] to execute UziScript programs. This implementation was chosen mainly because of its simplicity. Since the purpose of this language is educational, performance is not currently considered a high priority.

The stack and global variables are the only memory available to the user program. There is no heap or dynamic memory allocation. This allows for simpler virtual machine code and very compact object code [19]. The instruction set is designed to be compact. Almost all the instructions can be encoded using one byte for both the opcode and its arguments and just a few special instructions (such as branches) require an extra byte⁴.

Apart from the virtual machine, the firmware includes a monitor program that allows to interact with a computer through the serial port. Periodically, this monitor program

² These libraries can be found at: <https://github.com/RichoM/Uzi/tree/master/uzi/libraries>

³ Source code at: <https://github.com/RichoM/Uzi>

⁴ The instruction set specification can be found at: <https://richom.github.io/Uzi/ISA>

will send the status of the Arduino and receive commands, allowing the host computer to fully control the virtual machine.

By having these two programs running on the Arduino we can provide an interactive programming experience with a short feedback loop without sacrificing autonomy. Moreover, the monitor program permits the implementation of debugging tools that allow the user to stop the execution of any task, inspect the value of all the variables, explore the call stack, and execute instructions step by step.

6 Validation

We wrote three versions of the same program in order to compare the resulting code. The first version is written in standard Arduino code without using any concurrency library. The second version uses the SCoop library mentioned above. The third version is written in UziScript.

The chosen problem was inspired by the rules of the Robocup Junior Rescue competition [20]. However, the requirements have been slightly simplified. The objective of the challenge is to make a line follower robot [21] that can detect small patches of color representing “victims”. The line is made with standard electrical (insulating) tape, 1 - 2 cm wide. The victims are placed at random places along the circuit and every time the robot detects a victim it must stop and play an alarm to notify its finding. The number of beeps must be equal to the total number of victims found so far.

This challenge was chosen because it is representative of the problems a student learning robotics might face and its solution requires at least two concurrent tasks: following the line and detecting the victims. While each task on its own is trivial to solve, performing them at the same time can be challenging for a novice.

The same robot will be used for all versions of the program. This robot will be composed of an Arduino UNO, two DC motors, a single line sensor on its front, and a buzzer. The motors will be placed in a way that allows the robot to move forward, backwards and rotate. The line sensor will be made from an LED and a photoresistor. While not optimal, this type of sensor is inexpensive and easy to build.

For reasons of brevity and clarity, only extracts of the actual code will be included in this paper (the full source code will be made available online).

Arduino version of the Robot Rescue line follower⁵:

```
int victims, lastVictimTime, state;

void loop() { followLine(); detectVictims(); }

void followLine() {
  switch (state) {
    case 0:
      moveRight(); if (isOnBlack()) state = 1;
```

⁵ Full source code at: https://richom.github.io/Uzi/cacic2017/line_follower.ino

```

        break;
    case 1:
        moveRight(); if (isOnWhite()) state = 2;
        break;
    case 2:
        moveLeft(); if (isOnBlack()) state = 3;
        break;
    case 3:
        moveLeft(); if (isOnWhite()) state = 0;
        break;
    }
}

void detectVictims() {
    if (millis() - lastVictimTime < 100) return;
    if (isOnVictim()) {
        stopMotors(); victims++;
        for (int i = 0; i < victims; i++) {
            digitalWrite(buzzer, HIGH); delay(500);
            digitalWrite(buzzer, LOW); delay(500);
        }
        lastVictimTime = millis();
    }
}

```

Although the two tasks have been encapsulated into separate procedures (“followLine” and “detectVictims”), the execution of each procedure can interfere with the other. For this reason, the Arduino solution is implemented as a state machine. The state variable keeps track of the line follower state in order to know which direction to move at each step. Similarly, the `lastVictimTime` variable is needed to keep track of the last time in which a victim was found and avoid detecting the same victim more than once.

SCoop version of the Robot Rescue line follower⁶:

```

int victims;

void setup() {
    pinMode(sensor, INPUT); pinMode(buzzer, OUTPUT);
    mySCoop.start();
}

void loop() { yield(); }

defineTaskLoop(followLine) {
    while (!isOnBlack()) { moveRight(); yield(); }
}

```

⁶ Full source code at: https://richom.github.io/Uzi/cacic2017/line_follower.SCoop.ino

```

    while (!isOnWhite()) { moveRight(); yield(); }
    while (!isOnBlack()) { moveLeft(); yield(); }
    while (!isOnWhite()) { moveLeft(); yield(); }
}

defineTaskLoop(detectVictims) {
  if (isOnVictim()) {
    followLine.pause();
    leftMotor.stop(); rightMotor.stop();
    victims++;
    for (int i = 0; i < victims; i++) {
      digitalWrite(buzzer, HIGH); sleep(500);
      digitalWrite(buzzer, LOW); sleep(500);
    }
    followLine.resume();
    sleep(100);
  }
}

```

By allowing the user to express concurrency using the `defineTaskLoop` macro, `SCoop` encourages a clearer separation of responsibilities and simpler code. The `followLine` task, in contrast to the Arduino version, can now be expressed as a straightforward sequential procedure that can be paused and resumed as needed. Avoiding multiple detections of the same victims can be achieved using a simple `sleep(100)` instruction after resuming the `followLine` task. Since the `sleep()` function only blocks the currently executing task, the follow line procedure will keep running while the victim detection does not execute for 100 ms.

Even though the execution of each task is independent it is still possible for a task to affect another (as demonstrated by the `detectVictims` task) but this interaction must be declared explicitly while in the Arduino code any blocking operation might inadvertently break the behavior of another procedure.

This makes the `SCoop` solution more robust and scalable as the number of concurrent tasks increases. However, the abstraction presented by `SCoop` is leaked as soon as the user is forced to call the `yield` function on every loop. Failing to meet this requirement can result in incorrect behavior that could be hard to reproduce, hard to diagnose if the user does not understand the cooperative multitasking model used by `SCoop`, and hard to fix considering the limited support for debugging in the Arduino IDE.

UziScript version of the Robot Rescue line follower⁷:

```

var victims = 0;

task followLine() running {
  until isOnBlack() { moveRight(); }
  until isOnWhite() { moveRight(); }
  until isOnBlack() { moveLeft(); }
}

```

⁷ Full source code at: https://richom.github.io/Uzi/cacic2017/line_follower.uzi


```
    until isOnWhite() { moveLeft(); }
  }

  task detectVictims() running {
    until isOnVictim();
    pause followLine; stopMotors();
    victims = victims + 1;
    repeat victims {
      turnOn(buzzer); delay(500);
      turnOff(buzzer); delay(500);
    }
    start followLine; delay(100);
  }
}
```

UziScript version shares the same benefits as the SCoop version but it doesn't suffer from its weaknesses. By letting the virtual machine handle the scheduling without any involvement from the user, UziScript solution is both more robust and easier to use by a novice. Furthermore, the use of a virtual machine with debugging capabilities makes it easier for the user to diagnose and fix any incorrect behavior due to a concurrency bug.

While simple, this example shows how the Arduino code is more fragile and can introduce complexity as the program grows. Adding an extra concurrent task without disturbing the current behavior could be challenging for a novice, and the resulting code would be more complicated.

7 Future work

This project is still a work in progress. A lot of improvements can be made to address some of its current limitations. For instance, the only supported data type is 32-bit floating point. It would be desirable to support some other numeric types as well as aggregate data types such as arrays and strings. Also, the language does not support any dynamic memory allocation. It does not support objects or higher order functions. While the usefulness of these features for this particular language is not fully decided yet, we believe they could be integrated smoothly to some degree.

Performance can also be a problem. Our early measurements show the UziScript virtual machine to be 6 to 10 times slower than native Arduino code. While this performance might be good enough for certain applications, we believe there is still room for improvement.

Another potential problem is the program size. Considering that Arduino boards have very limited memory⁸, it is desirable that the virtual machine code and the user programs are as compact as possible. The instruction set, although designed to be compact, can be optimized further. And the current compiler does not perform any kind of optimization nor code compression.

⁸ For the Arduino UNO: 32 KB Flash, 2 KB SRAM, and 1 KB EEPROM

8 Conclusion

In this paper, we have analyzed the importance of concurrency in programming languages for educational robotics. We showed how Arduino, one of the most popular platforms lacks support for concurrency and we reviewed different approaches to solve this problem.

As an alternative solution, we proposed a concurrent programming language (called UziScript) supported by a virtual machine running on the Arduino. We described the implementation of the language and the virtual machine, and we tested its expressiveness by writing the same program using both standard Arduino code and UziScript. We conclude that the lack of concurrency in the Arduino language forces the user to be extra cautious when writing independent procedures to avoid accidental interactions between them. When the problem requires several simultaneous tasks, the resulting code tends to be complicated and difficult to extend. Using a concurrency library helps but it is not enough to solve the issue and can introduce problems if the user is not aware of how the scheduling works.

Concurrency is hard and not having proper support in the language makes it even harder, especially for beginners. UziScript's simple model for concurrency allows to express simultaneous behavior in a straightforward way. We believe this language has the potential to become an ideal environment for educational robotics.

References

- [1] Arduino, "Arduino Playground - Structure," [Online]. Available: <http://playground.arduino.cc/ArduinoNotebookTraduccion/Structure>. [Accessed 23 Julio 2017].
- [2] M. Resnick, "MultiLogo: A Study of Children and Concurrent Programming," *Interactive Learning Environments*, vol. 1, no. 3, 1990.
- [3] G. R. Andrews and F. B. Schneider, "Concepts and Notations for Concurrent Programming," *ACM Computing Surveys (CSUR)*, vol. 15, no. 1, pp. 3-43, March 1983.
- [4] Grupo de Investigación en Robótica Autónoma del CAETI (GIRA), "Physical Etoys," 2010. [Online]. Available: <http://tecnodacta.com.ar/gira/projects/physical-etoys/>. [Accessed 15 Junio 2017].
- [5] Citilab, "About S4A," 2015. [Online]. Available: <http://s4a.cat/>. [Accessed 15 Junio 2017].
- [6] "Ardublock | A Graphical Programming Language for Arduino," [Online]. Available: <http://blog.ardublock.com/>. [Accessed 15 Junio 2017].
- [7] F. Oudert, "fabriceo/SCoop: Simple Cooperative Scheduler for Arduino and Teensy ARM and AVR," 13 January 2013. [Online]. Available: <https://github.com/fabriceo/SCoop>. [Accessed 23 July 2017].

- [8] A. Wisner, "wizard97/ArduinoProcessScheduler: An Arduino object oriented process scheduler designed to replace them all," 15 January 2017. [Online]. Available: <https://github.com/wizard97/ArduinoProcessScheduler>. [Accessed 23 July 2017].
- [9] I. Seidel, "ivanseidel/ArduinoThread: A simple way to run Threads on Arduino," 15 May 2017. [Online]. Available: <https://github.com/ivanseidel/ArduinoThread>. [Accessed 23 July 2017].
- [10] K. Fessel, "fesselk/everytime: A easy to use library for periodic code execution.," 2 February 2017. [Online]. Available: <https://github.com/fesselk/everytime>. [Accessed 23 July 2017].
- [11] G. Bob, "HaikuVM: a small JAVA VM for microcontrollers," 2017. [Online]. Available: <http://haiku-vm.sourceforge.net/>. [Accessed 15 Junio 2017].
- [12] R. Suchocki and S. Kalvala, "Microscheme: Functional programming for the Arduino," in Scheme and Functional Programming Workshop, Washington, D.C., 2014.
- [13] "PyMite - Python Wiki," 2014. [Online]. Available: <https://wiki.python.org/moin/PyMite>. [Accessed 15 Junio 2017].
- [14] C. L. Jacobsen and M. C. Jadud, "The Transterpreter: A Transputer Interpreter," Communicating Process Architectures 2004, vol. 62, pp. 182-196, 2004.
- [15] C. L. Jacobsen, M. C. Jadud, O. Kilic and A. T. Sampson, "Concurrent event-driven programming in occam- π for the Arduino," Concurrent Systems Engineering Series, vol. 68, pp. 177-193, 2011.
- [16] M. Elizabeth and C. Hull, "Occam-A programming language for multiprocessor systems," Computer Languages, vol. 12, no. 1, pp. 27-37, 1987.
- [17] A. W. Roscoe and C. A. R. Hoare, "The laws of Occam programming," Theoretical Computer Science, vol. 60, no. 2, pp. 177 - 229, 1988 .
- [18] J. Smith and R. Nair, Virtual Machines: Versatile Platforms for Systems and Processes, San Francisco, CA: Morgan Kaufmann Publishers Inc., 2005.
- [19] T. Luerweg, "Stack based programming paradigms," in Seminar Concepts of Programming Languages (CoPL), 2015.
- [20] RoboCup, "RoboCup Junior - rescue," [Online]. Available: <http://rcj.robocup.org/rescue.html>. [Accessed 23 July 2017].
- [21] M. Pakdaman and M. M. Sanaatiyan, "Design and Implementation of Line Follower Robot," 2009 Second International Conference on Computer and Electrical Engineering, pp. 585-590, 2009.

Analyzing the design of Student Play, an educational module of Agent SocialMetric

Antoniet Kuz¹, Roxana Giandini^{1,2,3}

¹GIDAS, Universidad Tecnológica Nacional, Facultad Regional La Plata, La Plata, Buenos Aires, Argentina.

akuz@frlp.utn.edu.ar

²LIFIA, Facultad de Informática, Universidad Nacional de La Plata, La Plata, Buenos Aires, Argentina.

giandini@info.unlp.edu.ar

³CIC, Comisión de Investigación Científica y Tecnológica, La Plata, Buenos Aires, Argentina

Abstract. Currently, the incorporation of Information and Communication Technologies (ICT) into the educational process allows seeing the same from different dimensions. The educational software as an instructional support integrates the contents from a didactic and adapted to the teaching as a fundamental part in the environment of the society of the knowledge and the information. Seeking to deepen the relationship between teachers and students in the courtly context, we have developed a web tool called Agent SocialMetric based on two research areas, social networks and interface conversational agents. Integrating to Agent SocialMetric we extends the tool with a module called Student Play which is responsible for establishing interactive games for students to educate in values, through various interface conversational software agents. In the present research we present the tool and analyzing design considerations that we have taken into account. Finally, we will present the conclusions related to this compilation work.

Keywords: Design, Educational Software, Agent SocialMetric, Student Play

1 Introduction

Educational software technology is *"the study and ethical practice of facilitating learning and improving performance by creating, using, and managing appropriate technological processes and resources"*[13]. Educational software comes in several types. Educational software is more intended to educate kids and is easy to use adapted to the different age ranges. The most accepted educational software is normally known as edutainment, meaning it is a game in which users play and learn at the same time. Children in the world play, and this activity is so preponderant in its existence that it is said to be the *raison d'être* of childhood. Indeed, the game is vital; conditions of a harmonious development of the body, intelligence and affectivity, so the child who does not play is a sick child, body and spirit [2]. The war, the misery, by leaving the individual delivered only to the worry of survival, doing with it difficult or even impossible to play, make it wilt the personality.

As part of developing educational software, design has becoming an important research area, and is significant consider that exists a certain a number of dimensions that are particularly prone to lead to misunderstandings when considering educational design issues. Design of educational software may correspond to very different realities. As a consequence, many notions or issues may be addressed in different ways according to perspectives or matters of concern [16]. Software design is the process of defining the architecture, components, interfaces and other characteristics of a system. Centered on the latter we combine a thorough understanding of the function of a educative software with a thorough understanding of the needs of the user. In our research we show differently, aspects of interactions that appear to impact learners' perception of the educational software, and seen as part of a broader field of knowledge known as software ergonomics [12].

In our case we present a case study through a module Student Play of a tool Agent SocialMetric. The heart of our tool and his modules was developing in stage process which involves different activities and stages like learn, that involves understand and clearly define a problem, after defines imagine stage, by the end of the imagine stage a solution has been identified, and the make stage uses prototyping and testing to refine that solution [1]. Through the research process we studied as a part of design educational software, when the software is the more attractive and effective means facilitating the teaching and learning processes [3]. Trying to work on this need in this our research we show Student Play a software for develop emotional and social skills for preschoolers. The software is a tool for learners seems to depend on its degree of interactivity with the user and the level of complexity.

This paper is organized as follows: In section 2 we introduce the concept of educational software. In section 3 teaching-learning processes and the relation with human computer interaction. Section 4 presents a case study with Agent SocialMetric and in particular Student Play and explains some factors of human interaction. Finally, conclusions and future work are presented in Section 5.

2 Educational Software

The need of new software technologies grows and also facilitates the teaching learning process. At present some software platforms provides the facility and access to parents for a bird's view of the progress of their children in the classroom. More and more educational software online are coming up to help fill a need for more personalized, interactive educational experiences for students and teachers alike.

Educational software is also known as an educational or didactic program are all those computer programs created with the specific purpose of being used as a didactic means, that is, to facilitate the teaching and learning processes. This type of software integrates multimedia content and provides users a high interactivity level. The two features distinguish them from traditional teaching practices. Multimedia content, such as graphics, pictures, and sound helps engage students in their lessons and in the teaching-learning process. For instance, when it comes to learning history, students could go back and see videos or other online-based content related to it. Furthermore, online education software benefits teachers, allowing them to better connect with the

students and help them keep students interested in a lesson and also promotes a productive learning environment.

As part of educational software, games provide new way to explore reality and different strategies to work on it. It benefits students because we live in a society that is based on rules. In addition games enable learners to develop their imagination, think of numerous alternatives to a problem, discover different modes and styles of thinking, and encourage behavior change in addition to promoting group exchange. At present, pedagogues consider games as a fundamental tool for the teaching-learning process, to raise the independent work of students and to solve problematic situations in practical activity. A game is defined as the activity performed by one or more players, using their imagination or tools to create a situation with a certain number of rules, in order to provide entertainment or fun [17].

Educational games can be define as game technology for learning and teaching purposes which includes specific subject intended to meet desired learning results for students. Besides that, educational games have also opened a new door of learning with fun and excitement. Studies on educational games shows that the fun factor of a game inspire students to believe that they will learn better, at the same time they are captivated with educational games idea itself [7]

3 Reviewing design educational software

Interactivity is a quality that differentiates it from other means, it should be considered as the main indicator for software use. Human-computer interaction concerns design, evaluation, and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them [4]. The use of the TICs and software's in the classroom and in the teaching-learning process implies a greater degree of interactions, and abstraction of actions, an awareness and anticipation of what is often done "automatically", stimulating the passage of sensory-motor behaviors to the operative behaviors, generalizing the reversibility to all planes of thought.

Through software, the user can interact by textual stimuli, graphics, color, sound, animations; is able to process the information and show the result of what the user requested to do. As we see in the research of Andrew Sears, Julie A. Jacko [14] they consider educational software design must expand the concept of user-centered design (UCD) to one of learned-centered design (LCD). In this process there are several steps that involves analysis and their needs of learners and teachers, select pedagogy, select media/technology, prototype, assessment strategies, usability and design that involves cognitive considerations, interface design. User-centered design incorporates a continuum of research perspectives, according to Hinostroza et al. [6] they assert that in the design, development and evaluation of educational software, this areas converges to demonstrate the need for knowing and considering the context of use of educational software and for understanding users' perspectives about its roles and possibilities and hence supports a perspective on educational software.

As we see in the research of Gennaro Pellone [11] educational software it is meant the computer software program with which learners interact when being taught or

Agent SocialMetric tool has been expanded with the Student Play module which is in charge of establishing with the preschoolers interactive games to educate in values, through various conversational software agents, shown in Figure 3. We trying to incorporate the student in the interaction with the tool, we have developed a module called Student Play [8], in Figure 2 we show the model integrating Student play to Agent Socialmetric.

Student Play is software through which children acquire, in a playful way, basic knowledge about emotions: what they are, how they work and how they can be controlled. Provide strategies to identify emotions and the child understands the relationships between emotions and the situations in which they occur. Train in skills to express the feelings and needs associated with them. It seeks to provide guidelines to accept positive and negative emotional states and provide control techniques to regulate emotions without overdoing or avoiding them. An adequate management of emotions also promotes the development of good social skills, a reinforcement of self-esteem and a better capacity to face and solve problems that life may pose.

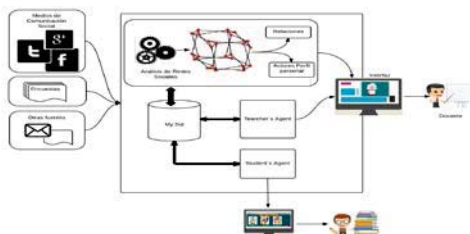


Fig. 2. Agent SocialMetric and Student Play



Fig. 3. Interface of Student Play

However, to develop a quality computerized educational material, in this case educational software, it is necessary to take into account the aspects considered important for its development, from several approaches like its environment, treatment and development methodology. On one hand from its environment we must consider the target population, area of knowledge, educational need, limitations, among others. Also it is necessary for the software to be effective to determine the target population to attend to the variable "age groups", considering the corresponding level of emotional maturity in each age group. In our case will be oriented to the preschool level in particular those that are in 5 years, since at this stage they develop the acquisition of values and principles indispensable for coexistence.

On the other hand centered in the design we consider some aspect as we considered important for HCI, which includes screens design looking that are easier to use for students. The development of this study is driven by the growing need to incorporate visual component in the field of design of educational interfaces. For that reason we taking into account cognitive structural, communicative design, aesthetic and visual criteria [15].

4.1 The interface design

In continuity with the aspects related to the requirements of software, it is interface design needed. The user interface in our project is a combination of its graphic elements and its navigation system. It is important to take care of the design, to ensure interest and ease of access. The user interfaces are those that include elements such as menus, windows, icons, buttons, scrollbars, tasks, displays and all those channels by which communication between the user and the computer is allowed. In our case the interface of software is the part of the application that the user sees and with which it interacts. We in development seek to achieve an intelligent interface that is easy to learn and use, designed specifically for those students who use it adjusting the user instead of having the latter to adjust to the software, as shown in Figure 4.

That is why we have considered the incorporation of interfaces software agents represented by an avatar to start to interact as shown in Figure 5. Interface agents are designed keeping in mind that a child may be attracted to certain objects, the way he perceives them, their shape, their colors, and the emotions it causes. This is why the child can select that personified agent that appeals to him, since affective attention is linked to the appearance of the application. It is based on the emotional bond that is established between the user and the application. In our case students perceives the information from the game, after system select and execute relevant knowledge and execute internal and external actions.

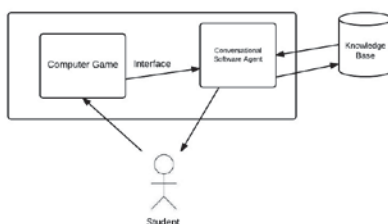


Fig.4. Model of interface design



Fig.5. Screen of selection of an avatar to start to play

4.2 Cognitive design

The subject of cognitive load theory has a number of active [10]. For the most part, cognitive load theory [10] has resided in the context of learning and instructional design and its use in the design of interactive entertainment, its usefulness in such a circumstance. The effective learning of a computer controls, games, interface, and general play should logically have a direct effect on how successfully a user can operate the game. The games embedded in the module are aimed for children in preschool age, while proposing ethical dilemmas that players-students should solve using their own scale of values, empower emotional intelligence.

When designing user interfaces and cognitive skills must be taken into account, perception of users, allow adapt the program to them. One of the most important things of a good interface, is to reduce dependence on users of their own memory, not forcing them to remember things unnecessarily (e.g. information that appeared on a previous screen) or to repeat operations already (e.g. example, enter the same data repeatedly).

In our case the user control it is given because the software does not require any pre-requisite in order to perform the tasks, and also by the talks gives by conversational software agent. This means that students are allowed to repeat the tasks as many times as necessary, in any sequence they prefer, giving them control over their own performance regarding the tasks separately. However, learners seem not to have a good control over the great scenario, that is, the software does not provide any kind of record containing scores and other relevant information for students to access in order to have a view of the whole learning process, becoming aware of which contents they are not mastering or what aspects of the tasks they need to improve.

Building embodied conversational agent, we wish helps users and must be able to interact with conversational behaviors that is, it must provide a way of realizing this set of conversational surface behaviors in a principled way. Also in our case agents adapted to the range of age to support conversational learning (see Figure 1), here, the emphasis is on the computer assisting rather than replacing the human teacher through a combination of computer (providing resources, tools and teaching) and human teacher. This involves a chat area, an online facility for learners to communicate at a distance, by typing and responding to messages.

4.3 Aesthetic design

In terms of web design, aesthetic consideration is the simplicity that refers to an approach to express something in a complete yet economic way and consistency. Reflecting on the intention of the system, the designer has to convey the messages to the reader thoroughly and get directly to the heart of the matter. We considered not put too many objects on the screen, and those that exist must be well distributed. Each visual element influences the user not only by himself, but also by its combination with the rest of elements present in the screen.

In order to fully this criterion, the software should allow learners to use the keyboard to replace some functions which are only made by using the mouse, such as

the selection and execution of some actions needed to perform the tasks. The software does not take into account the fact that students may be either experienced or novel computer users. The social experience of game play and the software agent creates the conditions to work on communication skills and personal traits from resilience to empathy, in our case story games and paintings can provide the motivation and engagement that helps students build knowledge and skills within a particular topic values such as friendship, etc.

The images on the screen are not just for decoration; they must help to encapsulate a mass of information. Screen layout should look clean as well. In our case we provides for the user tree types of animated images for the interface of the agent looking through these three options does not distracting and confusing. Just keep the content simple and to the point.

Then consistency means the agreement or harmony of parts or features to one another or a whole. In terms of our design, consistency involves the homogeneous pattern within the page and throughout all the system. The elements in the design are unified to produce a coherent whole. Such coherence ensures that your design will become visually apparent to the readers after their exploration. The background, text format, and color usage should be carefully chosen to produce a consistent screen layout for all our pages.

4.4 Visual criteria through the colors

It is important take into account that elements related to the use of colors in the design of educational software for the early ages [5]. A psycho pedagogical assessment is made, based on the influence of colors on educational software, as a means of teaching that favors the development of general intellectual abilities. It is explained how the achievement of the balance between colors favors an appropriate interaction of the children with the new technologies, as innovative resources and indisputable means for the fulfillment of the objectives in the educational field.

In the early ages, where preschool education can be framed, motivation is a fundamental component for the appropriation of knowledge. The children in these Ages, must acquire intellectual abilities related to the sensorial patterns, among which are the identification, verbalization and recognition of colors and their tonalities. The appropriation of these contents, will serve as a basis in the learning, writing, reading and mathematics. A psycho pedagogical characteristic in these ages is that the children are assimilating gradually the skills and contents for each year of life. The acquisition of knowledge comes through the institutions and the interaction with the adults that the surround. This makes it possible to create the appropriate conditions to promote maximum possible development in each of the children, as an end to Cuban education. To all of the above, as an indispensable element, the game is joined, which becomes a educational and pedagogical activity of vital importance. This favors, that the children of the first ages. Perform tasks, which encourage them to play, work, organize, socialize and learn with children and adults, in order to overcome the achievements of development. These tasks, they need the direction of the educators, as leaders of the process in the field of education.

Effective color choices go beyond just personal preference, because colors have an extraordinary ability to influence mood, emotions, and perceptions [5]; take on cultural and personal meaning; and attract attention, both consciously and subconsciously. The principal challenge is in balancing these complex roles that color plays to create an attractive, effective design. It is for that reason that we determine a set of balance colors, selecting a primary and secondary color palette (see Figure 3). This is due to the excess colors such as red, orange, gray or black, produce great levels of exaltation, fatigue and hyperactivity. This causes little assimilation of the contents proposed by the teaching in the fulfillment of the teaching tasks, since children do not reach the degree of concentration desired. In turn, the indiscriminate use of colors such as blue, green, violet, pink, gray and white in the educational software, propitiates a state of absolute rest and stillness, which favors the dream and exhaustion in children. This is detrimental to the objectives pursued by the software educational, for a favorable cognitive development.



Fig.3. Primary and secondary color palette

5 Conclusion

The design and implementation of educational software call into play two well established domains: software engineering and education. The research is based on showing some design considerations of the Student Play tool. This leads, to an exploration of the resource, in terms of visual design, which in most cases given the communicative features of the software. After identifying the target population, area of knowledge, educational need, limitations and resources for users of educational software and how to guide learning, for the designing educational software.

In this paper we present a set of design considerations of a software module as a computerized educational material or resource taking into account that the children construct their own knowledge through the interactions, thinking to find the balance between your current level of knowledge and the new.

The design of educational materials and resources such as Student Play involves the understanding of many aspects in order to be able to develop tools that effectively support the teaching-learning process. The use of new technologies opens extraordinary tending to improve the teaching-learning process.

As regards future work, we will try the interaction and between Agent SocialMetric and Student Play, emphasizing the design of application “character” (functionality and style) as well as the application’s structure, and expanding the notion of software to encompass the design of cognitive and social strategies.

References

1. *Thinking design with kids*. The Schoolhouse Centre for Progressive Education (2016).
2. K.S. Berger. *Psicología del desarrollo: infancia y adolescencia*. Médica Panamericana (2007).
3. G.G. Bitter and J.M. Legacy. *Using Technology in the Classroom*. Pearson Custom Education Series. Pearson/Allyn and Bacon (2008).
4. Elizabeth F. Churchill, Anne Bowser, and Jennifer Preece. Teaching and Learning Human-computer Interaction: Past, Present, and Future. *interactions*, 20(2):44–53, mar (2013).
5. E. Heller and J.C. Mielke. *Psicología del color: Cómo actúan los colores sobre los sentimientos y la razón*. Gustavo Gili (2015).
6. Enrique Hinostroza, Lucio E Rehbein, Harvey Mellar, and Christina Preston. Developing educational software: a professional tool perspective. *Education and Information Technologies*, 5(2):103–117 (2000).
7. Roslina Ibrahim, Samsudin Wahab, Rasimah Che Mohd Yusoff, Khalili Khalil, INSTEDT Desaru, and Azizah Jaafar. Student perceptions of educational games in higher education: An empirical study. *Issues in Information Systems*, XII(1):120–133 (2011).
8. Antonieta Kuz, Mariana Falco, Francisco Castellini, and Roxana Giandini. Student Play: un Módulo Educativo de Agent SocialMetric. In *4to Congreso Nacional de Ingeniería en Informática/Sistemas de Información (CONAISI)* (2016).
9. Antonieta Kuz, Mariana Falco, and Roxana Giandini. Understanding the teaching-learning environment through Agent SocialMetric. In *Information Systems and Technologies (CISTI), 2016 11th Iberian Conference on*, pages 1–4. IEEE (2016).
10. Fred Paas, Alexander Renkl, and John Sweller. Cognitive load theory and instructional design: Recent developments. *Educational psychologist*, 38(1):1–4 (2003).
11. Gennaro Pellone. Educational software design: A literature review. *Australian Journal of Educational Technology*, 11(1) (1995).
12. Gicele Vergine Vieira Prebianca, Vital Pereira dos Santos Junior, Christiane Fabiola Momm, Leonardo Furtado da Silva, and Hannelore Nehring. O uso de softwares educacionais como ferramentas mediacionais e de inclusão tecnológica. *Pesquisa Brasileira em Ciência da Informacao e Biblioteconomia*, 9(1) (2014).
13. Rhonda Robinson, Michael Molenda, and Landra Rezabek. Facilitating learning. *Educational technology: A definition with commentary*, pages 15–48 (2008).
14. A. Sears and J.A. Jacko. *Human-Computer Interaction: Designing for Diverse Users and Domains*. Human Factors and Ergonomics. CRC Press (2009).
15. Yen-Ru Shi and Ju-Ling Shih. Game Factors and Game-Based Learning Design Model. *International Journal of Computer Games Technology*, 2015:11 (2015).
16. P. Tchounikine. *Computer Science and Educational Software Design: A Resource for Multidisciplinary Work in Technology Enhanced Learning*. Springer Berlin Heidelberg, (2011).
17. Unesco and Unesco. Education Clearing House. *Educational Studies and Documents*. Number n.29-41 in Educational Studies and Documents. UNESCO (1978).

Applying fuzziness to enemy behavior in a serious game

Esteban Aitor Zapirain¹ and Stella Maris Massa¹,

¹ Facultad de Ingeniería, Universidad Nacional de Mar del Plata. Mar del Plata, Buenos Aires, Argentina.
{ezapirain, smassa}@fi.mdp.edu.ar

Abstract. This work presents the design and implementation of a real time fuzzy logic Mamdani-type controller for enemy artificial intelligence (AI) in a serious game. The game chosen is Power Down the Zombies, a serious game with educational purposes. A prototype of the game is coded in Javascript, and a comparison is made between the fuzzy enemy behavior and the classical, crisp one. The results of this work show the feasibility of fuzzy logic systems as a highly adequate choice for modelling human-like behaviour and decision making for an element in a videogame, with minimum processing costs.

Keywords: fuzzy logic, serious games, artificial intelligence, education.

1. Introduction

1.1 Game Artificial Intelligence (Game AI)

The term “Game AI” has many definitions, some broader than others. Game AI actually falls in the category of *weak AI* because it is not common to require that NPC (non-person characters) in a game show a full human intellect, but that they show some particular human-like response to events. This pseudo-intelligence is useful as long as it makes the game more fun and immersive, and totally superfluous and resource-demanding if it does not. In addition, game AI processes must run in real-time, along with highly demanding graphics and sounds. Therefore, game design must take into account what are the cost-benefit implications of implementing a simpler / more complex AI engine into a game. The motto of the game AI community is often thought to be *If the player cannot see it, why do it?* [1].

Due to all these constraints, game AI considers only the techniques and algorithms from academic AI that are appropriate for the context of videogames. Fuzzy logic is a fairly light, simple and adequate-for-decision-making technique that has been taken into account in several cases in the videogames industry.

1.2 Fuzzy logic

Fuzzy logic, first introduced by Lofti Zadeh in 1965 [2], is a superset of conventional logic, which allows truth values other than the classic binary “completely true” or “completely false”, by declaring a certain grade of membership of a variable to a value in the range between 0 and 1.

Fuzzy Logic Control (FLC) is an important application of fuzzy logic. It extends control theory by providing a better dynamic response to certain situations than classical controllers. A generic FLC block diagram is shown in Figure 1.

The two basic FLC approaches are Takagi-Sugeno and Mamdani controllers. Takagi-Sugeno controllers provide a smooth response to a system in which the data available are a set of inputs and its desired corresponding outputs [3]. Mamdani-type controllers, on the other hand, take advantage of human expertise in the operation of a system, without the need to know its underlying dynamics [4].

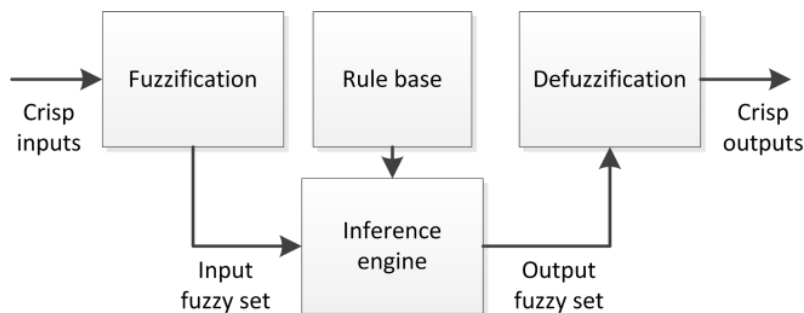


Fig. 1. Fuzzy Logic Controller block diagram [5].

1.3. Serious Games

The term Serious Games (SG) was introduced by Clark Abt in its homonymous book [6]. He defined SG as games with an implicit and carefully planned educational purpose, and which are not meant to be played for primary entertainment purposes.

At the time Abt was referring mostly to board and card games, but his definition can be extended to digital games. In 2005, Mike Zyda updates the term to refer to videogames. He defines SG as a mental challenge, played with a computer according to certain rules, which uses entertainment to promote citizen or corporate formation, with objectives in education, health, public policy and strategic communication [7].

2. Foundations

2.1. The Game

This work implements a Mamdani-type controller for enemy behaviour in a SG. The game chosen was *Power Down the Zombies*, a SG created by the Research Group in Interactive Technologies (GTI, Grupo de Investigación en Tecnologías Interactivas), based at the Faculty of Engineering of the National University of Mar del Plata in Argentina.

Power Down the Zombies (Fig. 1) is a 3D survival, tower defense action-strategy game with educational purposes. The educational objective is that players learn about the rational use of energy, clean alternative energy sources, electrical magnitudes and safety concepts. The narrative plot tells that in a close future, a virus has turned most of humanity into zombies, and it is up to the player, a scientist helped by a bunch of people who came to him looking for the protection of his fortress, to find a cure. In order to find this cure, he needs time, and every night the zombies strike from every front. He has learned that zombies die upon contact with direct light, so it is critical to construct and maintain a perimeter of lights during the day, and to generate enough energy to repel the zombie attacks at night.



Fig. 2. A screenshot from *Power Down the Zombies*. During the night, the power of lights is the only defense against zombie attacks.

3. Methods

3.1. Fuzzifying Enemy Behaviour

A simplified, sandbox version of Power Down the Zombies was coded in Javascript for this work (Fig. 2), for rapid prototyping and visualization of the effects of the fuzzy controller in enemy behaviour. Only the mechanics involved in player and zombie movement were kept functional, the game camera was set to top view at all times to be able to track zombies' trajectories in the whole terrain, and a simple house was drawn as a basic rectangle to model the home zone.

The "crisp" zombie movement algorithm is as follows: zombies always move in the direction of whoever is closer, the player or the house. This simple AI is interesting in the game dynamics, because it keeps the player alert for zombies around the perimeter of the house and discourages the strategy of keeping distance from everything to win.

This simple behaviour may be, however, somewhat predictable: a player can "control" a zombie by moving close to it, and then killing it by making it follow him into a light. This makes zombies feel "dumb" and lowers the immersion level.

A fuzzy controller applied to zombie movement results on a more "natural" response from the zombies as decision-makers, as they do not fully pursue the player or the house, rather making a more complex decision: a zombie at a timestep can decide, for example, to follow the player on 0.8 and target the house at 0.3. A subsequent calculation finally sets the target point of the movement, which will be an intermediate point between the player and the house. This determination is made every delta-time based on the current time positions, shaping a closed-loop controller.

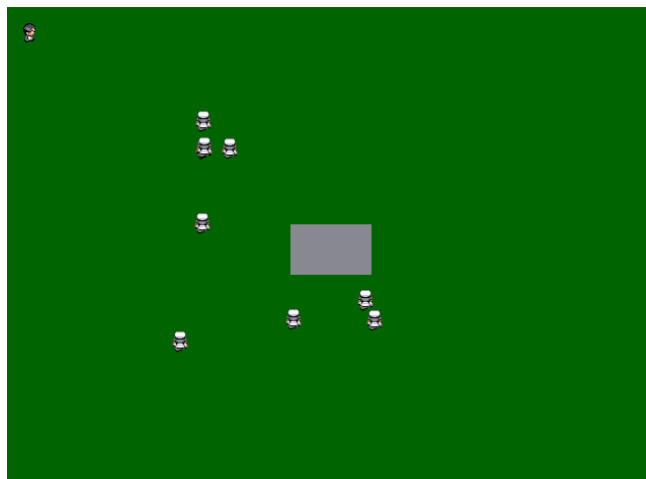


Fig. 3. A Javascript 2D simplified Power Down the Zombies was coded to test the fuzzy movement algorithm.

3.2. Fuzzy Model Design

The Fuzzy Inference System (FIS) was designed using MATLAB Fuzzy Logic Toolbox. A Graphical User Interface (GUI) was used to easily design FIS variables, membership functions, rules and visualize the system performance to variations of the inputs, as shown in Figure 3.

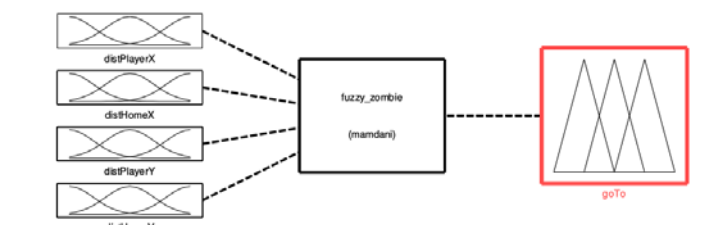


Fig. 4. Fuzzy controller block diagram, designed using MATLAB Fuzzy Logic Toolbox.

3.2.1. Variables

The terrain was set to 640 pixels wide and 400 pixels tall. In the actual Power Down the Zombies game, the field can be divided in as many cells as needed, so it is irrelevant which is the actual size of the terrain. The house was modelled as a rectangle at the center of the field, and actually zombies pointing to the house target the center pixel, which is (320,200).

Four input variables were considered: the horizontal (x-axis) and vertical (y-axis) distance from the zombie to the player (named `distPlayerX` and `distPlayerY`), and the corresponding distances to the house (`distHomeX` and `distHomeY`). Distances were directly considered in pixels. Due to the field size, the distance between the zombie and the player can vary from 0 to 640 in the x-axis, and reach a maximum of 400 in the y-axis. However, as the house stands still in the center of the field, the distance from the zombie to the house can be at most 320 in the x-axis, and 200 in the y-axis (half the field size).

Three fuzzy sets were defined for each input variable, named “Close”, “Medium” and “Far”. Trapezoidal membership functions were used in all cases, to lower the computational cost of the controller computations. The membership functions for the input variable `distPlayerX` are shown in Figure 4.

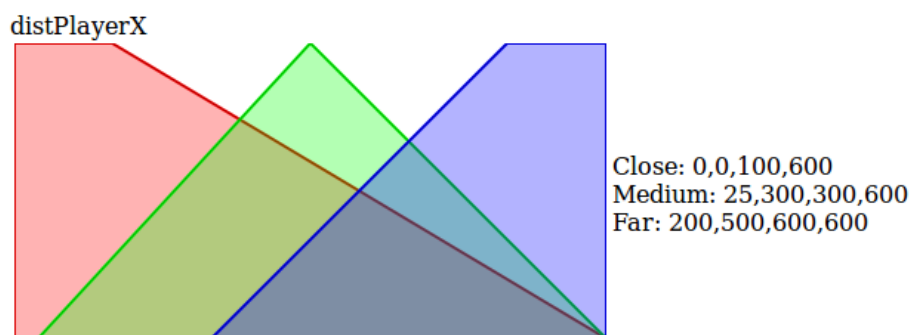


Fig. 5. Membership functions for the input variable `distPlayerX`.

A single fuzzy output variable was defined as `goTo`, with fuzzy sets “Player” and “Home”. The support range was from 0 to 100, as a generic percentage indicator. Trapezoidal membership functions were used for both sets.

3.2.2. Rule set

The rule set was simple. If the distance between the zombie and the player or the house was “close” or “medium”, then track that element. If the distance was “far”, then go to the other. There were a total of 12 rules in the system.

3.2.3. Output

The library was modified to return the fuzzified outputs of each fuzzy set, without performing the step of defuzzification. After the fuzzified outputs are calculated, target for the zombie gets updated. The first fuzzy output corresponds to the linguistic term “go to Player” and the second means “go to House”, so the following equation is evaluated for coordinates x and y .

$$\text{target.x} = \text{player.x} * \text{fuzzy_gotoPlayer} + \text{fuzzy_goHome} * (\text{house.x} - \text{player.x}) . \quad (1)$$

$$\text{target.y} = \text{player.y} * \text{fuzzy_gotoPlayer} + \text{fuzzy_goHome} * (\text{house.y} - \text{player.y}) .$$

The overall result is that zombies tend to “block” the way between the player and the house, a human-like behaviour that adds to the experience of the game. Resulting trajectories for the zombies are shown in Figure 5.

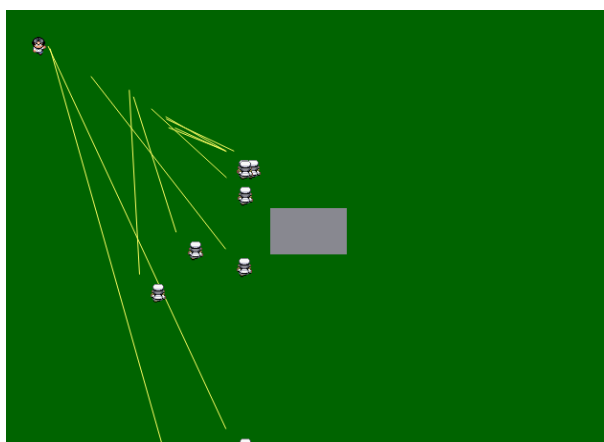


Fig. 6. Zombie target paths outputted by the fuzzy logic controller.

4. Results

The aim of this work was to evaluate the possibilities of applying FLC systems to game AI in a serious game. To this end, a FLC algorithm based on an open fuzzy logic library was implemented on a sandbox version of an existing serious game.

The performance of the controller was very satisfying. The processing cost of the FLC computations was minimal, and customizable by changing the rate of update of the target point. In the tests taken, the algorithm ran every 0.2 seconds (5 times per second).

Fuzzy logic showed as very appropriate for game AI design. Mamdani FLC systems are great for extracting expert knowledge and transferring it into an algorithm. This is very helpful when the goal is to model realistic human-like behaviour and decision making in a game element. Also, it eases the development process as it provides a link to meet design specifications with functional code.

In the course of this work, this particular implementation of a FLC system needed to obtain the fuzzified outputs for every output set, instead of the crisp defuzzified output. For this reason, the fuzzy library had to be slightly modified, to skip the

defuzzification process and return the fuzzified output set. This customization was possible because of the existence of an open fuzzy library, and the simplicity of the algorithm allows its migration to other programming languages.

5. Conclusions and future work

A real time Fuzzy Logic Controller was successfully implemented for non-player character decision making in a serious game. A prototype of this game was coded as a sandbox to easily deploy the fuzzy AI algorithm, and its behavior was compared against traditional AI.

The results of this work show the potential of fuzzy logic applied to serious games. Further research is necessary to fully exploit this potential, and this paper is intended as a kickstart for future lines of research.

The immediate next steps are to implement this algorithm in the actual Power Down the Zombies game, and to explore other suitable models and technologies for game AI improvement.

6. References

- [1] Ian Millington. Artificial Intelligence for Games (The Morgan Kaufmann Series in Interactive 3D Technology). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2006.
- [2] Zadeh, Lotfi A.(1990) 'FUZZY SETS AND SYSTEMS', International Journal of General Systems, 17: 2, 129 — 138.
- [3] T. Takagi and M. Sugeno, “Fuzzy identification of systems and its applications to modeling and control,” IEEE Trans. Syst., Man, Cybern., vol. 15, pp. 116–132, 1985.
- [4] Mamdani, E. H. (1977). Application of fuzzy logic to approximate reasoning using linguistic synthesis. IEEE Transactions on Computers, 100(12), 1182-1191.
- [5] Panduru, K. K., Riordan, D. and Walsh. J (2014). Fuzzy Logic Based Intelligent Energy Monitoring and Control for Renewable Energy. 25th Irish Signals and Systems Conference (ISSC 2014). Conference Paper.
- [6] Abt, C. (1970). Serious Games. New York, Estados Unidos: Viking Press.
- [7] Zyda, M. (2005). From visual simulation to virtual reality to games. In Computer 38, pp. 25-32.

Caso de aplicación de representación del conocimiento utilizando grafos conceptuales en un sistema de corrección automatizado de exámenes

María Alejandra Paz Menvielle (pazmalejandra@gmail.com), Mario Alberto Groppo (proyale@groppo.com.ar), Marcelo Martín Marciszack (marciszack@gmail.com), Analía Guzmán (analía.guzman@the-group.com.ar), Karina Ligorria (karinaligorria@hotmail.com), Martín Casatti (mcasatti@gmail.com)

CIDS-Centro de Investigación, Transferencia y Desarrollo de Sistemas de Información, Departamento de Ingeniería en Sistemas de Información, Facultad Regional Córdoba – Universidad Tecnológica Nacional.
Maestro Marcelo López esq. Cruz Roja Argentina – Córdoba 0351 – 4686385

Resumen El presente trabajo describe la implementación de una base de conocimientos, implementada como un grafo conceptual, utilizado como parte de un sistema de corrección de exámenes automatizado, sobre respuestas redactadas en forma de texto en lenguaje natural. Se presentan los conceptos básicos de representación del conocimiento y las características de los grafos conceptuales aplicados a tal fin y se describe el mecanismo utilizado para evaluar una respuesta utilizando un grafo conceptual como repositorio de los conceptos del dominio de la materia Paradigmas de Programación, utilizada en este caso de aplicación en particular. Por último, se explica la utilización de conceptos, relaciones y equivalencias dentro del proceso de evaluación, las reglas que el sistema implementa para simplificar las expresiones, minimizar el almacenamiento necesario y optimizar los tiempos de búsqueda y se detallan los pasos del proceso de evaluación.

Keywords: análisis de textos; grafos conceptuales; representación del conocimiento; detección de patrones

1. Contexto

El presente trabajo forma parte del proyecto de investigación y desarrollo homologado por la Secretaría de Investigación, Desarrollo y Posgrado de la Universidad Tecnológica Nacional, desarrollado en el ámbito del CIDS – Centro de Investigación, Desarrollo y Transferencia en Sistemas de Información, dentro del Departamento de Ingeniería en Sistemas de Información de la Facultad Regional Córdoba.

El dominio de aplicación seleccionado para la validación de la presente propuesta, se corresponde con los contenidos mínimos fijados para la asignatura Paradigmas de Programación[1], que pertenecen al bloque de tecnologías básicas dentro del área programación y están principalmente referidos a los paradigmas lógicos, funcional y de orientación a objetos.

2. Introducción

Un grafo conceptual[2] es un sistema de notación simbólica y de representación del conocimiento, impulsado principalmente por John F. Sowa. El concepto se basa en los gráficos existenciales[3] de Charles Sanders Peirce, en las estructuras de redes semánticas y en datos de la lingüística, la filosofía y la psicología[4].

Se utilizará, para la detección de rutas dentro de la información almacenada en el grafo, el enfoque de búsqueda inexacta, el cual a través de la especificación de un cierto umbral puede determinar tanto una validez total como una parcial para las respuestas analizadas. De esta manera se podrán detectar respuestas válidas aún con conceptos faltantes, con relaciones similares aunque no idénticas y con órdenes alterados o invertidos en las relaciones[5].

La elección de este enfoque de búsqueda de patrones, durante las primeras etapas de investigación, se realizó considerando la naturaleza inherentemente variable de las respuestas de los alumnos frente a un mismo conjunto de preguntas de examen.

Actualmente el equipo de investigación ha desarrollado el prototipo de un sistema que trabaja con la búsqueda de patrones en una base de datos de grafos que representa, de manera adecuada y simple, las estructuras del lenguaje natural con el fin de determinar si la respuesta a una pregunta es correcta, dentro de un dominio especificado y aplicando las técnicas y las reglas gramaticales relacionadas con los lenguajes estructurados.

El objetivo del presente trabajo es describir el mecanismo de corrección diseñado, utilizando grafos como base de conocimiento para la corrección automatizada de exámenes escritos por estudiantes en formato de texto y mostrar ejemplos de la aplicación de éste enfoque.

3. Los grafos como herramientas de representación del conocimiento

Dentro de los esquemas de representación del conocimiento encontramos distintas vertientes que buscan la manera más eficiente de representar los hechos de manera tal que la utilización de los mismos para la solución de problemas sea eficiente[6].

Pero cada dominio tiene conjuntos distintos de problemas y por ende los mecanismos de representación van a variar de uno a otro por lo que no existe un único mecanismo válido para dicha representación, sino que cada problema a resolver debe ser analizado antes de resolver el mecanismo de representación a utilizar, planteando dos interrogantes:

1. ¿Cómo escoger el mecanismo de representación que nos permita hacer una traducción fácil de los conceptos del mundo real a la representación?
2. ¿Cómo ha de ser esta representación para que pueda ser utilizada de forma eficiente?

Es importante destacar en este punto la diferencia entre información y conocimiento, debido a que un sistema de corrección automatizado no busca de ninguna manera

evaluar la información que el alumno posee, sino su conocimiento de un tema dado. En el presente trabajo llamaremos “conocimiento” al conjunto de datos que modelan de forma estructurada la experiencia que se tiene sobre un cierto dominio o que surgen de interpretar datos básicos[6].

3.1. ¿Qué es la representación del conocimiento?

En[7] Randall Davies menciona que la representación del conocimiento es un “sustituto de la realidad”. Esto se debe a que el proceso de razonamiento se realiza de manera interna al individuo, pero los objetos o entidades sobre los que se razona son externos al mismo; luego surge la necesidad de que el individuo consiga una representación interna sobre la cual pueda razonar, efectivamente “sustituyendo” la realidad por su representación.

Esto no es distinto a la necesidad que tiene un sistema automatizado de representar de manera interna las entidades sobre las que debe trabajar.

3.2. Estructuras de redes para la representación del conocimiento

Tal como se menciona en[8], Quillian[9] y Shapiro & Wodmansee[10] fueron los primeros en definir los esquemas de representación formalizados.

Los elementos básicos que se encuentran en este tipo de representaciones, comúnmente denominadas *redes semánticas*[11], son:

- Una estructura de datos formada por nodos, que representan conceptos, unidos por arcos que representan las relaciones entre esos conceptos.
- Un conjunto de procedimientos de inferencia que operan sobre las estructuras de datos.

Entre los tipos más utilizados de redes semánticas encontramos las *Redes IS-A*[12], en las que los enlaces entre los nodos están etiquetados, los *Grafos conceptuales*[13], en las que existen dos tipos de nodos, de Concepto y de Relación y las *Redes de marcos*[14], en las que los puntos de unión de los enlaces son parte de la etiqueta del nodo.

Como base para el sistema de corrección automatizado que aplica los conceptos del presente trabajo se optó por el uso de grafos conceptuales como mecanismo de representación del conocimiento por dos cuestiones fundamentales:

1. Es relativamente simple representar los conceptos y relaciones propios de una materia con contenido estructurado, en una estructura de éste tipo.
2. Existen sistemas de persistencia y bases de datos, que aplican directamente el esquema de grafos, lo que posibilita una mayor simplicidad al momento de la implementación de la solución informática.

4. Validación conceptual de una respuesta a examen

La riqueza expresiva de la lengua española es uno de los principales obstáculos a vencer en el planteo de un sistema de corrección automatizado.

Si bien el idioma nos permite expresar el mismo concepto de diversas formas, esto obstaculiza la labor de una computadora debido a que la misma debe reconocer el contenido conceptual de una respuesta de forma independiente de los términos utilizados para expresar dicho concepto, existiendo infinidad de expresiones para el mismo concepto, todas ellas válidas y en cierta medida equivalentes.

Es por eso que el enfoque utilizado en el presente trabajo de investigación busca evaluar la *similitud conceptual* entre dos respuestas a una pregunta de examen, una proporcionada por el docente y la otra por el alumno, descartando cualquier tipo de comparaciones textuales.

El abanico de posibilidades se amplía al considerar diversas respuestas posibles y correctas, así como a la capacidad del sistema de tener en cuenta sinónimos y términos equivalentes en la redacción de la respuesta por parte del alumno y/o del docente.

Al momento de establecer esas relaciones conceptuales y, seguidamente, compararas para obtener una evaluación de una respuesta dada, es necesario que todos los conceptos y relaciones involucrados se encuentren en el mismo espacio de dominio. Es decir, que todos esos conceptos y relaciones se hallen representados en el mismo grafo.

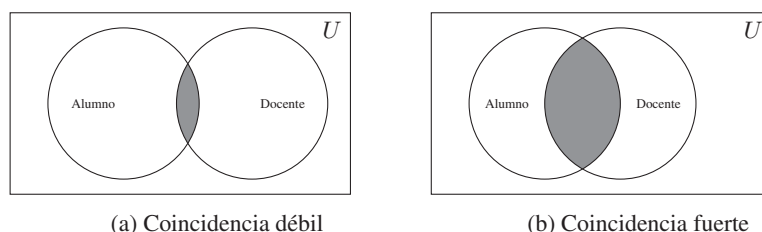


Figura 1: Dominios de la respuesta del alumno y del docente

La evaluación surge de comparar la exactitud de la intersección (ver Figura 1) que se encuentra entre los conceptos incluidos por el docente en su respuesta ideal (denominada *respuesta base*) y los conceptos suministrados por el alumno en su respuesta propuesta (denominada *respuesta candidata*).

En las secciones posteriores veremos en detalle cómo se determinan los conceptos y relaciones en cada una de estas respuestas, como se amplía el espacio expresivo a través de sinónimos y equivalencias y como se simplifican y almacenan las respuestas a fin de optimizar el proceso de comparación y evaluación.

5. Definición y detección de conceptos y relaciones

En [2] John F. Sowa define el *Grafo Contextual*, en el que basaremos nuestro estudio, como:

“Un grafo conceptual es un grafo finito, conectado, no dirigido y bipartito con nodos de un tipo denominados conceptos y nodos de otro tipo denominados relaciones conceptuales. Un grafo conceptual puede consistir de un único concepto pero no puede contener relaciones conceptuales con aristas sin enlazar”

En el marco del presente trabajo denominaremos *Conceptos* a todos aquellos términos contenidos en una respuesta textual que tengan un valor particular para la evaluación de los conocimientos de un alumno, de acuerdo al dominio de la asignatura Paradigmas de Programación. En este mismo sentido denominaremos *Relación* a los nexos que unen dos Conceptos, constituyendo una unidad de información.

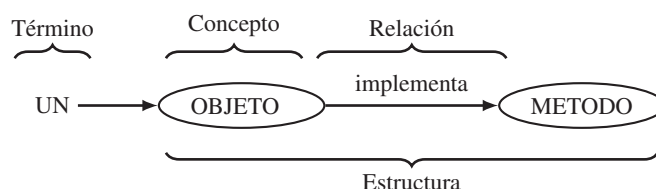


Figura 2: Mínima estructura de representación

En esta implementación en particular se simplificará la representación de relaciones mediante etiquetas apropiadas en el propio enlace entre dos conceptos (Figura 3a), haciendo de esta forma innecesaria la existencia de los nodos relación de la definición original (Figura 3b) pero manteniendo su utilidad y funcionamiento originales.

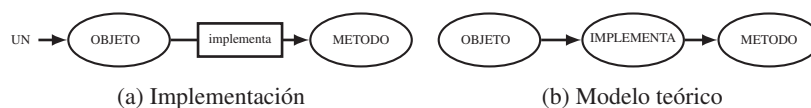


Figura 3: Modelo teórico

Existen dos escenarios posibles para la detección de conceptos y relaciones. Uno de ellos se desarrolla cuando el docente ingresa las respuestas válidas a una pregunta de examen. El segundo escenario se desarrolla cuando el alumno ingresa su respuesta candidata, previa a la evaluación[15].

5.1. Proceso de la respuesta del docente y del alumno

Este escenario comienza con la redacción de la respuesta en formato libre, por parte del docente, en idioma castellano, dando lugar a un texto que luego se envía al módulo de corrección ortográfica que realiza todas las revisiones necesarias hasta obtener un texto correcto.

Una vez validada la ortografía el conjunto de términos se envía a la base de conocimientos la cual marca todos los conceptos ya existentes e informa al docente de los términos que no reconoce.

El docente tiene, en este punto, la posibilidad de completar la base de conocimientos, indicando qué conceptos considera que deben ser incluidos en la base, cuáles se

deben considerar equivalencias, y qué términos no aportan ningún valor y pueden ser ignorados.

A modo de guía, el sistema automático sugiere que los sustantivos y adjetivos no reconocidos sean considerados conceptos, pero esto de ninguna manera es una decisión automática, sino que siempre recae en el docente la responsabilidad de ampliar la base de conocimientos.

Para poder construir una ruta que permita evaluar la respuesta, los conceptos deben estar correctamente encadenados, siguiendo la secuencia *Concepto* \rightarrow *Relacion* \rightarrow *Concepto*. El sistema analiza la existencia de estos encadenamientos y en caso de detectar algún error se detiene el procesamiento hasta tanto todas las secuencias involucradas sean válidas. En esta etapa el docente tiene la posibilidad de agregar relaciones faltantes o determinar si son equivalentes de algunas ya existentes.

El sistema brinda asistencia, también en este caso, para las relaciones faltantes, sugiriendo que los verbos sean considerados relaciones entre conceptos.

La operatoria del docente finaliza en esta etapa y la respuesta está en condiciones de ser almacenada en la base de conocimientos.

El escenario correspondiente al procesamiento de la respuesta del alumno es en gran medida similar al escenario correspondiente al docente, con la excepción de que, en el paso final, frente a conceptos o relaciones inexistentes el sistema no sugiere su agregado a la base de conocimientos, sino que los descarta.

Una vez detectados todos los conceptos y relaciones existentes en la base de conocimientos, el sistema procede a ejecutar el proceso de evaluación, posteriormente descripto en detalle.

6. Sinónimos y equivalencias

A fin de aumentar la potencia expresiva de las respuestas posibles, tanto generadas por los docentes como proporcionadas por los alumnos, el sistema de corrección automatizado implementa un mecanismo tendiente a aumentar el contenido y variedad textual de las respuestas, pero manteniendo su validez conceptual.

Para ello implementa el concepto de *Sinónimos y Equivalencias*. Los sinónimos y las equivalencias son dos términos relacionados pero no intercambiables.

En este contexto, un Sinónimo es un término que tiene una equivalencia del 100 % con respecto al concepto al que hace referencia. Por otra parte, un término Equivalente guarda una relación con el concepto al que referencia, pero esta relación no es de una exactitud plena, sino que puede tener un grado de ponderación variable de acuerdo a lo aproximado que esté su significado al concepto original. Podemos entender un sinónimo como un caso particular de equivalencia, un término con una equivalencia exacta al concepto original y que es intercambiable con el mismo.

El sistema de corrección realiza su análisis contemplando todo el abanico de posibilidades que proporcionan los conceptos registrados en la base de conocimientos, sus sinónimos, y sus equivalentes, aumentando de esta manera en gran medida la expresividad de las respuestas que puede analizar.

En la sección correspondiente al proceso de evaluación se presentará en mayor detalle el análisis que se realiza con los sinónimos y las equivalencias.

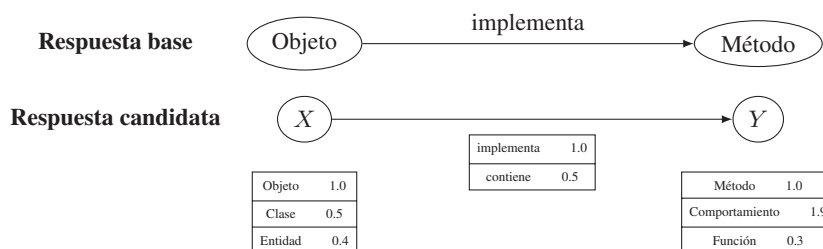


Figura 4: Ejemplo de conceptos y relaciones con equivalencias

7. Simplificación y almacenamiento

A fin de mantener una estructura coherente en la base de conocimientos, todas las respuestas, ya sean las provistas por el docente como las elaboradas por el alumno como parte de la evaluación, se someten a un proceso de simplificación previo a su almacenamiento en la base de datos de grafos.

Este proceso busca eliminar información accesoria que no aporta valor a la corrección y al mismo tiempo minimizar tanto el tamaño de la base de datos como los tiempos de búsqueda asociados a la corrección.

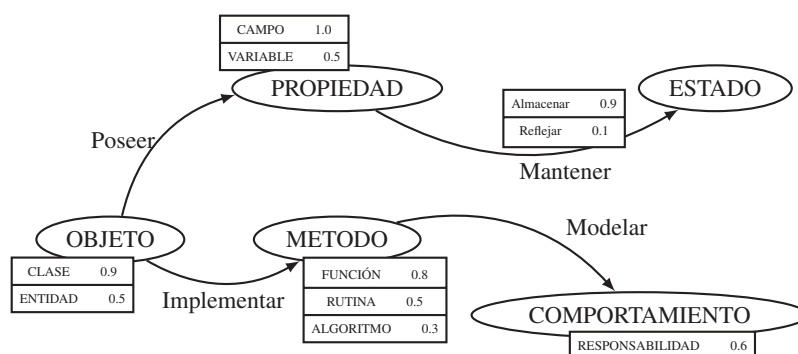


Figura 5: Almacenamiento de una respuesta

Como parte del proceso de simplificación previo al almacenamiento el sistema unifica el formato de los conceptos y relaciones siguiendo las siguientes reglas:

1. Todos los *Conceptos* se almacenan en mayúsculas.
2. Todas las *Relaciones* se almacenan en formato CamelCase.
3. Todos los *Verbos* se transforman en su forma infinitiva.
4. Los *Conceptos compuestos* se concatenan separados por un espacio.
5. Todas las *Relaciones compuestas* por varios términos se componen en una única frase antes de ser almacenada, y la misma se transforma en CamelCase.

Se pueden ver ejemplos de aplicación de las reglas de simplificación y almacenamiento en el Cuadro 1 y la estructura almacenada en la Figura 5.

Regla	Tipo	Término	Resultado
Regla 1	Concepto	Objetos	OBJETO
Regla 2,3	Relación (Verbo)	contiene	Contener
Regla 4	Concepto compuesto	Lógica binaria	LÓGICA BINARIA
Regla 5	Relación compuesta	formado por	FormadoPor

Cuadro 1: Reglas de simplificación y almacenamiento. Ejemplos.

8. El proceso de evaluación

Para establecer una línea base con la cual realizar la evaluación de la respuesta de un alumno se debe contar con el valor de la respuesta del docente, denominada *respuesta base*. La valoración de la respuesta base del docente es sencilla ya que es simplemente el resultado de contar la cantidad de conceptos y relaciones que la componen, según se ve en la fórmula 1.

$$V_r = C + R \quad (1)$$

Siendo V_r el Valor de la respuesta, C la cantidad de conceptos presentes y R la cantidad de relaciones.

Posteriormente se procede a evaluar la respuesta del alumno, denominada *respuesta candidata*, lo que da inicio a la evaluación propiamente dicha.

Se realiza una enumeración de los conceptos contenidos en la respuesta del alumno, mediante una consulta a la base de conocimientos. La lista de conceptos se procesa, en el paso siguiente, para determinar si existe alguna ruta de Conceptos y Relaciones que incluya los conceptos vertidos por el alumno en su respuesta.

Una vez obtenida la ruta, cada uno de los conceptos y relaciones se evalúa para obtener el peso de cada uno de ellos, teniendo en cuenta si es un Concepto o Relación exacta o si es una equivalencia, las cuales poseen menor peso con respecto al total de la respuesta.

Existen dos escenarios al calcular el peso de los distintos conceptos que componen la respuesta.

Misma posición El concepto o la relación es igual al existente en la respuesta base y tiene la misma ubicación. En este caso el valor es 1, igual que en la respuesta del docente.

Desplazado El concepto o la relación es igual al existente en la respuesta base pero su ubicación en la respuesta no es igual a la ubicación en la respuesta del docente. En este caso el valor del concepto desplazado se calcula por la fórmula 2

$$C_d = 1 - \left| \frac{P_c - P_b}{n} \right| \quad (2)$$

Siendo C_d el valor del concepto desplazado, obtenido a partir de P_c , posición del concepto en la respuesta candidata, P_b , posición del concepto en la respuesta base y n la cantidad de conceptos en la respuesta base.

Este planteo considera que mientras más alejado esté el concepto de su ubicación correcta, su valoración en la respuesta candidata disminuye.

En este paso se gestionan las equivalencias, para las cuales cada concepto tendrá una tabla, definida por el docente, cuyo fin será ponderar cada una de las equivalencias a un valor numérico entre 0 y 1, donde 1 será una equivalencia completa, que indicará que ambos términos pueden usarse indistintamente. De esta forma el valor del concepto se obtiene directamente de la tabla, según $C_e = C_{E_i}$, siendo C_e el valor del concepto equivalente y C_{E_i} el valor de la equivalencia obtenido de la tabla correspondiente (ver Figura 4).

Los conceptos existentes en la respuesta del alumno que no se encuentran en la base de conocimientos, o que no son equivalentes a conceptos existentes, se descartan y no tienen incidencia en el cálculo del valor de la respuesta.

Una vez ajustados los pesos, se procede a calcular el valor de cada una de las respuestas (candidata y base). Estos valores se comparan entre sí (según la ecuación 3) y el resultado nos indica el grado de acercamiento de la respuesta candidata a la respuesta base. Este valor se informa al docente a modo de *calificación sugerida*.

$$C_r = \frac{V_c}{V_b} \quad (3)$$

Siendo C_r la calificación relativa, V_b el valor de la respuesta base (del docente) y V_c el valor de la respuesta candidata (del alumno).

9. Conclusiones y trabajos futuros

Se ha expuesto un caso de técnicas para la representación de conocimiento por medio de un grafo conceptual, en el contexto de una materia de la carrera de Ingeniería en Sistemas de Información, de la Universidad Tecnológica Nacional, Facultad Regional Córdoba.

Las técnicas expuestas en el presente trabajo han demostrado ser eficaces, considerando que existe un prototipo de sistema de corrección automático que, utilizando los mecanismos aquí mencionados, se encuentra en fase de pruebas. Dicho prototipo explota las características aquí descritas con respecto a valoración conceptual de las respuestas, detección de conceptos y relaciones, gestión de un universo expresivo amplio, minimización de expresiones para su almacenamiento, y persistencia de la base de conocimientos.

El prototipo mencionado ha podido evaluar un conjunto de respuestas en un nivel similar a como lo haría un docente, lo que refuerza la hipótesis de que un grafo conceptual es un buen esquema de representación del conocimiento para sistemas de corrección automatizados.

Está previsto extender el presente estudio mediante la búsqueda de patrones sobre la base de conocimientos existente, de manera tal que permita realizar análisis tanto estadísticos como topológicos sobre la distribución de los conceptos de la materia, tanto desde el punto de vista del cuerpo docente, representado por las respuestas elaboradas para las evaluaciones, como desde la perspectiva de los alumnos, representados en función de las respuestas a exámenes que los mismos suministran.

Referencias

- [1] Marcelo Marciszack y col. *Material de estudio de la Cátedra Paradigmas de Programación*. Ed. por Cátedra de Paradigmas de Programación Ingeniería en Sistemas de Información. Universidad Tecnológica Nacional, Facultad Regional Córdoba, 2016.
- [2] John F Sowa. "Conceptual graphs for a data base interface". En: *IBM Journal of Research and Development* 20.4 (1976), págs. 336-357.
- [3] W Hartshorne. *Burks, editors. Collected Papers of Charles Sanders Peirce, Cambridge, Massachusetts, 1931-1935*.
- [4] Frank Van Harmelen, Vladimir Lifschitz y Bruce Porter. *Handbook of knowledge representation*. Vol. 1. Elsevier, 2008.
- [5] Ivan Olmos, Jesus A Gonzalez y Mauricio Osorio. "Inexact Graph Matching: A Case of Study." En: *FLAIRS Conference*. 2006, págs. 586-591.
- [6] Marcello Balduccini, Chitta Baral y Yuliya Lierler. "Knowledge representation and question answering". En: *Foundations of Artificial Intelligence* 3 (2008), págs. 779-819.
- [7] Randall Davis, Howard Shrobe y Peter Szolovits. "What is a knowledge representation?" En: *AI magazine* 14.1 (1993), pág. 17.
- [8] M Chantal Pérez Hernández. "Explotación de los corpórea textuales informatizados para la creación de bases de datos terminológicas basadas en el conocimiento". En: *Estudios de lingüística del español* 18 (2002).
- [9] Allan M Collins y M Ross Quillian. "Facilitating retrieval from semantic memory: The effect of repeating part of an inference". En: *Acta Psychologica* 33 (1970), págs. 304-314.
- [10] Stuart C Shapiro y GH Woodmansee. "A Net Structure for Semantic Information Storage, Deduction and Retrieval." En: *IJCAI*. 1971, págs. 512-523.
- [11] John F Sowa. "Semantic Networks". En: *Encyclopedia of Cognitive Science*. John Wiley & Sons, Ltd, 2006. ISBN: 9780470018866.
- [12] R.J. Brachman. "What is-a is and isn't: an analysis of taxonomic links in semantic networks". En: *Computer; (United States)* 10 (oct. de 1983).
- [13] John F Sowa. "Conceptual graphs as a universal knowledge representation". En: *Computers & Mathematics with Applications* 23.2-5 (1992), págs. 75-93.
- [14] Marvin Minsky. "A framework for representing knowledge". En: (1975).
- [15] Maria Alejandra Paz Menvielle y col. "Text format written questions evaluation Methodology". En: *Information Systems and Technologies (CISTI), 2016 11th Iberian Conference on*. IEEE. 2016, págs. 1-4.

Construcción de modelos y simuladores con Squeak-Etoys como recurso de aprendizaje en la escuela media

Ricardo P. Salvador¹, Claudia Pons², Guillermo L. Rodríguez³

¹ Esc. Superior de Comercio "Libertador. Gral. San Martín" (U.N.R.), Balcarce 1240, 2000 Rosario, Santa Fe, República Argentina

² Laboratorio de Investigación y Formación en Informática Avanzada (U.N.L.P.), 50 y 120, La Plata, Buenos Aires, Argentina

³ Fac. de Ciencias Exactas, Ingeniería y Agrimensura (U.N.R.), Av. Pellegrini 250, Rosario, Santa Fe, República Argentina

rsalvado@unr.edu.ar, cpons@info.unlp.edu.ar, guille@fceia.unr.edu.ar

Resumen. En este trabajo de investigación se estudia la influencia en el aprendizaje de cinemática de alumnos de secundaria utilizando la construcción de una simulación en el entorno de programación *Squeak-Etoys*. Al considerar *a)* los efectos de usar simulaciones como recurso educativo, *b)* el surgimiento de nuevos entornos de programación que combinan usabilidad, manipulación de multimedia y programación por bloques, y *c)* la revaloración en el ámbito educativo de las ciencias de la computación y la programación, hizo intuir que construir una simulación sería un recurso útil para el aprendizaje. Se realizó un trabajo experimental con dos grupos de alumnos de secundaria. Uno de los grupos utilizó un tutorial interactivo con indicaciones y recursos para programar la simulación. La experiencia se evaluó mediante un protocolo pre-test / post-test. Los resultados indican que los alumnos aumentaron las calificaciones y se interesaron en la programación como recurso didáctico. La recepción que la actividad tuvo en los alumnos anima a diseñar actividades similares y explorar su uso en otras disciplinas.

Palabras clave: aprendizaje, Física, educación media, simulación, programación.

1 Introducción

El objetivo de este trabajo es indagar la utilidad del diseño y desarrollo de modelos y simulaciones informáticas por parte de los alumnos como recurso de aprendizaje en la escuela media, en una disciplina distinta de las relacionadas a la informática misma, utilizando un entorno de programación amigable *Squeak-Etoys*. El aporte que se busca es proponer una metodología de diseño de actividades en la que los alumnos deban modelar y simular contenidos escolares.

En la sección 2, marco teórico, se explican las posibilidades que brinda la construcción de simulaciones como recurso didáctico, se menciona la Taxonomía de Bloom como marco de referencia para habilidades intelectuales, se recopilan aportes del construccionismo a la enseñanza mediante computadoras, se exponen referencias a la correlación entre programar y el desarrollo de pensamiento computacional, se destacan iniciativas de alcance nacional e internacional promoviendo las ciencias de la computación y la programación, y se caracteriza el entorno de programación *Squeak-Etoys*. El marco metodológico del trabajo de campo, los recursos y los instrumentos utilizados se exponen en la sección 3. En la sección 4, se describen los resultados: las características de la escuela donde se realizó el trabajo de campo, de los grupos de alumnos participantes, de los encuentros que conformaron la actividad experimental y los resultados de los diagnósticos y evaluaciones, previas y posteriores a la misma. En la sección 5 se exponen las conclusiones alcanzadas, y para finalizar, las perspectivas actuales.

2 Marco teórico

2.1 Computadoras, simulación y educación

Frente a tecnologías educativas que ponen al alumno en un rol relativamente pasivo, tales como el libro, el audio y el video, las computadoras como máquinas interactivas aportan la posibilidad de un aprendizaje especialmente activo, tema sobre el que contamos con notables aportes de Seymour Papert¹ y Alan Kay².

Construir una simulación implica *a)* comprensión del sistema que se va a simular, *b)* abstracción de los elementos pertinentes a la modelización, *c)* puesta en juego de habilidades relacionados al pensamiento computacional, *d)* representar el conocimiento e interactuar con esa representación, involucrando el ensayo/error y la puesta a prueba de hipótesis, *e)* interactuar con otros (pares, profesores) en el desarrollo y en la comunicación de la actividad.

Esto es importante ya que el alumno (o grupo) realiza una tarea compleja e integral respecto del objeto de estudio, llevando a cabo un conjunto de habilidades cognitivas

¹ Fundador del Laboratorio de Inteligencia Artificial del Massachusetts Institute of Technology, y del Grupo de Epistemología y Aprendizaje del MIT-MediaLab, creador del lenguaje Logo, asesor en Squeak-Etoys, scratch en el proyecto One Laptop per Child.

² Investigador del Xerox Palo Alto Research Center, ARPANet, Atari, desarrollo de la primera computadora personal gráfica orientada a objetos, SmallTalk, Squeak-SmallTalk, Etoys, Hewlett Packard, presidente del Instituto Viewpoints Research.

que abarca desde las más simples a las de mayor complejidad intelectual, y desde las más pasivas a las que le exigen al alumno más compromiso intelectual.

2.2 Aportes del Construccinismo a la enseñanza mediante computadoras

Seymour Papert creó el concepto de construccionismo, al que define como dar a los niños buenas cosas para hacer de manera que el aprendizaje tenga lugar por el hecho de hacerlas, ya que pensaba que en lugar de sólo interactuar con la tecnología deberían aprender a programar sus propias animaciones, juegos y simulaciones, y en el proceso, aprender habilidades para resolver problemas y estrategias para el diseño de proyectos [1].

Sostiene que el aprendizaje se da indirectamente, al utilizar el conocimiento y no sólo memorizándolo. En otras palabras, la forma en que mejor aprendemos es a través de la acción de construir algo externo a nosotros mismos: construir una torre, escribir un cuento, construir un artefacto robótico, programar un videojuego, o realizar una animación, son ejemplos de construcción. Este tipo de actividades tiene valor respecto de lo que Papert llama “prueba de la realidad”: si no funcionan, son un reto para comprender por qué, y para superar los obstáculos. Pueden compartirse y discutirse con otras personas. Y sirven como objetos de transición para la apropiación personal de estas ideas [2].

Al elaborar el enfoque construccionista, Papert sentó las bases de la enseñanza con tecnologías digitales reconociendo un mundo de cambios cada vez más rápidos, dando un lugar preferencial a los intereses del alumno, para el que los saberes nuevos deben tener sentido actual, valorando la confrontación de modos de pensar diversos y alternativos, dando a la programación un lugar clave, asumiendo que los alumnos pueden aprender a programar desde edades tempranas, donde se presentarán situaciones nuevas aún para el docente, que se transforma en un maestro que aprende a la par que sus alumnos y rescatando la capacidad de simulación de las computadoras como recurso educativo.

2.3 El pensamiento computacional

El pensamiento computacional es la actividad intelectual relativa a la formulación de un problema de manera que pueda admitir una solución por medios computacionales, a la vez que la solución pueda ser llevada a cabo por un ser humano o una máquina, o de ambas maneras [3].

Kafai y Burke [4], señalan que la programación, hoy se reconoce como una respuesta a la necesidad de desarrollar pensamiento computacional en los estudiantes ya que éstos pueden resolver problemas, diseñar sistemas para la vida diaria, y generar progresos en otras disciplinas.

Resnick y Brennan [5], en estudios sobre el diseño de medios interactivos realizado por jóvenes, elaboraron un marco de referencia para el pensamiento computacional y cómo evaluar logros sobre el mismo; elaboraron una lista de siete conceptos computacionales comunes a muchos lenguajes de programación y presentes en el pensamiento computacional: *secuencia*, *ciclo*, *evento*, *paralelismo*, *condicional*, *operador*, y *dato*. También observaron prácticas de diseño relativas a cómo aprendían los participantes de sus estudios: *ser incremental e iterativo*, *ensayar y depuración*, *reusar y remezclar*, *abstraer* y *modularizar*. La conexión entre programación y pensamiento computacional parece corroborarse en otras fuentes consultadas [6] [7].

2.4 Iniciativas promoviendo las Ciencias de la Computación y la Programación en las escuelas

Actualmente se está revalorizando la enseñanza de las Ciencias de la Computación en el ámbito escolar y la enseñanza de la Programación y su uso como recurso educativo y como competencia general en distintos lugares del mundo. Por ejemplo, los proyectos *Raspberry Pi* (<http://www.raspberrypi.org/about/>), *Scratch* (<http://scratch.mit.edu/about/>), *Squeak-Etoys* (<http://squeakland.org/about/>), y *Code.org* (<http://code.org>).

En Argentina, la Ley de Educación Nacional N° 26206 (2006) promueve las competencias relativas las tecnologías de la información y la comunicación (capítulo 11, inciso *m*), y la enseñanza de la “Programación” es declarada de importancia estratégica por la 65° Asamblea del Consejo Federal de Educación (Resolución 263/15, art. 1°). *Program.AR* es una iniciativa de la Fundación Sadosky, surgida en 2012, y su objetivo es llevar el aprendizaje de Ciencias de la Computación a las escuelas (<http://program.ar/quienes-somos/>). *La hora del código* (iniciativa conjunta de *Program.AR* y *Code.org*) ofrece recursos para aprender a programar de forma amena y divertida como una forma de aproximarse a las Ciencias de la Computación.

2.5 Squeak-Etoys: un entorno de programación diseñado para la educación

Squeak-Etoys es un sistema de autoría multimedia especialmente dirigido a aprender ideas construyéndolas. Inspirado en LOGO, *Smalltalk*, *Hypercard* y *Starlogo*, se desarrolló a partir de *Squeak* en Walt Disney Imagineering (la rama de investigación y desarrollo de The Walt Disney Company) y luego fue apoyado por el Viewpoints Research Institute (VPRI) como uno de sus proyectos. *Etoys* presenta un estilo unificado, interfaz de usuario, medios de comunicación y entorno de programación para construir cosas con computadoras [8]. Integrado en el software de las laptops del proyecto One Laptop Per Child, esta aplicación multiplataforma (Linux, Windows y Mac) soporta objetos (texto, imágenes bitmap y vectores, sonidos) que pueden ser editados, sensar eventos, comunicarse con otros objetos, e incluir código de programa. Su simplicidad es proporcional a la posibilidad de obtener resultados de

gran calidad didáctica: facilitando la escritura de código, permite al usuario concentrarse más en el problema que debe resolver al implementar su simulación.

3 Marco metodológico

El trabajo de campo consistió en evaluar el efecto del uso del modelado y la simulación en la construcción de un programa como recurso para el aprendizaje, mediante un protocolo pre-test / post-test en el que participaron un grupo control y un grupo experimental. Se contactó a un docente de Física quien facilitó la participación dos de sus cursos, de 30 y 24 alumnos, en los roles de grupo control y grupo experimental. Los alumnos de este último grupo desarrollaron la simulación sobre un ejercicio de “encuentro” de Movimiento Rectilíneo Uniforme. En vínculo con esto, existen diversas experiencias de investigaciones utilizadas en la enseñanza de la Física [9].

Previamente a la realización del trabajo experimental, se realizó en ambos grupos *a)* un diagnóstico de los grupos control y experimental con el objeto de conocer características del grupo en cuanto al uso de recursos tecnológicos digitales, forma de abordar problemas, conocimiento sobre modelado y opinión de su utilidad en el aprendizaje, y *b)* una encuesta sobre contenidos de Cinemática, con preguntas sobre Movimiento Rectilíneo Uniforme.

La actividad experimental se realizó en el salón donde habitualmente tiene clase el grupo; se utilizaron netbooks del Plan Conectar Igualdad, que incluyen *Squeak-Etoys*. Se previó que el trabajo se hiciera en equipos, especialmente para favorecer la cooperación ante un recurso nuevo.

Se llevó a cabo utilizando un tutorial interactivo diseñado para ofrecer la guía para la actividad y el espacio de trabajo en una misma ventana, de manera que se evitara cambiar de ventanas continuamente. Las consignas y orientaciones para la actividad también se imprimieron para ofrecer una fuente alternativa, por si se experimentaba alguna dificultad en la lectura, o si se deseaba eliminar la guía del tutorial para contar con más espacio en la pantalla (sobre todo considerando que las pantallas de las netbooks son de sólo 10,1 pulgadas).

Debido a que los alumnos debían intervenir modificando y creando guiones y código, se esperaba que surgieran numerosos imprevistos, por lo que la guía fue redactada con la expectativa de que promoviera el trabajo autónomo y que la intervención del profesor estuviera dispuesta sobre todo para la observación del trabajo de los grupos e intervenir en situaciones emergentes.

El tutorial interactivo se desarrolló utilizando un recurso de *Squeak-Etoys* llamado “libro” que permite añadir páginas, en forma similar a una presentación multimedia, con la diferencia de que cada diapositiva también puede contener objetos con sus guiones y se puede intervenir la presentación mientras se está “leyendo”.

Se organizó en una serie de problemas parciales, en grado creciente de complejidad respecto del uso del entorno de programación a la vez que cada etapa progresa hacia la resolución del ejercicio construyendo etapas de la simulación.

4 Resultados

4.1 La Escuela

El trabajo de campo se realizó en la Escuela Superior de Comercio “Libertador General San Martín”, dependiente de la Universidad Nacional de Rosario. La Escuela ofrece la orientación “Comercialización”. Situada en el centro de la ciudad de Rosario, Santa Fe, tiene una población de aproximadamente 1200 alumnos agrupados en comisiones de treinta, que se distribuyen en dos turnos (mañana y tarde), cinco años y cuatro divisiones.

La estructura académica de la Escuela agrupa espacios curriculares afines en Departamentos encabezados por un Jefe y un Adscripto o Coordinador de Área, que coordina cada espacio curricular en particular.

El espacio curricular “Física” está presente de 3° a 5° año, con una carga horaria de 4 horas cátedra semanales; la unidad didáctica correspondiente a “Cinemática” es la primera de 4° año.

La Escuela cuenta con el espacio curricular “Informática” de 1° a 4° año, con una carga horaria de 2 horas cátedra semanales en cada año. Los contenidos no incluyen Programación ni Algoritmia.

4.2 Actividad experimental

4.2.1 Indicaciones previas

Al principio de la primera de las tres clases que duró la actividad, se señalaron los siguientes criterios, aunque se recordaron en forma general y a cada grupo en particular, durante toda la experiencia, y cada vez que se consideró necesario: *a)* se destacó que habría cosas no explicadas, a propósito para que intuyan, imaginen y propongan (y lo pongan en práctica), o para que interactúen efectuando preguntas, *b)* se insistió en que la actividad funciona mejor con una actitud proactiva, por lo que se pidió a los alumnos que prueben, intenten, pongan a prueba este material y sus propias suposiciones, sin esperar hacer todo "correcto"; ni esperar a tener una idea brillante para poner manos a la obra: ¡no se va a romper nada! *c)* también se señaló

que la actividad contiene indicaciones en cada página del tutorial para que cada grupo pueda trabajar a su propio ritmo, y cuando lo necesite, por supuesto, llame al profesor.

4.2.2 Desarrollo en el tiempo

El inicio de la actividad experimental estaba previsto para fines de abril, se esperaba una duración de 2 a 3 módulos de 80 minutos, con una frecuencia semanal, para finalizarla. Por imprevistos que alteraron el calendario escolar, la actividad se desarrolló durante tres clases, el 9 de junio, y el 5 y el 12 de agosto de 2016:

Primera clase

Al comienzo había aún 3 netbooks para desbloquear (aunque todos los días de la semana previa se acudió al curso para desbloquear las netbooks, y solicitárselo a los que aún las tuvieran bloqueadas). Para el desarrollo de la actividad se requería cargar el tutorial interactivo (un archivo de 1,1 Mb), por lo que se previó descargarlo de Internet, o desde dos pendrives, opción que se utilizó ya que no se disponía de WiFi y la señal de telefonía móvil no era suficiente para implementar un punto de acceso desde un celular. Se contaba con 10 netbooks, y la tarea de copiar los tutoriales insumió casi 40 minutos de los 80 de la clase, en parte, debido a que algunas máquinas no reconocían los pendrives (o tardaban mucho tiempo en hacerlo) y otras directamente no tenían la configuración de plataformas original y hubo que instalar *Etoys*.

En los poco más de 40 minutos en que se trabajó en la actividad, los 10 grupos que se formaron trabajaron evidenciando interés, con algunas distracciones pero superando más del 50% de la actividad.

Este encuentro incluyó la presentación de la actividad, sus objetivos, la forma de trabajar y el entorno de programación que se usó, sobre el cual el tutorial incluye indicaciones.

Segunda clase

Se continuó el trabajo utilizando una versión modificada del proyecto, en la que se realizaron correcciones en la redacción.

En esta ocasión, la mayoría de los grupos llegó a la instancia del trabajo en que se simula completamente el comportamiento de los autos “protagonistas” de la simulación. Como los dos autos que intervienen en el ejercicio responden al mismo modelo de comportamiento, una vez terminado uno de los autos, sólo es necesario duplicarlo y modificar su nombre, los parámetros relativos a velocidad, a tiempo, y el aspecto (esto último, para diferenciarlo visualmente del primer auto). Casi todos, a la hora de modificar el aspecto (uno de los autos del ejercicio es azul y el otro es verde) hicieron un paréntesis para cambiar algo más que el color a su nuevo auto.

Tercera clase

La mayoría de los grupos resolvió los desafíos de la penúltima diapositiva, en que simula la situación planteada en el ejercicio; sin embargo no todos terminaron la actividad de la última diapositiva, donde debían obtener y mostrar las respuestas al ejercicio sobre hora y distancia del encuentro entre los dos autos. Al llegar a la mitad

de la clase, se realizó una evaluación del trabajo de cada grupo y al observar que la mayoría encontraba dificultades en los dos últimos desafíos se resolvió explicarlos a todo el grupo utilizando un proyector, para que posteriormente cada grupo intentara terminar su trabajo.

En las clases experimentales se evidenció motivación y se manifestaron inquietudes acerca de cómo cambiar otros aspectos de la simulación, que en principio no estaba previsto como parte de la actividad: *a*) cómo modificar el aspecto (colores, formas) de la simulación, *b*) simular otro tipo de movimiento (realizar cambios de dirección, describir curvas, por ejemplo), *c*) también surgieron consultas sobre la posibilidad de crear simulaciones en otros espacios curriculares, sobre el entorno de programación utilizado y sobre cómo se aprende a programar. Se observó que el aprendizaje de la interface de usuario, al inicio de la experiencia, insumió una cantidad de tiempo que disminuyó en las siguientes clases.

4.3 Resultados de las encuestas

4.3.1 Grupo control

En este grupo, compuesto por un 62% de alumnas y 38% de alumnos, el 70,37% manifestó en el pre-test que había usado software en la escuela para representar situaciones que habitualmente se hacen con lápiz y papel, el 48,15% para resolver problemas y el 18% para construir aplicaciones. El 88,88% afirmó usar tecnología informática para resolver problemas cotidianamente y el 55,56% prefiere resolver problemas en forma grupal.

Frente a la pregunta sobre el concepto de simulación, en el pre-test el 29,63% contestó correctamente, mientras que en el post-test lo hizo el 46,67%. Respecto del concepto de modelo, en el pre-test contestó correctamente el 18,52% y el 40% lo hizo en el post-test. Se observó en el post-test una reducción de la incertidumbre respecto a estos conceptos, de 37,04% a 16,67% y de 40,74% a 6,67% respectivamente. El 77,78%, en el pre-test, contestó afirmativamente acerca de si el uso y/o construcción de modelos y/o simulaciones podría contribuir a sus aprendizajes escolares, y en el post-test lo hizo el 76,67%. En la evaluación de contenidos de Cinemática, en el pre-test, el promedio de respuestas correctas fue de 59,26%, mientras que en el post test, fue de 57,06%.

4.3.2 Grupo experimental

El grupo experimental estuvo compuesto por un 79% de alumnas y 21% de alumnos, de los cuales en el pre-test, el 54,17% manifestó haber usado software en la escuela para representar situaciones que habitualmente se hacen con lápiz y papel, y el 37,5% para resolver problemas. El 66,67% afirmó usar tecnología informática para resolver problemas cotidianamente y el 70,83% prefiere resolver problemas en forma individual.

En cuanto a la pregunta sobre el concepto de modelo, el 25% contestó correctamente en el pre-test, mientras que en el post-test lo hizo el 36,36%. Respecto del concepto de simulación, las respuestas correctas en el pre-test fueron el 33,33%, siendo el 59,59% en el post-test. Se observó en el post-test una reducción de la incertidumbre respecto a estos conceptos, de 70,83% a 33,33% y de 58,33% a 8,33% respectivamente. En la evaluación pre-test, acerca de si el uso y/o construcción de modelos y/o simulaciones podría contribuir a sus aprendizajes escolares, el 54,17 contestó afirmativamente y alcanzando al 72,73% estas respuestas en el post-test. En la evaluación de contenidos de Cinemática, en el pre-test, el promedio de respuestas correctas fue de 56,25%, mientras que en el post test, fue de 60,42%.

5 Conclusiones

La construcción de una simulación mediante *Squeak-Etoys* favoreció el aprendizaje de contenidos de Cinemática, evidenciado en el aumento en las calificaciones de las evaluaciones post-test. Los resultados de la actividad experimental fueron satisfactorios aún tratándose de un grupo heterogéneo.

Programar una simulación se mostró como un recurso didáctico válido aún para alumnos sin conocimiento de programación ni algoritmia. Tras realizar la actividad experimental, los alumnos que comprendían el concepto de simulación aumentó casi tres veces.

Construir la simulación del ejercicio de Cinemática generó motivación en los alumnos, así como interés y expectativas positivas respecto de usar la misma forma de trabajo en otros aprendizajes.

Los alumnos, además de valorar el tipo de actividad que realizaron, demostraron curiosidad por conocer modos de expresarse y poner su sello personal en la actividad usando las herramientas del entorno de programación utilizado.

La actividad presentó numerosos imprevistos y desafíos, lo que significó mayor intensidad en la tarea a la vez que promovió una flexibilidad en la asimetría entre los roles docente y alumno, al encarar juntos esos desafíos.

El trabajo en torno al tutorial interactivo deja una serie de recursos reutilizables en nuevas actividades, y retroalimentación para mejorarlas, reduciendo el tiempo de nuevos desarrollos.

Proponer la experiencia con características lúdicas, desafíos, contenidos significativos para los alumnos y mediatizada por las nuevas tecnología, favoreció el aprendizaje y el interés de los alumnos en los medios utilizados, siendo nuestro desafío actual formalizar su utilización dentro de la escuela.

6 Bibliografía

1. Resnick, M. (2012), "Reviving Papert's Dream", "Educational-technology – the magazine for managers of change in education", Volume 52, Number 4, July/August 2012, <http://web.media.mit.edu/~mres/papers/educationaltechnology2012.pdf>
2. Papert, S. (1999), "Logo Philosophy and Implementation", Logo Computer Systems Inc. Introduction.
3. Wing, J. M. (2010), "Computational Thinking: What and Why?", Scholl of Computer Science, Carnegie Mellon University, <http://www.cs.cmu.edu/~CompThink/resources/TheLinkWing.pdf>
4. Kafai, Y. and Burke, Q. (2014), "Connected Code – Why Children Need to Learn Programming", The MIT Press
5. Resnick, M. and Brennan, K., (2012), "New frameworks for studying and assessing the development of computational thinking", http://web.media.mit.edu/~kbrennan/files/Brennan_Resnick_AERA2012_CT.pdf
6. Lye, S.Y. and Koh, J.H.L. (2014), "Review on teaching and learning of computational thinking through programming: What is next for K-12? ", artículo en "Computers in Human Behavior", Dec2014, Vol. 41, p51-61. 11p. URL: <http://www.sciencedirect.com/science/article/pii/S0747563214004634>
7. Yohan Hwang, Kongju Mun, Yunebae Park (2016), "Study of Perception on Programming and Computational Thinking and Attitude toward Science Learning of High School Students through Software Inquiry Activity: Focus on using Scratch and physical computing materials", Journal of the Korean Association for Science Education, URL: http://koreascience.or.kr/article/ArticleFullRecord.jsp?cn=GHHOBX_2016_v36n2_325
8. Kay, Alan (2007) "Children Learning By Doing – Etoys on the OLPC XO", Viewpoints Research Institute, <http://wiki.laptop.org/images/2/28/OLPCEtoys.pdf>
9. Taub, Armoni, Bagno, Ben-Ari (2015), "The effect of computer science on physics learning in a computational science environment", Computers & Education Volume 87, September 2015, Pages 10–23, URL: <http://www.sciencedirect.com/science/article/pii/S0360131515000913>

Desafiate: Juego Serio para la Autoevaluación de los Alumnos y su Integración con un Entorno Virtual de Enseñanza y Aprendizaje

Federico Héctor Archuby¹, Cecilia Sanz^{1,2}, and Patricia Pesado¹

¹Instituto de Investigación en Informática LIDI
Facultad de Informática, Universidad Nacional de La Plata
La Plata, Argentina

² Investigador Asociado de la Comisión de Investigaciones Científicas de la Prov. de Buenos Aires
{farchuby, csanz, ppesado}@lidi.info.unlp.edu.ar

Abstract. En este trabajo se presenta Desafiate, un juego serio para móviles, que se integra a las posibilidades los entornos virtuales de enseñanza y aprendizaje, a partir de las herramientas de autoevaluación que estos disponen. La propuesta parte de una investigación relacionada con el uso de juegos serios en escenarios educativos. Desafiate ha formado parte de tres experiencias realizadas con estudiantes y docentes, en las que se ha integrado el juego como estrategia de autoevaluación para los estudiantes. Los primeros resultados muestran una aceptación del juego y de la propuesta de autoevaluarse a través de este tipo de estrategias. Al mismo tiempo, se revelan aspectos de mejora tanto en la dinámica del juego como en propuestas que realizan los estudiantes para las historias que se presentan. De igual manera, los estudiantes han manifestado que quisieran contar en otras materias con autoevaluaciones similares, lo que también resulta una muestra de motivación.

Keywords: Juegos Serios, Educación, Autoevaluación, Dispositivos Móviles, Unity 3D, Entornos Virtuales de Enseñanza y Aprendizaje

1 Introducción

Entre los años 90 e inicios del 2000, las universidades se hicieron testigos del avance y proliferación de entornos virtuales de enseñanza y aprendizaje (EVEA). Estos sistemas centrados en la web abrieron el camino para que docentes y alumnos pudieran compartir y vivenciar procesos educativos mediados por estas tecnologías, con la utilización de un conjunto de herramientas, integradas en un mismo espacio y con posibilidad de realizar un seguimiento de las actividades de los alumnos (Sanz, 2015). Actualmente existen muchos EVEA, algunos con gran difusión en la comunidad académica y otros que han sido desarrollos propios de las universidades. Estos entornos permiten complementar la tarea realizada en el aula. Muchos de ellos cuentan con herramientas de autoevaluación con capacidad de generar preguntas que son aprovechadas por numerosos docentes.

Por otra parte, los juegos —y los videojuegos, en particular— han constituido una herramienta para el aprendizaje del comportamiento y de la actitud, y han sido utilizados en la educación desde hace años (Marcano Lárez, 2014). A la hora de categorizar los juegos es útil tener en cuenta su función principal. Así los juegos serios según Michael & Chen (2005) son aquellos que se usan para educar, entrenar e informar (Abt, 1970). En Marcano Lárez (2014) se afirma que los últimos avances en las ciencias cognitivas en general apoyan a los principios del aprendizaje que los juegos incorporan. Destacan el carácter activo y el rol protagónico que adquiere el alumno en la resolución de problemas en tiempo real y con un *feedback* inmediato. Explican que los videojuegos siguen estrategias de diseño para generar ambientes virtuales que atrapen las capacidades perceptivas del usuario, que les produzcan gratificación sensorial (Crawford, 2003) y que generen la sensación de inmersión en la que el jugador “sienta” que puede “participar”. Esta interactividad que ofrecen los videojuegos, logran la atracción, la inmersión, la emoción y motivación que son factores fundamentales en el proceso de aprendizaje de una persona.

No obstante, en Sung & Hwang (2013) se citan varios estudios que demuestran que sin un diseño apropiado el uso de juegos serios podrían no tener el efecto esperado. La parte psicológica del usuario, especialmente la relacionada con la motivación y el disfrute, es un punto importante a tener en cuenta cuando se diseñan juegos, como puede verse en el análisis de diferentes *frameworks* de diseño, en Mora Carreño, Riera, Gonzales, & Arnedo-Moreno (2015).

Dentro de la industria del videojuego, algo que ha ganado mucha fuerza es el área de juegos móviles. Uno de los motivos es el incremento que ha tenido la penetración de los *smartphones* en el mundo. Por lo que en varios ámbitos académicos se está aprovechando esto para el desarrollo de juegos serios educativos. Esto puede verse en la investigación de Baek, Park, & Han (2016) en la que se utilizan tres juegos serios para probar su utilidad a la hora de enseñar sobre conceptos de ciencia, en particular de la física. Dos de estos tres juegos podían usarse en dispositivos móviles. En esta investigación se les pidió a diferentes docentes que evalúen los juegos, y se llegó a la conclusión de que estos eran herramientas potenciales para enseñar sobre conceptos de ciencia, demostrando también el potencial de los juegos serios desarrollados para móviles.

A partir de los antecedentes indagadas, en este trabajo se presenta el diseño de un juego serio para dispositivos móviles llamado Desafiate y su integración con un entorno virtual de enseñanza y aprendizaje (IDEAS) para acompañar a la autoevaluación del estudiante en la aplicación de conceptos, recuperación de conocimientos abordados en un determinado proceso educativo, e integración de temáticas, entre otros.

De aquí en más este artículo se organiza de la siguiente manera: en la sección 2 se presentan antecedentes relevantes para el trabajo realizado; en la sección 3 se describe el diseño de Desafiate, para luego presentar en la sección 4 aspectos de su implementación; en la sección 5, se explican las experiencias ya realizadas utilizando el juego y los resultados obtenidos. Finalmente, la sección 6, detalla las conclusiones y trabajos futuros.

2 Antecedentes

Como se dijo anteriormente, a lo largo de los años se han analizado las ventajas que traen consigo los juegos serios. Una de las más analizadas en las investigaciones suele ser el impacto que tienen en la generación de conocimiento del alumno. Se han utilizado juegos serios para el aprendizaje de las personas incluso en ámbitos no académicos. Un ejemplo de esto es la investigación realizada por Chittaro & Buttussi (2015), en donde se utiliza un juego serio para enseñar sobre las medidas de seguridad a tomar en un accidente de avión y un aterrizaje forzoso, logrando de esta forma conseguir uno de los objetivos, el cual era, que este aprendizaje sea más duradero en el tiempo que el conseguido mediante un método tradicional.

Un ejemplo del uso de un juego serio para mejorar el aprendizaje en el ámbito académico es *It's a Deal!* desarrollado para la investigación realizada por Guillén-Nieto & Aleson-Carbonell (2012). Este juego fue creado para enseñar sobre la comunicación intercultural en un ambiente de negocios. En la investigación que realizaron se buscó analizar el impacto de los juegos serios en el aprendizaje. Para ello, se utilizó un proceso que consistía en comparar los resultados arrojados por dos pruebas de tipo cuestionario, una anterior y otra posterior al uso del juego. La investigación arrojó resultados positivos, ya que se notaron mejoras estadísticamente significativas en las tres variables que se analizaron.

Esta hipótesis es reforzada por la investigación realizada por Giannakos (2013). Aquí se desarrolló un juego serio cuya finalidad era la de enseñar sobre la suma y resta matemáticas. En este contexto se realizaron dos estudios diferentes, uno con la intención de demostrar que los juegos serios son igual o mejores a la hora de generar conocimiento con respecto a estrategias de aprendizaje más tradicionales, y otro para analizar si el disfrute, la intención de uso, y la sensación de felicidad al usarlo tienen algún impacto en el rendimiento a la hora de generar conocimiento.

Si bien pocos autores se enfocan en analizar el posible impacto de los juegos serios en la evaluación, existen investigaciones interesantes que intentan explorar esta relación. Una de ellas es la realizada por Kickmeier-Rust & Buttussi (2014) en la cual se desarrolla un juego llamado 1x1 Ninja con el objetivo de que el alumno pueda practicar las multiplicaciones. Lo destacable de esta investigación es que utiliza al juego no solo como una herramienta para enseñar y poner en práctica la multiplicación matemática, sino que también lo utiliza como una herramienta para la evaluación formativa. En esta investigación se analizaron diferentes tipos de *feedback*, incluido uno de tipo formativo, el cual también incluyó características adaptativas que consistían en acomodar la dificultad de las multiplicaciones realizadas en base a la *performance* del jugador, o de presentarle retos con los números con los que el jugador tuvo mayor cantidad de problemas. Los resultados arrojaron que este tipo de *feedback* es más efectivo que otros más tradicionales como *feedback* visuales o auditivos simples, y que el juego desarrollado puede usarse como una herramienta para la evaluación formativa.

El estudio realizado por Kiili, Devlin, Perttula, Tuomi, & Lindstedt (2015) también es interesante con respecto a la evaluación ya que utiliza al juego serio

Semideus para poder evaluar los conocimientos sobre números racionales de los alumnos. Para validar los aportes, los participantes realizaron una prueba escrita, y luego de una semana utilizaron el juego cuyos desafíos son de un formato similar a los que se encontraban en la prueba escrita. Los resultados arrojaron datos similares en ambas pruebas lo que marca una correlación estadística.

Los antecedentes aquí presentados han servido para la toma de decisiones en el diseño de Desafiate, que se presenta a continuación.

3 Diseño de Desafiate

Desafiate es un juego serio para dispositivos móviles orientado a la autoevaluación de los alumnos. Se trata de un juego de preguntas y respuestas, que tiene la particularidad de tomar dichas preguntas de las autoevaluaciones disponibles para el alumno en un EVEA, en particular, en la actualidad se lo ha integrado con el EVEA IDEAS¹. El alumno personifica a un pirata que se encuentra en busca de tesoros y por esto decide salir en busca de diferentes aventuras. Cada aventura ocurre en un archipiélago de islas, las cuales tienen la particularidad de que están habitadas por diferentes tipos de personajes. Estos personajes encontraron los tesoros deseados por el protagonista y sólo a cambio de información cederán el preciado bien. En el puerto de cada isla, un habitante le hará una pregunta de la autoevaluación al protagonista, y solo le dará el cofre si éste responde bien. El juego se sitúa en el Caribe, lugar relacionado tradicionalmente con historias de piratas.

La relación que se pensó entre los EVEA y Desafiate, hizo necesario que cada una de las partes que componen las autoevaluaciones de los EVEA, encuentren su contraparte en Desafiate. De esta forma, se planificó que cada una de las autoevaluaciones disponibles para el alumno, representen una aventura distinta para el jugador. A su vez, cada una de las preguntas que componen una autoevaluación está representada por el viaje que realiza el protagonista a una isla en particular. Por último, la cantidad de desafíos que el jugador superó es lo que determina el puntaje total que el alumno ha conseguido en la autoevaluación. En la Fig. 1, se pueden ver estas relaciones tomando como referencia al EVEA IDEAS.

Actualmente Desafiate está compuesto de cuatro escenas principales que serán explicadas a continuación: la pantalla de inicio, el menú principal, la resolución de preguntas y la visualización de los resultados.

La pantalla de inicio de sesión tiene la intención de introducir la estética y el ambiente del juego. Toda la vista está completamente llena de árboles para dar la sensación de encontrarse en una selva además de poseer detalles pertinentes al juego como algunos cocos, o un cofre del tesoro cerrado. Lo principal de esta pantalla es el menú que muestra el nombre del juego y proporciona los *inputs*

¹ Este EVEA es una evolución surgida del proceso de reingeniería al que fue expuesto el EVEA WebUnlp, el cual fue desarrollado en la Facultad de Informática de la Universidad Nacional de La Plata, al igual que IDEAS. Se lanzó su primera versión al público general en Febrero de 2017. <https://ideas.info.unlp.edu.ar/>



Fig. 1. Se describe en el esquema la relación que existe entre las partes de la herramienta de autoevaluación en el EVEA IDEAS y sus contrapartes en Desafiate.

necesarios para poder iniciar sesión en el sistema (Fig. 2). El diseño del menú ha sido creado para adaptarse a la temática del juego. Es por esto que los bordes han sido diseñados como cañas de azúcar atadas con hilos. Esta visual se mantendrá a lo largo de juego en todas las pantallas, incluyendo a las ventanas modales que sirven para brindar información al jugador.

En el menú principal se puede visualizar una playa con un muelle en el cual se encuentra el barco con el que se realizará la aventura. Además se puede ver el listado de autoevaluaciones que es proporcionado por el servidor del EVEA, en el orden que éste determine. Cada autoevaluación se representa con un mapa que indica que se trata de una aventura diferente. Accediendo a cada mapa aparece la opción de iniciar dicha aventura y la información pertinente a cada autoevaluación: nombre de la autoevaluación, la cantidad de preguntas que la componen (desafíos), el estado de ésta (si ya fue realizada), y la calificación si corresponde (ver Fig. 3).

La resolución de las preguntas ocurre a medida que el personaje recorre las diferentes islas que se corresponden con las preguntas de la autoevaluación. El tipo de isla donde ocurrirá la acción es seleccionada de forma aleatoria en base a unos modelos ya creados. La ambientación y estética de la isla varían entre cada uno de los escenarios. Cada isla tiene una serie de elementos que la hacen única con respecto a las demás. Además cada isla posee un conjunto de historias predeterminadas, y se elige en forma aleatoria cuál será la historia a mostrar. En la Fig. 4 se puede ver el ejemplo de unas de las islas.

La acción en esta pantalla ocurre a través de varias etapas. Se comienza visualizando al barco llegar a la isla, para reunirse con un personaje que contará



Fig. 2. Diseño del menú de inicio de sesión.



Fig. 3. Diseño del inicio de la aventura.

la historia que ha sido seleccionada. A continuación sigue la resolución del desafío, como puede verse en la Fig. 5 (actualmente solo existen los desafíos de tipo *multiple choice* y verdadero y falso), donde una vez terminado se podrá ver un *feedback* del resultado junto a un texto creado previamente por el docente para mostrarse como retroalimentación de la pregunta. Por último, mientras el barco abandona la isla, el juego guardará los datos correspondientes a la resolución de la pregunta y esto también se verá reflejado en el EVEA.



Fig. 4. Ejemplo de una de las posibles islas.

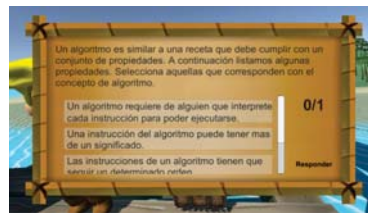


Fig. 5. Diseño de la interfaz para responder preguntas.

Cuando el jugador ya ha respondido las preguntas se pasa a la escena de fin de la aventura donde se muestra al protagonista llegar al puerto donde empezó todo. Aquí se visualizará un cofre que simboliza las ganancias obtenidas a lo largo de la aventura, además de poder ver un listado de todas las preguntas incluidas como desafíos que se le fueron haciendo al protagonista, en el orden que las fue contestando. Para cada pregunta se lista la cantidad de monedas que el jugador ha obtenido en relación al máximo que se podía obtener.

4 Aspectos de Implementación de Desafiate

Para la realización de Desafiate se analizaron distintos motores de juego en base a requisitos deseables para poder elegir el motor adecuado. De este análisis surge la decisión de utilizar el motor *Unity*, el cual es actualmente uno de los motores más utilizados por desarrolladores de juegos. Al estar enfocado al público general

desde su creación, la comunidad que utiliza el motor fue creciendo rápidamente, con lo cual, cuenta actualmente con la comunidad más grande entre los motores de juego.

Debido a que Desafiante está pensado para usarse en conjunto con un EVEA que puede ser intercambiable según el requisito, se decidió dividir la arquitectura del juego en tres diferentes capas, para de esta forma, poder cambiar o reemplazar estas capas sin que las demás se vean afectadas. La Fig. 6 muestra como está compuesta esta arquitectura usando al EVEA IDEAS como ejemplo.

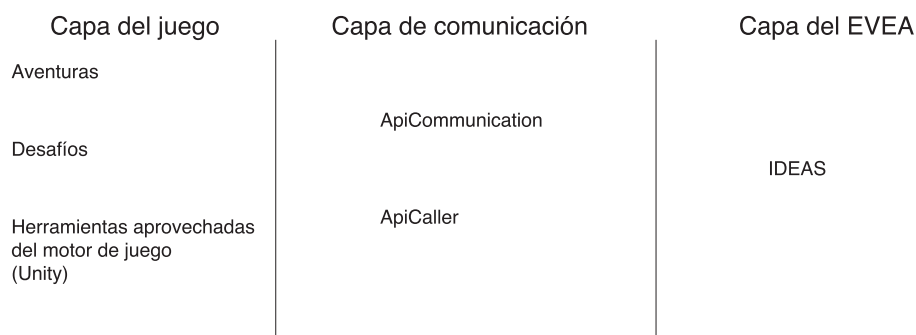


Fig. 6. Las diferentes capas de la arquitectura y sus componentes.

La capa del juego representa los elementos que se utilizan para el funcionamiento interno del juego. Abarca el motor de juego elegido, los assets externos a utilizar, y los componentes desarrollados. La capa del EVEA hace referencia al entorno relacionado con el que se hará toda la comunicación y del cual se obtendrán los datos necesarios para la realización de las aventuras. Todo lo referente al funcionamiento de esta capa excede del control del desarrollo de Desafiante. Por último, la capa de comunicación es la que se encarga de poder comunicar al EVEA con Desafiante. Fue desarrollada especialmente con la idea de poder configurarse y extenderse fácilmente en el futuro, para hacer frente a un posible cambio de EVEA con el cual comunicarse.

Dentro de la capa del juego, se realizaron diferentes interfaces y jerarquías siguiendo patrones de diseño de objetos para facilitar el desarrollo y la extensión de Desafiante. Una de las partes centrales de esta capa es la resolución de desafíos. Realizar la programación necesaria para cada isla sería un trabajo tedioso y largo, es por eso que se buscaron puntos en común para lograr generar una estructura genérica a la que solo haga falta cambiar algunos atributos y de esa forma simplificar la creación de las diferentes islas.

Para lograr esto, se creó una jerarquía que representa el comportamiento del barco en las diferentes etapas en las que incurre: llegar al muelle de la isla, enfocar al isleño, resolver el desafío, irse de la isla y guardar los datos pertinentes. Incluso algunas de estas etapas coinciden con las llevadas a cabo por la escena

de visualización de resultados, con lo que se resolvió agregar el comportamiento propio de esta escena a la jerarquía.

5 Primeras Experiencias con Desafiate

Con el fin de analizar a Desafiate se llevaron a cabo una serie de pruebas, de manera tal de comprobar qué aspectos de Desafiate le resultan interesantes al alumno y cuáles son los aspectos que le gustaría mejorar o agregar para que la experiencia sea más disfrutable. Al mismo tiempo, se buscó conocer la opinión de estudiantes y docentes en relación a los juegos como estrategia de evaluación, y como parte de los procesos de enseñar y aprender, en general. Las experiencias realizadas han aportado también a la mejora de la usabilidad del juego, ya que se recogieron datos sobre las dificultades encontradas a lo largo de las sesiones, las cuales fueron corregidas para no dificultar las sesiones siguientes.

Se realizaron tres diferentes sesiones, dos de ellas con alumnos del curso de ingreso de la Facultad de Informática de la UNLP, y la restante con alumnos de primer año de la carrera de Ingeniería en Computación en la misma institución. Para cada sesión se crearon junto a los docentes autoevaluaciones que abordaron los temas vistos por los alumnos en clase. Para poder usar Desafiate, los alumnos tenían la posibilidad de descargar la aplicación de internet, o utilizar *tablets* preparadas para la ocasión. Luego de realizada la aventura en el juego, se les entregó a los participantes un cuestionario, para poder obtener sus opiniones acorde a los objetivos planteados para las experiencias con anterioridad.

En total participaron 23 alumnos de entre 17 y 37 años, concentrándose un 74% de ellos entre los 17 y 20 años. Un 82% de los participantes eran de sexo masculino y un 74% de ellos provenían de escuela de gestión pública. Todos los participantes afirman utilizar recursos digitales como apoyo para el estudio.

Con respecto a la motivación se les consultó a los participantes qué tan de acuerdo se encontraban con respecto a diferentes afirmaciones relacionadas a la motivación al trabajar con Desafiate, y en relación a los juegos serios y su utilización en escenarios educativos, en particular para su autoevaluación. La escala utilizada fue: Muy de acuerdo, De acuerdo, Indeciso, En desacuerdo, Muy en desacuerdo. El cuadro 1 refleja algunos de los diferentes tópicos consultados y los resultados de las respuestas obtenidas.

En general, se visualizan resultados positivos, que al ser ampliados en diálogo con los alumnos durante las sesiones, muestran un fuerte interés en la propuesta, algo que refuerza lo dicho anteriormente en la sección de antecedentes en cuanto a la motivación acerca del uso de juegos serios en educación y en particular en lo referente a la autoevaluación.

Cuando se les consultó a los participantes sobre cuáles fueron los elementos que más y menos les gustaron, 22 participantes estuvieron de acuerdo en que la idea de usar un juego para autoevaluarse resulta interesante. Otro punto a favor resaltado por 14 de los participantes fue el uso de los dispositivos móviles como plataforma. Por el lado contrario, 9 de los participantes opinó que la dinámica

Table 1. Tabla de resultados de las encuestas realizada en las sesiones.

Tópicos	Muy de acuerdo	De acuerdo	Indeciso	En desacuerdo	Muy en desacuerdo
1- El uso de juegos me resulta un complemento interesante para mi aprendizaje.	6	15	1	1	0
2- El uso de juegos para autoevaluarme me resulta motivante.	11	8	4	0	0
3- El uso de Desafiate resultó estimulante para continuar aprendiendo sobre la temática de la materia.	8	9	4	1	1
4- El uso de Desafiate me resultó menos estresante que una autoevaluación en lápiz y papel.	14	6	2	1	0

de juego es algo a ser mejorado a partir de la inclusión de nuevas historias en las islas.

Cuando se les consulto a los participantes sobre qué cuestiones les gustaría ver en Desafiate en sus próximas versiones, 17 de ellos aportaron que la idea de poder desafiar a un compañero sería interesante en el futuro. A su vez, otras cuestiones en las que por lo menos 10 participantes estuvieron de acuerdo en incorporar son: el agregado de sonido y música al juego, el recibir un mejor *feedback* a la hora de contestar las preguntas, y poder puntuar la dificultad de las preguntas.

6 Conclusión

En este trabajo se ha presentado parte de un proceso de investigación relacionado con los juegos serios, y su utilización en escenarios educativos. Más particularmente, se pone el foco en juegos serios en móviles y sus posibilidades para la autoevaluación de los estudiantes. Como parte de este proceso, se diseñó y desarrolló Desafiate, un juego serio, que se integró a un EVEA. Las primeras experiencias realizadas con Desafiate, reafirman varios de los antecedentes indagados, reforzando la idea de que los estudiantes se motivan con el uso de este tipo de estrategias en sus procesos de aprendizaje, y en particular, en la autoevaluación. Desafiate ha resultado de interés para los estudiantes y también para los docentes que participaron de las experiencias. Todos indicaron que volverían a utilizarlo y señalaron la necesidad de agregar aspectos de competencia entre los participantes. Al mismo tiempo, un aspecto que surgió se vincula con la posibilidad de desafiar algún compañero con una pregunta de una aventura del juego (autoevaluación). Esto marca también un aspecto de motivación por parte de los participantes, que han aportado ideas no sólo para la dinámica del juego sino también para las historias y en relación a la usabilidad.

References

- Abt, C. (1970). *Serious games*. Viking Press.
- Baek, S., Park, J.-Y., & Han, J. (2016). Teacher assessment of simulation-based serious games for science education. *International Journal of Serious Games*, 3(3). doi: 10.17083/ijsg.v3i3.123
- Boyle, E., Connolly, T. M., & Hainey, T. (2011). The role of psychology in understanding the impact of computer games. *Entertainment Computing*, 2, 69-74.
- Chittaro, L., & Buttussi, F. (2015, April). Assessing knowledge retention of an immersive serious game vs. a traditional education method in aviation safety. *IEEE Transactions on Visualization and Computer Graphics*, 21(4), 529-538. doi: 10.1109/TVCG.2015.23918532
- Crawford, C. (2003). *On game design*. New Riders.
- Giannakos, M. (2013). Enjoy and learn with educational games: Examining factors affecting learning performance. *Computers & Education*, 68, 429 - 439.
- Guillén-Nieto, V., & Aleson-Carbonell, M. (2012). Serious games and learning effectiveness: The case of it's a deal! *Computers & Education*, 58(1), 435 - 448. doi: <http://dx.doi.org/10.1016/j.compedu.2011.07.015>
- Kickmeier-Rust, M. D., & Buttussi, F. (2014). Gamification and smart, competence-centered feedback: Promising experiences in the classroom. *International Journal of Serious Games*, 1(1). doi: 10.17083/ijsg.v1i1.7
- Kiili, K., Devlin, K., Perttula, A., Tuomi, P., & Lindstedt, A. (2015). Using video games to combine learning and assessment in mathematics education. *International Journal of Serious Games*, 2(4). doi: 10.17083/ijsg.v2i4.98
- Marcano Lárez, B. E. (2014). *Factores emocionales en el diseño y la ejecución de videojuegos y su valor formativo en la sociedad digital.: El caso de los videojuegos bélicos* (1st ed.). Ediciones Universidad de Salamanca.
- Michael, D., & Chen, S. (2005). *Serious games: Games that educate, train, and inform*. Thomson Course Technology.
- Mora Carreño, A., Riera, D., Gonzales, C. S., & Arnedo-Moreno, J. (2015). A literature review of gamification design frameworks. In *7th international conference on games and virtual worlds for serious applications (vs-games)*.
- Sanz, C. V. (2015, Junio). Los objetos de aprendizaje, un debate abierto y necesario. *Bit & Byte*, 1(1), 33-35.
- Sung, H.-Y., & Hwang, G.-J. (2013). A collaborative game-based learning approach to improving students' learning performance in science courses. In *Computers & education* (Vol. 63, p. 43-51).

Estimación del nivel de habilidad en sistemas tutores inteligentes utilizando una metodología multiatributo

Sonia G. Sosa-León^{1,2}, Julio Waissman², Jose A. Olivas¹ y Manuel E. Prieto¹

¹ Departamento de Tecnologías y Sistemas de Información. Universidad de Castilla-La Mancha. Ciudad Real, España

² Departamento de Matemáticas. Universidad de Sonora. Hermosillo, México
sonia@mat.uson.mx

Resumen. Para el funcionamiento ideal de un sistema tutor inteligente es indispensable poder estimar el nivel de habilidad de los estudiantes de acuerdo a objetivos complejos de aprendizaje. En este trabajo se propone una arquitectura para la evaluación del nivel de habilidad del estudiante, basada en la teoría de la utilidad multiatributo, utilizando como operador de agregación a la integral de Choquet. El método toma en cuenta los objetivos de aprendizaje planteados por el tomador de decisiones (académicos, representantes instituciones, etc.) representados por relaciones complejas que se pueden dar entre los criterios considerados para la evaluación.

Palabra Clave: Sistemas tutores inteligentes, MAUT, Integral de Choquet.

1 Introducción

Los sistemas tutores inteligentes (ITS por sus siglas en inglés) dan un complemento a la educación formal, permitiendo a los estudiantes reforzar sus conocimientos a su ritmo, con libertad de horarios y de lugar de trabajo. La personalización de las actividades propuestas por el ITS es un factor clave para su aprovechamiento [1]. En [2] se señalan tres problemas principales en la personalización de actividades: la estimación de la dificultad de los ejercicios, el algoritmo de personalización y la estimación del nivel de habilidad del estudiante. En este trabajo se propone un método para la estimación del nivel de habilidad del estudiante, para su aplicación en la plataforma Sofía XT.

Sofía XT es una plataforma desarrollada e implementada en México como complemento a la enseñanza clásica, buscando mejorar las habilidades matemáticas y fomentar el interés en ellas para los grados de enseñanza básica. Actualmente se utiliza en diferentes instituciones, tanto públicas como privadas [3]. El desempeño de Sofía XT depende en gran medida de la medición de la evolución del estudiante, por lo que la estimación del nivel de habilidad del estudiante es fundamental.

Establecer los criterios que permitan evaluar el nivel de habilidad de los estudiantes es una tarea compleja que requiere el manejo de niveles de abstracción elevados [4]. Típicamente este cálculo se realiza utilizando el modelo de Rasch [5] o una versión modificada de éste [6], sin embargo estos métodos no se ajustan a los criterios particu-

lares de cada institución. Por ejemplo, en educación matemática hay estudios que muestran que la ansiedad matemática puede causar bajos desempeños [7] mientras que otros sustentan un punto de vista opuesto en el cual, es el bajo desempeño matemático lo que causa la ansiedad matemática [8].

La teoría de la utilidad multiatributo (MAUT) es una metodología adecuada para asignar un valor global a una situación que puede ser observada desde diferentes puntos de vista, tomando en cuenta, además, que dichos puntos de vista pueden ser de muy diferente naturaleza [9].

La contribución principal de este trabajo es la propuesta de una arquitectura para la evaluación del nivel de habilidad de los estudiantes después de realizar actividades en Sofia XT, en el cual los tomadores de decisión (académico, responsables institucionales o escolares, ...) puedan expresar las relaciones complejas por las cuales medir el nivel de habilidad de acuerdo a diferentes criterios.

El artículo se organiza de la siguiente manera: En la siguiente sección se presenta la idea general de la metodología MAUT de ayuda a la decisión. En la sección 3 se presenta la operación general de la plataforma y la arquitectura del sistema de evaluación del nivel de habilidad propuesto. En las secciones 4 y 5 se presentan los operadores propuestos para el cálculo de los indicadores de desempeño marginales y globales respectivamente. Por último, se presentan las conclusiones, así como los trabajos futuros.

2 Metodología MAUT

Sea $X \subseteq X_1 \times X_2 \times \dots \times X_n$, $n \geq 2$, un conjunto de situaciones de interés descrito por un conjunto $N = \{1, 2, \dots, n\}$ de atributos, los cuales se denominan índices. La intención, dentro de la MAUT, es modelar las preferencias de un tomador de decisión (DM) representada por un operador binario \succsim en X , por medio de una función global de utilidad $U: X \rightarrow [0,1]$ tal que

$$x \succsim y \Leftrightarrow U(x) \geq U(y), \quad \forall x, y \in X.$$

La función U es generalmente determinada a partir de la información proveniente del DM mediante la cual expresa sus preferencias sobre un subconjunto relativamente pequeño y prototípico de situaciones y de atributos. La función global de utilidad U es entonces una representación numérica de la relación \succsim en X y puede utilizarse como un modelo del conocimiento experto del DM.

El operador binario de preferencia \succsim es una relación completa y transitiva. El modelo MAUT más general es el llamado modelo transitivo decomposable, propuesto por Krantz [7]. En este, U se obtiene por

$$U(x) = g(u_1(x_1), u_2(x_2), \dots, u_n(x_n)), \quad \forall x = (x_1, x_2, \dots, x_n) \in X,$$

donde las funciones $u_i: X_i \rightarrow [0,1]$ se conocen como funciones marginales de utilidad, y la función no decreciente $g: [0,1]^n \rightarrow [0,1]$ es llamada operador de agregación.

La capacidad de descomponer el proceso de calificación en dos etapas depende de la selección de las funciones u_i las cuales deben de ser conmensurables. Dos funciones

de utilidad marginal son conmensurables si los criterios son similares en cuanto a magnitudes, esto es que, de acuerdo al tomador de decisiones, los atributos i y j cumplen en el mismo grado con sus criterios marginales.

3 Arquitectura propuesta

La plataforma Sofia XT funciona en términos generales de la siguiente manera:

1. En base a la información de registro, al estudiante se le asigna un nivel de habilidad por cada concepto y en base a su nivel, el sistema habilita los conceptos que el estudiante puede seleccionar.
2. Una vez que el estudiante selecciona un concepto, se genera una actividad, que consiste en 5 ejercicios en orden creciente de dificultad, a partir del conjunto de ejercicios asociados al concepto.
3. El estudiante realiza la actividad y se reevalúa el nivel de habilidad del estudiante sobre el concepto y se asignan recompensas para el estudiante, en relación a su progreso.
 - a. Si el estudiante no cumple con un nivel de habilidad mínimo, se le presenta la opción de seleccionar uno de los conceptos antecedentes al actual.
 - b. En caso que el estudiante tenga un nivel de habilidad alto en este concepto, se le presenta una selección de conceptos adecuados a su nivel.
 - c. En otro caso se mantiene la misma selección de conceptos.

Un método de estímulos y recompensas busca guiar al estudiante, al mismo tiempo que le brinda la libertad de tomar decisiones sobre su aprendizaje, convirtiéndolo en un sujeto activo en la toma de decisión. Este método de estímulos funciona basado en un avatar, y la otorgación de créditos para que los estudiantes personalicen su avatar. De acuerdo a lo observado por los profesores, el sistema de créditos ha estimulado en los estudiantes el uso de la plataforma.

La base para la operación correcta de la plataforma es la estimación del nivel de habilidad del estudiante, la cual se realiza tomando en cuenta el nivel de habilidad anterior y la evaluación de la última actividad. En dicha evaluación es donde se requiere expresar los objetivos de aprendizaje.

En la Fig. 1 se presenta un esquema con la adaptación del método MAUT, propuesto para la calificación de las actividades resueltas por estudiantes. Una vez que un estudiante resuelve una actividad, se genera un vector de cinco entradas $X_A = [x_1, x_2, x_3, x_4, x_5]$, las cuales a su vez son estructuras complejas de datos. La estructura de la entrada incluye como información, el identificador del estudiante que la realizó $e(x_k)$, la dificultad del problema $d(x_k)$, el tiempo en segundos que tardó en realizar el problema $t(x_k)$ y los intentos que requirió el estudiante para realizar el problema $i(x_k)$, además de la información genérica de la actividad. Esta información se procesa para obtener dos conjuntos de datos: por una parte, se obtiene un conjunto de atributos específicos a la actividad realizada, los cuales servirán para la calificación como atributos en un proceso MAUT; por otra, un conjunto complementario de datos (concepto, grado, área de conocimiento), para establecer los criterios de los indicadores de desempeño

marginal. Los atributos específicos a la actividad son la dificultad, el tiempo utilizado y el número de intentos realizados por cada uno de los ejercicios, definidos por $\bar{X}_k = (d(x_1), t(x_1), i(x_1), \dots, d(x_5), t(x_5), i(x_5))$.

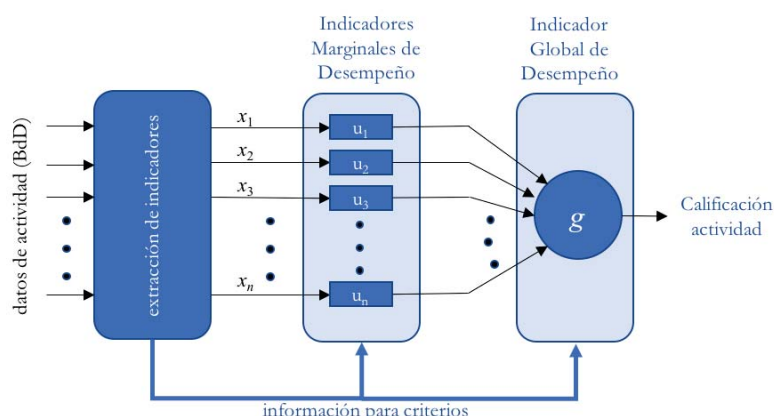


Fig. 1. Arquitectura tipo MAUT de evaluación de actividades

Estos atributos son llevados a las diferentes funciones u_i las cuales se designan como indicadores marginales de desempeño (MPI). Los MPI se pueden tratar y analizar como funciones de pertenencias, las cuales definen 15 subconjuntos borrosos. Si bien, se calcula cada uno independiente de las otras mediciones de la actividad, debido al requerimiento de conmesurabilidad de índices en el proceso MAUT, todos los MPI similares serán calculados de la misma forma. Así, los MPI se calculan como funciones de pertenencia difusas parametrizadas a partir del conjunto complementario de datos.

Una vez que los datos se han tratado, se obtiene un vector de indicadores, los cuales se van a utilizar para generar un indicador global de desempeño. Los operadores de agregación g son funciones globalmente estrictamente crecientes, tales que $g(0, \dots, 0) = 0$ y $g(1, \dots, 1) = 1$. En las secciones siguientes se desarrollan los indicadores marginales y global de desempeño propuestos.

4 Indicadores marginales de desempeño

Se desarrollaron 3 tipos básicos de funciones de pertenencia borrosas: la dificultad, el tiempo utilizado para contestar el reactivo y el número de intentos realizado.

El grado de dificultad de cada ejercicio se calcula en forma estadística, modelando el grado de dificultad de los reactivos como una variable aleatoria normal $\mathcal{N}(0,1)$, donde $d(x_k) = 0$ es la dificultad media de los ejercicios. A partir de esta información, el MPI respecto a la dificultad de un reactivo se puede realizar como una aproximación de la función de probabilidad dada por

$$u_{3k-2}(d(x_k)) = \frac{1}{2} \left(1 + \text{sign}(d(x_k)) (1 - e^{-2d(x_k)/\pi})^{1/2} \right).$$

Para el cálculo de los MPI correspondientes al tiempo que le tomó al estudiante realizar el reactivo, es necesario parametrizar en relación a los datos complementarios (el concepto, subtema y grado), debido a que estos difieren entre sí. Como podemos observar en el ejemplo de la Fig. 2, la distribución de los tiempos presenta una regularidad entre todos los conceptos, variando su valor medio en función del nivel de dificultad del concepto.

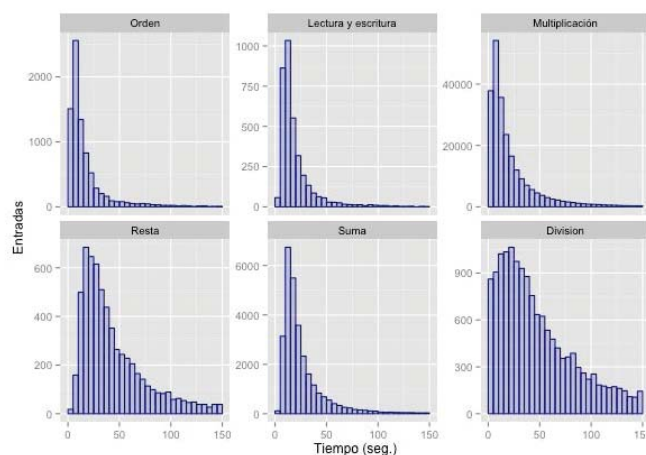


Fig. 2. Distribución del tiempo de resolución del subtema *Números naturales de cuatro cifras*

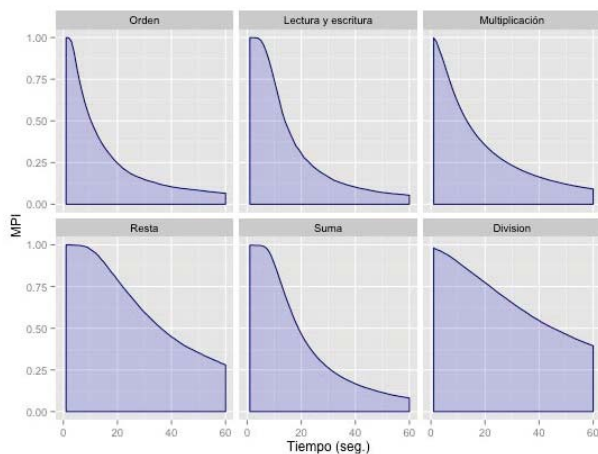


Fig. 3. Índices marginales de desempeño en tiempo para el subtema *Números naturales de cuatro cifras*

Sea φ_t^C el número de ocurrencias en que el conjunto de entradas históricas pertenecientes al concepto C tuvo un intervalo de tiempo de resolución t , en segundos. El cálculo de los MPI de tiempo se realiza mediante

$$u_{3k-1}(t(x_{i,k})) = 1 - \frac{\sum_{k=1}^{|t(x_k)|} \varphi_k^C}{\sum_{\forall t} \varphi_t^C}.$$

Si bien el índice propuesto solo contempla tiempos en una granularidad de segundos, esta parece ser una escala de detalle suficientemente suave, tal como se aprecia en la Fig. 3.

En la Fig. 4 podemos ver el número de intentos de los 6 conceptos que conforman el subtema *Números naturales de cuatro cifras* de tercer grado. La gran mayoría de las entradas se resolvió en el primer intento como se puede observar. Otro comportamiento interesante de observar es el número de entradas que registran 0 intentos. En la plataforma Sofia XT los estudiantes pueden decidir saltar un ejercicio sin resolverlo, a pesar de las ayudas ofrecidas. Cada vez que un estudiante salta un ejercicio, su número de intentos se marca como 0. Si bien en algunos conceptos revisados existe un número mayor que 4 de intentos significativos, estos nunca (en todos los conceptos de la plataforma) son mayores a 10.

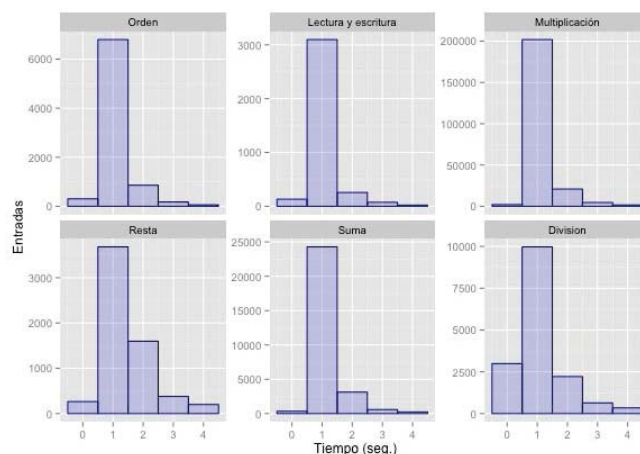


Fig. 4. Distribución de intentos del subtema *Números naturales de cuatro cifras*

Sea θ_i^C el número de entradas del concepto C que se registró con i intentos, entonces el índice de desempeño marginal se define como

$$u_{3k}(i(x_k)) = \begin{cases} i(x_k), & \text{si } i(x_k) < 2, \\ \left(\frac{\theta_{i(x_k)}^C}{\sum_{l=0}^{10} \theta_l^C} \right)^{1/r}, & \text{en otro caso,} \end{cases}$$

donde $r \geq 1$ es una constante de autoestima. Entre mayor sea el valor de r , menor será la exigencia del entorno de aprendizaje para que los estudiantes resuelvan los ejercicios al primer intento. Como se observa en la función, solamente si se resuelve el ejercicio

en el primer intento, el valor del índice será de 1 y solamente si el estudiante decidió no resolver el ejercicio se asignará el valor de 0.

5 Indicador global de desempeño

Una vez procesada la información de la actividad por los MPI se obtienen quince valores de desempeño marginal. Éstos se arreglan de la forma siguiente:

$$u = (d_1, t_1, i_1, \dots, d_5, t_5, i_5),$$

donde d_k, t_k y i_k representan al valor del MPI de dificultad, tiempo e intentos de la entrada k de la actividad a calificar. Usualmente, las actividades de estudiantes se evalúan por medias ponderadas. Esto implica que las relaciones entre la dificultad, el tiempo y los intentos de un mismo ejercicio, o entre los ejercicios de la actividad se consideran irrelevantes para la evaluación. Sin embargo, en [10] y [11] se muestra con ejemplos que aún en casos muy sencillos, esta hipótesis no se cumple generalmente.

En este trabajo se propone como operador de agregación la integral de Choquet. Sea $N = \{1, 2, \dots, n\}$ el conjunto de n índices, de manera que cada índice represente a un atributo de una situación $x = (x_1, x_2, \dots, x_n) \in X$. Una capacidad borrosa es una función del conjunto potencia de N al intervalo unitario, $\mu: \mathcal{P}(N) \rightarrow [0, 1]$ que satisface las condiciones siguientes:

- $\mu(\emptyset) = 0, \mu(N) = 1$ (condiciones de frontera);
- para todo $S, T \subseteq N$, tal que $S \subseteq T$ entonces $\mu(S) \leq \mu(T)$ (monotonía).

Para cada subconjunto de índices $S \subseteq N$, el número $\mu(S)$ representa el peso o importancia de S . La monotonía de μ significa que el peso o importancia de un subconjunto de índices no puede disminuir si se agregan nuevos índices. La integral de Choquet respecto a una capacidad μ es una función $C_\mu: [0, 1]^n \rightarrow [0, 1]$ definida por

$$C_\mu(x) := \sum_{i=1}^n x_{\sigma(i)} [\mu(A_{\sigma(i)}) - \mu(A_{\sigma(i+1)})],$$

donde $\sigma = (\sigma(1), \dots, \sigma(n))$ es una permutación de N tal que $x_{\sigma(1)} \leq x_{\sigma(2)} \leq \dots \leq x_{\sigma(n)}$. Igualmente, para todo $i \in N$, $A_{\sigma(i)} = \{\sigma(i), \dots, \sigma(n)\}$ y $A_{\sigma(n+1)} = \emptyset$.

El análisis del comportamiento de la integral de Choquet en cuanto a relevancia de criterios e interacciones, se hace mediante dos de los índices más reportados en la literatura. Para evaluar la importancia de cada uno de los atributos respecto a los demás, utilizamos los valores de Shapley, definidos como

$$\phi_\mu(i) := \sum_{T \subseteq N - \{i\}} \frac{(n - |T| - 1)! |T|!}{n!} (\mu(T \cup \{i\}) - \mu(T)),$$

y pueden verse como el valor promedio de la contribución marginal de i a todas las capacidades.

La interacción básica entre dos índices i y j se puede obtener calculando la diferencia $[\mu(\{i, j\}) - [\mu(\{i\}) + \mu(\{j\})]]$. Si la diferencia es positiva, se dice que son complementarios; si la diferencia es 0 se dice que son independientes; y si la diferencia es negativa, se dicen redundantes. La idea básica del valor de Shapley se puede generalizar a la relación con los dos índices, permitiendo calcular un valor de interacción de la forma

$$I_{\mu}(i, j) = \sum_{T \subseteq N - \{i, j\}} \frac{(n - |T| - 2)! |T|!}{(n - 1)!} [\mu(T \cup \{i, j\}) - \mu(T \cup \{i\}) - \mu(T \cup \{j\}) + \mu(T)].$$

El valor $I_{\mu}(i, j) \in [-1, 1]$ puede por lo tanto ser considerado como el promedio de la interacción marginal de i y j , cuyo valor es 1 si existe una complementariedad máxima y -1 si los índices son completamente redundantes.

Frente a problemas de evaluación, es poco factible solicitar a la institución o pedagogos reglas generales que permitan obtener el operador de agregación. Para los expertos puede resultar más fácil expresar su conocimiento en casos concretos. En este trabajo se propone un método interactivo donde los expertos expresen su conocimiento como relaciones de preferencias, que irán tomando una complejidad creciente a medida que se interacciona con el método de adaptación. El experto expresa su conocimiento a partir de ejemplos de actividades que pueden ser reales o casos prototípicos.

Sea u y u' dos actividades prototípicas descritas por un vector de MPI de dimensión 15, se dice que u debe estar mejor calificada u' y se denota $u \succcurlyeq u'$ si y solo si

$$C_{\mu}(u) - C_{\mu}(u') \geq \delta_c,$$

donde δ_c es una constante pequeña. De la misma forma, se dice que dos actividades deben tener una calificación parecida (denotada por $u^{(p)} \sim_A u^{(q)}$) siempre que $|C_{\mu}(u) - C_{\mu}(u')| \leq \delta_c$.

Una vez establecido el conjunto de casos particulares y un orden parcial de ellos, se establece un orden entre los valores de importancia de los índices, lo que permite expresar relaciones de importancia entre la dificultad de los ejercicios y la importancia del tiempo respecto a los intentos en cada ejercicio.

Se dice que un índice i se prefiere respecto a un índice j ($i \succcurlyeq_N j$) si $\phi_{\mu}(i) - \phi_{\mu}(j) \geq \delta_{sh}$ donde δ_{sh} es una constante muy pequeña. De la misma forma dos índices i y j tienen una importancia semejante ($i \sim_N j$) si $|\phi_{\mu}(i) - \phi_{\mu}(j)| \leq \delta_{sh}$.

Si es necesario, se pueden establecer criterios que fijen las interdependencias entre los diferentes criterios. La relación entre los índices i y j se obtiene directamente al imponer el criterio $I_{\mu}(i, j) < 0$ para redundancia $I_{\mu} > 0$ para complementariedad y $|I_{\mu}(i, j)| < \epsilon$ para independencia. Por otra parte, las relaciones de preferencia entre relaciones de índices, tales como $i, j \succcurlyeq_I k, l$ están dadas por $I_{\mu}(i, j) - I_{\mu}(k, l) \geq \delta_I$.

El proceso de especificación se hace en forma incremental, buscando representar cada vez un objetivo de aprendizaje más complejo. El método de construcción de g se desarrolló en el lenguaje estadístico R usando librería *kappalab* [9].

Con el fin de ilustrar el proceso de generación de operador de agregación se presenta un ejemplo ilustrativo, el cual se basa en cinco casos prototípicos representados en el

cuadro 1. El orden de preferencia establecido por el tomador de decisiones es $e_1 \succ e_2 \succ e_3 \succ e_4 \succ e_5$. Este criterio busca fomentar la competitividad matemática.

Cuadro 1. Casos prototípicos del ejemplo ilustrativo

Estudiante	Índices marginales de desempeño														
	d_1	t_1	i_1	d_2	t_2	i_2	d_3	t_3	i_3	d_4	t_4	i_4	d_5	t_5	i_5
e_1	0.2	0.4	0.9	0.4	0.4	0.9	0.6	0.4	0.5	0.8	0.4	0.5	1.0	0.4	0.5
e_2	0.2	0.8	0.9	0.4	0.8	0.5	0.6	0.4	0.9	0.8	0.4	0.5	1.0	0.4	0.5
e_3	0.2	0.8	0.5	0.4	0.8	0.9	0.6	0.4	0.9	0.8	0.4	0.5	1.0	0.4	0.5
e_4	0.2	0.8	0.5	0.4	0.8	0.5	0.6	0.4	0.9	0.8	0.4	0.9	1.0	0.4	0.5
e_5	0.2	0.8	0.5	0.4	0.4	0.5	0.6	0.8	0.5	0.8	0.8	0.9	1.0	0.4	0.9

Una vez parametrizada la integral de Choquet se evalúan los casos prototípicos obteniendo la estimación del nivel de habilidad para cada estudiante siguiente: $g(e_1) = 0.9$, $g(e_2) = 0.8$, $g(e_3) = 0.7$, $g(e_4) = 0.6$ y $g(e_5) = 0.5$ respectivamente.

Con el fin de entender el comportamiento de la evaluación, se revisan los valores de Shapley en cada criterio y los índices de interacción. Dado que se desea establecer un criterio para que la plataforma promueva la competitividad matemática, se revisan principalmente los índices que se relacionan con el número de intentos. Al obtener los valores de Shapley se encuentra que $\phi_\mu(i_1) = 0.375$, $\phi_\mu(i_2) = \phi_\mu(i_3) = 0.125$ y los valores de Shapley para los problemas con mayor dificultad $\phi_\mu(i_4) = \phi_\mu(i_5) = 0$.

Esta conducta es inaceptable desde el punto de vista de la competitividad matemática, ya que no le da importancia a requerir muchos intentos en los ejercicios de mayor dificultad. Con este fin se agrega al algoritmo de parametrización del operador de agregación la consideración

$$i_1 \sim_N i_2 \sim_N i_3 \sim_N i_4 \sim_N i_5.$$

Una vez obtenida la nueva parametrización, se obtienen valores de Shapley $\phi_\mu(i_k) = 0.125$ para $k = 1, \dots, 5$. Al recalcular los niveles estimados de habilidad obtenemos lo siguiente: $g(e_1) = 0.68$, $g(e_2) = 0.63$, $g(e_3) = 0.58$, $g(e_4) = 0.53$ y $g(e_5) = 0.48$ respectivamente. Estas evaluaciones cumplen con todos los criterios establecidos y se ajustan mejor a las evaluaciones que consideran correctas los tomadores de decisiones. Es interesante resaltar que, al estudiar la relación de interacción entre índices, se encontró que el número de intentos entre los diferentes problemas son independientes entre sí salvo por los ejercicios 3 y 5 en el cual $I_\mu(i_3, i_5) = 0.25$.

6 Conclusiones y trabajos futuros

En este trabajo se propone un método basado en la metodología MAUT para la evaluación del nivel de habilidad de un estudiante, en el cual los objetivos de aprendizaje se puedan establecer en términos de comparación de casos reales o prototípicos. El método propuesto utiliza la información histórica de la plataforma para establecer los índices de desempeño marginal, así como el conocimiento experto y las preferencias de los

tomadores de decisiones, a través de un método interactivo para construir un operador de agregación. Para esto último, se adaptó la metodología MAUT que permite expresar un conocimiento experto en función de relaciones particulares tanto entre niveles de habilidad de estudiantes como entre los criterios considerados para la evaluación.

Actualmente el método propuesto se encuentra en etapas de prueba en la plataforma Sofía XT. Con esta propuesta se espera poder personalizar la plataforma de enseñanza a las necesidades y objetivos de cada institución de enseñanza que la utilice.

Agradecimientos

Este trabajo ha sido parcialmente financiado por la Agencia Estatal de Investigación (AEI) y el Fondo Europeo de Desarrollo Regional (FEDER) dentro del proyecto MERINET: TIN2016-76843-C4-2-R (AEI/FEDER, UE). Sonia G. Sosa-León, agradece el apoyo otorgado por PRODEP (México) mediante beca con folio UNISON-344. Julio Weissman agradece a la Universidad Castilla – La Mancha por su financiamiento a través del programa de apoyo a profesores visitantes.

Referencias

1. Wauters, K., Desmet, P., Van Den Noortgate, W.: Adaptive item-based learning environments based on the item response theory: Possibilities and challenges. *Journal of Computer Assisted Learning*, 26(6), 549–562 (2010).
2. Klinkenberg, S., Straatemeier, M., Van Der Maas, H. L.: Computer adaptive practice of Maths ability using a new item response model for on the fly ability and difficulty estimation. *Computers and Education*, 57(2), 1813–1824 (2011).
3. Cacho-Carranza, Y.: Sofía XT, una web para mejorar el aprendizaje matemático en niños. *Agencia Informativa CONACYT AIC*. Ciudad de México. <http://conacytprensa.mx/index.php/tecnologia/tic/>, ultimo acceso 2017/06/01.
4. Kamvysi, K., Gotzamani, K., Andronikidis, A., Georgiou, A. C.: Capturing and prioritizing students' requirements for course design by embedding Fuzzy-AHP and linear programming in QFD. *European Journal of Operational Research*, 237(3), 1083–1094 (2014).
5. Rasch, G.: Probabilistic models for some intelligence and achievement tests. Danish Institute for Educational Research, Copenhagen (1960).
6. Maris, G., van der Maas, H.: Speed-Accuracy Response Models: Scoring Rules based on Response Time and Accuracy. *Psychometrika*, 77(4), 615–633 (2012).
7. Chinn, S.: Mathematics Anxiety in Secondary Students in England. *Dyslexia*, 15, 61–68 (2009).
8. Ma, X., Xu, J.: The causal ordering of mathematics anxiety and mathematics achievement: A longitudinal panel analysis. *Journal of Adolescence*, 27(2), 165–179 (2004).
9. Bigaret, S., Meyer, P.: Evaluation and Decision Models with Multiple Criteria: Case Studies. Springer-Verlag, Berlin Heidelberg (2015).
10. Shieh, J. I., Wu, H. H., Liu, H. C.: Applying a complexity-based Choquet integral to evaluate students' performance. *Expert Systems with Applications*, 36(3):5100–5106, (2009).
11. Grabisch, M., Kojadinovic, I., Meyer, P.: A review of methods for capacity identification in Choquet integral based multi-attribute utility theory. Applications of the Kappalab R package. *European Journal of Operational Research*, 186(2):766–785, (2008).

Experiencia de enseñanza-aprendizaje de Cloud Computing y Cloud Robotics en la UNLP

Armando De Giusti^{1,2}, Ismael Rodriguez¹, Sebastián Rodríguez Eguren¹

¹ Instituto de Investigación en Informática LIDI (III-LIDI), Facultad de Informática, Universidad Nacional de La Plata – Comisión de Investigaciones Científicas de la Provincia de Buenos Aires.

² Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)
Argentina
{degiusti, ismael, seguren} @lidi.info.unlp.edu.ar

Resumen. Se presenta y analiza el conjunto de experiencias obtenidas en la enseñanza de los temas relacionados con Cloud Computing y Cloud Robotics, en las carreras de Licenciatura e Ingeniería de la Facultad de Informática de la UNLP, considerando el impacto tecnológico de los paradigmas actuales. Asimismo, se exponen algunos resultados concretos de la experiencia obtenida en los últimos años.

Palabras Clave: Cloud Robotics, Cloud Computing, Internet of Things, Informática, Enseñanza, Amazon Web Service (AWS).

1 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 y también como un esquema de procesamiento masivo de datos y de servicios remotos de gran impacto en la Informática actual. Esta tecnología provee una abstracción de los recursos de hardware permitiendo ejecutar simultáneamente múltiples instancias de Software en máquinas virtuales sobre un único hardware físico (centralizado o distribuido).

Cloud Computing proporciona grandes conjuntos de recursos físicos y lógicos (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”. Estos recursos son proporcionados como servicios (“As a Service”) y pueden ser dinámicamente re-configurados para adaptarse a una carga de trabajo variable (escalabilidad), logrando una mejor utilización y evitando el sobre o sub dimensionamiento (elasticidad). [1] [2].

Al mismo tiempo, considerando la gran influencia de la robótica en la sociedad y la diversidad de servicios ofrecidos por robots, nos encontramos que los mismos presentan grandes limitaciones en consumo de energía, poder de cómputo, capacidad de almacenamiento, toma de decisiones, tareas cognitivas, etc., por lo tanto, surge la

necesidad de integrar las tecnologías de Cloud Computing con los sistemas de Robots. [3][4][5].

En el año 2010, comenzaron a surgir proyectos de investigación (ej.: RoboEarth [6]), que integraban las tecnologías de Cloud con los sistemas de robots. Es así, que James Kuffner propone el concepto de Cloud Robotics, basado en combinar las tecnologías de robots con el paradigma de Cloud Computing [7].

La idea de Cloud Robotics, permite por medio de aplicaciones tratar los datos de los componentes de hardware del robot (sensores, actuadores, cámaras, microcontroladores, memoria, etc.), sin importar las limitaciones de cómputo de las placas de desarrollo con microcontroladores y la capacidad de almacenamiento de las mismas [8]. En otras palabras, este concepto permite a los robots obtener resultados de tareas de cómputo intensivo, tales como: procesamiento de imágenes, reconocimiento de voz, determinación de rutas, confección de mapas, acciones cognitivas, etc., sin tratamiento local, sino en el Cloud.

Este paradigma brinda la capacidad de establecer escenarios para sistemas de multi-robot, donde cada robot se integra de un hardware mínimo; una placa microcontroladora con conectividad WiFi permite al robot comunicarse con el Cloud y otros robots. Los datos de sensores y/o sistemas de adquisición de imágenes se procesan en el Cloud y los actuadores de cada robot llevarán a cabo las operaciones necesarias [9].

El presente trabajo está estructurado de la manera siguiente: en la Sección 2, se mencionan los antecedentes en las experiencias de enseñanza de temas de Cloud Computing y Cloud Robotics; en la Sección 3, se presenta la propuesta general aplicada en la asignatura “Cloud Computing. Cloud Robotics” de las carreras de la Facultad de Informática de la UNLP. En la Sección 4, se describen las actividades desarrolladas en los cursos dictados. Por último, en la Sección 5, se presentan algunos comentarios finales y proyecciones a futuro.

2 Antecedentes

En el año 2013, el Honorable Consejo Directivo de la Facultad de Informática, aprobó la propuesta de enseñanza de la asignatura “Cloud Computing. Aplicaciones en Big Data” [10], con régimen de cursada semestral, de carácter optativa, con una carga horaria semanal de seis horas, válida para alumnos de 4to/5to año de los planes de estudio de las carreras de Licenciatura en Sistemas, Licenciatura en Informática e Ingeniería en Computación de la Facultad.

Los objetivos generales de la asignatura eran:

- Profundizar el conocimiento de la tecnología Cloud y su utilización.
- Presentar algunos de los problemas principales en las aplicaciones sobre Cloud Computing.
- Estudiar los fundamentos de Big Data.
- Resolver problemas de Big Data, utilizando arquitecturas Cloud.

Para alcanzar tales objetivos, se propuso incorporar actividades colaborativas de carácter teórico-práctico que permitieron a los alumnos relacionar los contenidos

presentados en las clases teóricas y las actividades/talleres, con el fin de resolver un caso problema.

La práctica de la materia se llevó a cabo realizando todos los trabajos colaborativos sobre el Cloud público de Amazon Web Services (AWS) [11]. Para esto, la asignatura contó con una beca de investigación y educación proporcionada por AWS, que permitió a cada alumno disponer de un acceso vía WEB, con un usuario con perfil de administrador. Sobre AWS, los alumnos desplegaron instancias de máquinas virtuales configuradas en Clúster, ejecutaron aplicaciones de HPC y realizaron tratamiento masivo de datos utilizando Amazon Elastic MapReduce (Amazon EMR) [12].

Luego de tres años de dictado de la asignatura, visto el gran interés manifestado por los alumnos y atento a los requerimientos de los desafíos actuales, se consideró necesario ampliar la propuesta de enseñanza, desdoblando la misma en dos asignaturas que consideraran los temas en forma más amplia, denominadas “Conceptos y Aplicaciones de Big Data” y “Cloud Computing. Cloud Robotics”; en esta última, se busca profundizar en el conocimiento de la integración de tecnologías del paradigma de Cloud Computing y sistemas de Robots, logrando conectar Robots al Cloud.

Paralelamente, en el año 2015 se dictó el curso de “Cloud Computing con AWS (Amazon Web Services)” en la Escuela Internacional de Informática en el marco del XXI Congreso Argentino de Ciencias de la Computación (CACIC 2015) [13], destinado a alumnos avanzados de grado.

Asimismo, en el año 2016 y 2017 se dictó el curso de “Cloud Robotics” en el marco de las Jornadas de Cloud Computing & Big Data [14] [15], destinado a alumnos avanzados de grado, alumnos de Postgrado e investigadores del área.

3 Propuesta general

En vista a los antecedentes mencionados, en el año 2016, se presentó al Honorable Consejo Directivo de la Facultad de Informática, la propuesta de una nueva asignatura que mantiene el régimen de cursada semestral y su carácter optativo, para alumnos avanzados de 4to/5to año, presentando un nuevo paradigma de integración de las tecnologías de Cloud Computing y los sistemas de Robótica, denominada “Cloud Computing. Cloud Robotics”, entrando en vigencia para la oferta académica de Optativas 2017.

Los objetivos generales de esta asignatura son:

- Profundizar el conocimiento de las arquitecturas CLOUD y su utilización.
- Analizar los problemas principales en las aplicaciones de Cloud Computing.
- Estudiar los fundamentos de los problemas de Robótica.
- Resolver los sistemas de Robots utilizando arquitecturas Cloud.

Los contenidos de la asignatura son:

A. Fundamentos de Cloud Computing

- Evolución: de Clusters y GRID a Cloud.
- Análisis de beneficios del empleo de arquitecturas Cloud: Escalabilidad, Disponibilidad, Confiabilidad.

- Riesgos y desafíos del empleo de arquitecturas Cloud: Seguridad, Portabilidad, Eficiencia.
- Características de un Cloud.
- B. Arquitecturas en Cloud Computing
 - IaaS. Infraestructura como Servicio.
 - PaaS. Plataforma como Servicio.
 - SaaS. Software como Servicio.
 - Tipos de arquitecturas Cloud. (públicas, privadas, híbridas).
 - Virtualización.
 - Manejo de las Comunicaciones en Cloud.
 - Manejo de Datos en Cloud.
 - Seguridad en Cloud.
- C. Software de Base y de aplicación en Cloud Computing
 - Hipervisores.
 - Balance de carga.
 - Administración de recursos distribuidos.
 - Configuración de máquinas virtuales.
 - Programación de aplicaciones sobre Cloud.
- D. Conceptos Fundamentales de IoT
 - Introducción: ¿Qué es Internet de las cosas?
 - Casos de estudio: Ciudades Inteligentes (Smart Cities) y Logísticas Inteligentes (Smart Logistic).
 - Tecnologías: NFC, RFID, Bluetooth 4.0, WiFi.
 - Tecnología M2M: Protocolo MQTT.
 - Seguridad en IoT.
 - Estudios de casos: Robots y Vehículos Autónomos.
- E. Robots conectados al Cloud
 - Placas basadas en Linux embebido: Arduino, Raspberry Pi, Intel Galileo entre otras.
 - Conectividad: USB, Ethernet y WiFi, y Bluetooth.
 - Dispositivos sensores: de Humedad Relativa y Temperatura (DTH22), Ultrasonido (HC-SR04), Infrarrojo PIR, de Humedad de Superficies, de Imagen (cámara VGA y HD).
 - Lenguajes de programación: C, Python y Node-Red.
 - Programación básica en Arduino y Raspberry Pi, conectados al Cloud.
 - Programación del protocolo MQTT para interactuar con el Cloud e inter-robots.
 - Estudios de casos: Robots Conectados al Cloud.

La propuesta metodológica de la asignatura, propone desarrollar actividades colaborativas de carácter teórico-práctico, que permitan a los alumnos relacionar los contenidos presentados en las clases teóricas con las actividades prácticas. Estas actividades se llevan a cabo sobre el Cloud público de Amazon Web Services (AWS). Para ello, la asignatura cuenta con una nueva beca de investigación y educación proporcionada por AWS, que permite a los alumnos acceder vía WEB a los servicios del Cloud.

La comunicación adicional con los alumnos y el espacio de consulta de contenidos del curso se efectúa bajo el entorno virtual de enseñanza y aprendizaje IDEAS [16],

desarrollado por el Instituto de Investigación en Informática III-LIDI de la Facultad de Informática de la UNLP. Este entorno ofrece un trabajo en áreas: Bienvenida, Comunicación, Información General y Contenidos, Trabajo Colaborativo, Recursos Educativos y Evaluación y Seguimiento. Cada área tiene, a su vez, diferentes herramientas para el trabajo entre los asistentes al curso virtual. Por ejemplo: en el área de Comunicación se puede elegir trabajar con mensajería, cartelera de novedades, foros y/o chat.

La asignatura presenta dos modalidades de cursada: presencial y semipresencial. En la modalidad presencial, los alumnos deben cumplir con la entrega de las actividades colaborativas y la aprobación de las mismas; el cumplimiento de estos requerimientos habilita al alumno a acceder a una modalidad de promoción, rindiendo un coloquio integrador al finalizar la cursada. En la modalidad semipresencial, los alumnos deben cumplir los mismos requerimientos que la modalidad presencial, debiendo rendir un examen final o elaborar un trabajo final experimental.

4 Actividades experimentales

Para abordar los contenidos teórico-prácticos, se realizan clases con modalidad de laboratorio, donde los alumnos interactúan en grupo, aplicando los conceptos dados sobre el Cloud público AWS.

A lo largo del curso, se realizan cinco actividades colaborativas, consecutivas e interrelacionadas, con un aumento en la graduación de la dificultad de las mismas:

- Actividad N° 1: Introducción a Amazon Elastic Compute Cloud (Amazon EC2). El objetivo de esta actividad es que el alumno se familiarice con el Cloud Público AWS, administrando instancias EC2, gestionando accesos a las mismas por SSH y aplicando el concepto de IP elástica.
- Actividad N° 2: Gestión de almacenamiento Amazon Simple Storage Service (Amazon S3) y Amazon Elastic Block Store (Amazon EBS). El objetivo de la actividad es que el alumno sepa diferenciar los sistemas de almacenamiento y adquiera los conocimientos necesarios para seleccionar la mejor herramienta a utilizar de acuerdo al problema a resolver, ya sea, almacenamiento en buckets (S3) o en volúmenes (EBS).
- Actividad N° 3: Utilización de AWS desde la Interfaz de Línea de Comandos (Amazon CLI). Se espera como resultado de esta actividad que el alumno logre familiarizarse con la interfaz de línea de comandos (CLI), como herramienta para gestionar y controlar los servicios de AWS, mediante la automatización de secuencias de comandos.
- Actividad N° 4: IoT utilizando AWS IoT. El objetivo de esta actividad es introducir al alumno al mundo de “Internet de las Cosas” utilizando los servicios de AWS IoT.
- Actividad N° 5: Robots conectados al Cloud. Se espera como resultado de esta actividad que el alumno implemente una solución mediante la cual conecte sensores a un microcontrolador y los pueda controlar por medio del Cloud.

Cada actividad práctica está planificada para desarrollarse en el transcurso de dos semanas.

5 Comentarios finales

Se presentan los resultados de las experiencias en el dictado de temas relacionados con Cloud Computing y Cloud Robotics, en los cursos de los últimos años.

En el curso dictado en el año 2014 [17], asignatura “Cloud Computing. Aplicaciones en Big Data”, se inscribieron 27 alumnos, donde 13 alumnos no iniciaron nunca la cursada, y participaron de la misma 14 alumnos: 12 en modalidad presencial y 2 en modalidad semipresencial. Los resultados obtenidos fue del 85% de alumnos que alcanzaron la resolución de las actividades colaborativas presentadas; como así también, el 100% de alumnos que promocionaron/aprobaron la materia por medio de la instancia de coloquio integrador/final.

En el curso dictado en el año 2015, asignatura “Cloud Computing. Aplicaciones en Big Data”, se inscribieron 15 alumnos, donde 5 alumnos no iniciaron nunca la cursada, participaron de la misma 10 alumnos: 8 en modalidad presencial y 2 en modalidad semipresencial. Los resultados obtenidos fue del 80% de alumnos que alcanzaron la resolución de las actividades colaborativas presentadas; como así también, el 100% de alumnos que promocionaron/aprobaron la materia por medio de la instancia de coloquio integrador/final.

En el curso dictado en el año 2016, asignatura “Cloud Computing. Aplicaciones en Big Data”, se inscribieron 65 alumnos, donde 15 alumnos no iniciaron nunca la cursada, participaron de la misma 48 alumnos: 35 en modalidad presencial y 13 en modalidad semipresencial. Los resultados obtenidos fue del 98% de alumnos que alcanzaron la resolución de las actividades colaborativas presentadas; como así también, el 85% de alumnos que promocionaron/aprobaron la materia por medio de la instancia de coloquio integrador/final.

Actualmente, en el curso del año 2017, asignatura “Cloud Computing. Cloud Robotics”, se inscribieron 33 alumnos, donde 8 alumnos no iniciaron nunca la cursada, participan de la experiencia 25 alumnos: 22 en modalidad presencial y 3 en modalidad semipresencial; a la fecha, el 88% de los alumnos cumple con los requisitos de entrega y aprobación de las actividades colaborativas, restando aún la presentación de la actividad N° 5, como así también la instancia del coloquio integrador o el examen final, según corresponda a la modalidad de cursada obtenida.

A lo largo del período analizado, los alumnos han manifestado gran interés sobre los temas abordados en la asignatura, y conformidad con la metodología, como así también, con las actividades colaborativas presentadas, derivando un trabajo final de carrera de grado de Licenciatura en Sistemas, con nota diez [18]; y actualmente, cuatro alumnos se encuentran desarrollando su trabajo final de carrera de grado con temas relacionados al área, bajo la dirección y asesoramiento de los docentes de la asignatura.

Como trabajo futuro a corto plazo, se proyecta incorporar a las actividades prácticas la interacción con un dispositivo de vuelo (Drone/Cuadrícóptero) con el fin de ampliar el sistema multi-robot con el que lo alumnos puedan trabajar.

Además, se considerará la posibilidad de incorporar el acceso al Cloud de Microsoft Azure, con el fin de comparar los servicios de los Cloud Públicos mencionados en este trabajo.

Referencias

1. Thomas Erl “Cloud Computing. Concepts, Technology & Architecture”. Prentice Hall 2013.
2. Rothon J. “Cloud computing explained: Implementation handbook for enterprises”. Recursive Press 2013.
3. Terrissa L.S., Ayad S. “Towards a new cloud robotics approach”. IEEE Cist Conference 2016.
4. Chen Y., Du Z., Garcia-Acosta M. “Robot as a Service in Cloud Computing”. Fifth IEEE International Symposium on Service Oriented System Engineering 2010.
5. Doriya R., Chakraborty P., Nandi G. C. “Robotic Services in Cloud Computing Paradigm” International Symposium on Cloud and Services Computing 2012.
6. RoboEarth. In: <http://www.roboearth.org>.
7. Kuffner J.: “Cloud-enabled robots”. IEEE-RAS International Conference on Humanoid Robot. Nashville, USA, 2010.
8. Wang L., Liu M., Meng M., Siegwart R.: “Towards Real-Time Multi-Sensor Information Retrieval in Cloud Robotic System”. IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI). Hamburgo, Alemania, 2012.
9. Turnbull L., et al.: “Cloud Robotics: Formation Control of a Multi Robot System Utilizing Cloud Infrastructure”. Proceedings of IEEE – Southeastcon. Jacksonville, USA, 2013.
10. De Giusti A., Naiouf M., De Giusti L., Chichizola F., Pousa A., Sanz V., Rodriguez I., Rucci E. “Enfrentando desafíos de la currícula en Informática: Concurrencia, Paralelismo, Cloud Computing, Multicores y GPUs. Proceedings del X Congreso de Tecnología en Educación y Educación en Tecnología (TE&ET 2015). Corrientes, Argentina. (2015).
11. Amazon Elastic Compute Cloud (AmazonEC2). <http://aws.amazon.com/es/ec2>.
12. Amazon Elastic MapReduce (Amazon EMR). <https://aws.amazon.com/es/emr/>.
13. XXI Congreso Argentino de Ciencias de la Computación. CACIC 2015. <http://cacic2015.unnoba.edu.ar/>.
14. IV Jornadas de Cloud Computing & Big Data. JCC 2016. <http://www.jcc2016.info.unlp.edu.ar/wordpress/>
15. V Jornadas de Cloud Computing & Big Data. JCC 2017. <http://www.jcc.info.unlp.edu.ar/wordpress/>
16. Entorno virtual de enseñanza y aprendizaje IDEAS. <https://ideas.info.unlp.edu.ar>.
17. De Giusti A. Rodríguez I., Hasperue W., Rodríguez Eguren S. “Estrategias para la enseñanza de Cloud Computing en la UNLP”. II Jornadas de Cloud Computing & Big Data (JCC 2015). La Plata. Argentina (2015).
18. Tesina de grado. “Incidencia de idiomas populares en la lengua española con Big Data: análisis masivo de datos mediante Amazon Elastic MapReduce y Google N-grams”, De Luca J., <http://catalogo.info.unlp.edu.ar/meran/opac-detail.pl?id1=7178>.

Extensión, Docencia e Investigación destinada a los Comedores Barriales: un caso testigo de cómo se combinan los pilares de la Universidad en forma multidisciplinaria

Viviana Harari¹, Mariela Stávale² e Ivana Harari¹

¹Facultad de Informática, Universidad Nacional de La Plata. La Plata 1900, Buenos Aires, Argentina

{vharari, iharari}@info.unlp.edu.ar

²Facultad de Humanidades y Ciencias de la Educación, Universidad Nacional de La Plata. La Plata 1900, Buenos Aires, Argentina

mari_stavale@yahoo.com.ar

Resumen. Este artículo detalla las actividades de investigación y docencia que se integraron dentro del proyecto de extensión El Barrio va a la Universidad, aprobado en sucesivas convocatorias desde el 2010. El mismo se caracteriza por sus diferentes líneas de acción que permiten acercar a niños y jóvenes de los comedores barriales de las zonas marginales de La Plata a la Universidad, descubriendo sus talleres, carreras y cursos. En el marco del proyecto se aplicó un trabajo en conjunto con la Facultad de Humanidades y Ciencias de la Educación que se va a detallar en el artículo, donde se abordó el tema de la Memoria en el BIM 3, lugar que funcionó como un Centro Clandestino de Detención, Tortura y Exterminio (CCDTE), durante la última dictadura militar. Los trabajos de investigación fueron adaptados para ser transmitidos en forma adecuada y pedagógica a los niños, que culminaron con producciones periódicas y de relatos, utilizando todos los recursos tecnológicos aprehendidos.

Keywords: Alfabetización informática, ONGs, brecha digital, TICs

1 Introducción

La Facultad de Informática¹ de la Universidad Nacional de La Plata (UNLP), se creó en el año 1999. Desde sus inicios ha trabajado en línea con la misión de la UNLP expresada en el preámbulo de su estatuto, en el cual se reconoce como funciones primordiales el desarrollo y fomento de la enseñanza, la investigación y la extensión.

¹ www.info.unlp.edu.ar

En el marco de ese objetivo, ha trabajado en pos de mejorar los vínculos con la comunidad en relación a diversas áreas de interés común, en involucrar a los estudiantes en las nuevas problemáticas de la sociedad y en promover políticas públicas dirigidas a una nueva comunidad estudiantil².

Se ha implementado numerosos programas con diferentes objetivos que abarcan desde temas relacionados con el medio ambiente, accesibilidad hasta vínculo con escuelas tanto primarias como secundarias. En particular, desde el año 2007, viene trabajando con sectores sociales vulnerables de la ciudad de La Plata y alrededores, en el marco de un programa denominado "Informática Inclusiva" [1]. En el año 2017, se cumplen 10 años de un trabajo continuo e ininterrumpido con estos sectores con quienes se ha forjado una relación estrecha que permite un "intercambio" de saberes.

En el marco de dicho programa se han ejecutado diferentes proyectos que han tenido diferentes denominaciones a lo largo de los años, pero que han conservado el objetivo principal, que es el de acortar la brecha digital en sectores sociales vulnerables de la ciudad de La Plata y alrededores. Entre ellos, se encuentra el proyecto de El Barrio va a la Universidad, vigente desde el año 2010 [2]. Si bien, el trabajo está destinado a diferentes grupos etarios, la mayoría de los destinatarios son niños de edad escolar primaria y, de los primeros años de la escuela secundaria. Se destaca este punto porque dada las características de los actores, las capacitación informática se basan en el uso de TICs aplicadas a la educación.

Los proyectos desarrollados fueron enriqueciéndose y ampliándose, producto de las demandas y necesidades que surgían de las interacciones permanentes con los diferentes sectores sociales. En esta línea, uno de los aportes más significativos que se han realizado es la incorporación de otro objetivo al programa, relacionado con sensibilizar a los niños y jóvenes, sobre la idea de "Universidad" como institución posible de elección, a la hora de pensar en su futuro.

A través de este artículo se contará una experiencia en donde se usaron diferentes recursos informáticos, se enseñaron otros y, se realizaron diferentes tareas de investigación relacionada con la historia reciente de nuestro país. En este caso, se instanció el trabajo sobre la Memoria en el BIM 3, Batallón de Infantería de la Marina Número 3, articulando con investigadores de la Facultad de Humanidades de la UNLP.

Fue una actividad que vinculó la docencia, la investigación y la extensión, en un contexto de inclusión e involucrando diferentes disciplinas.

2 El proyecto El Barrio va a la Universidad

Desde el año 2007, docentes y estudiantes voluntarios de la Facultad de Informática, concurren a organizaciones barriales para estrechar vínculos genuinos entre dichas entidades y la Universidad, a través de la enseñanza de las tecnologías de información y comunicación (TICs) aplicada a diferentes áreas disciplinares.

Se inició con 3 comedores barriales y un grupo de 38 niños, donde se enseñaban los conceptos básicos de computación, ofimática, Internet, entre otros y, se alquilaba

² www.unlp.edu.ar

cybers cercanos para dar la capacitación. A lo largo de los años el vínculo fue afianzándose, el proyecto fue creciendo y expandiéndose según las necesidades y demandas detectadas.

Actualmente se alcanzó un trabajo permanente con 10 instituciones, destinado a 120 niños y adolescentes que acuden a dichas entidades. Las actividades de enseñanza de Computación se acrecentaron, transformándose en propuestas pedagógicas y de extensión que integran diferentes temáticas multidisciplinares.

Entre las diferentes líneas de trabajo que fueron evolucionando a lo largo del tiempo, se pueden mencionar algunas de ellas como:

- Integración con el proyecto de e-Basura, que permitió la donación de computadoras a los comedores barriales, pudiendo conformar en sus instalaciones salas de PC para poder llevar a cabo las actividades allí [3].
- Alfabetización Informática articulada con actividades de apoyo escolar sobre asignaturas de la escuela primaria. Se trabajó en conjunto con el apoyo escolar de las diferentes asociaciones civiles, incluyendo actividades que afianzaran la lecto escritura y comprensión de textos.
- Cambiando el imaginario social respecto a la Universidad, donde se trabajó sobre el descubrimiento de la UNLP, conociendo sus carreras de grado, posgrado y talleres de oficio. Se llevaron a cabo visitas, entrevistas, actividades conjuntas con distintas unidades académicas y se trabajó sobre producciones informáticas relacionadas al respecto [4].
- Profundización en temas informáticos innovativos, como ciberseguridad, programación en robots, realidad aumentada. En muchos de estos casos se articularon actividades en conjunto con otros proyectos de extensión como también con materias de 5to año de las carreras de Informática de la UNLP [5].
- Iniciación en la Programación. A partir del 2016, se comenzó con la formación en esta temática con los alumnos más avanzados. Se los introdujo a la temática desde diferentes estrategias didácticas, que van desde la experimentación misma, donde el docente simulaba ser un robot y los niños, con el editor de textos, escribían las órdenes para que realice una tarea determinada, hasta la utilización de una herramienta para programación como Scratch [6].

Estas actividades y muchas más, dan cuenta de la evolución, dinamismo y transformación que el proyecto de extensión va adquiriendo, adaptándose a las problemáticas detectadas en la comunidad destinataria, dejando de ser simples cursos de Computación en los comedores, a ofrecer propuestas innovativas donde la enseñanza de las TICs se aplican a diferentes temáticas multidisciplinares que requieren la articulación con la investigación y extensión.

Estas actividades permiten aplicar las tecnologías al servicio de otras disciplinas y en dicho proceso aprender a utilizarlas. Las producciones de los chicos fueron desde notas en editores de textos, al desarrollo de informes, folletos, presentaciones de diapositivas, estadísticas en planillas de cálculos, blogs, uso de ofimática en la nube, experimentación de aplicaciones de realidad aumentada y hasta alcanzar a aprender inicios de programación.

Las siguientes secciones, se van a explicar las actividades de investigación, docencia y extensión realizadas durante el primer semestre del ciclo lectivo 2017, en donde se trabajó con la temática de la Memoria en el BIM 3, un establecimiento ubicado a cuadras de la Facultad de Informática, donde se emplazaron las Facultades de Psicología y Humanidades, que fue testigo de la detención, tortura y desaparición de personas durante la última dictadura militar.

3 Actividades de Investigación

En el marco del proyecto, se programó una visita a la Facultad de Humanidades y Ciencias de la Educación. La misma tenía un doble objetivo, por un lado, que los niños y jóvenes de sectores vulnerables puedan conocer la entidad y las carreras que en ella se dictan y, por otro lado, que puedan aprender sobre la historia reciente, relacionada con el golpe de estado más cruel y violento que sufrió la Argentina, en los años 1976-1983. Es importante destacar que tanto la Facultad de Humanidades como la de Psicología de la UNLP, se encuentran emplazadas en el predio donde funcionaba el BIM 3, uno de los tantos centros clandestinos de detención activos durante la dictadura militar.

La Facultad de Humanidades y Ciencias de la Educación, a través del proyecto "Estudiar en el nuevo predio de la FaHCE, un sitio cargado de sentidos" desde 2015 y durante el comienzo del ciclo lectivo, realiza recorridos destinados a los alumnos ingresantes a la entidad, con el objetivo de acercarlos a la historia y las memorias del predio. A partir del conocimiento de esta actividad, el equipo de trabajo que integra el programa de inclusión informática, se pone en contacto con la Facultad de Humanidades y coordina una visita para un grupo de 40 alumnos. Para el caso del proyecto de investigación, la visita programada resultaba diferente en relación con las edades de los destinatarios con los que estaban acostumbrados a trabajar. Esto provocó que revisaran la forma de realizar el recorrido para que resultara acorde a las edades del grupo. Como consecuencia de la coordinación se programó la visita para el mes de mayo.

Desde el lado de Informática, se propuso la visita a dos de las asociaciones civiles con las que se trabaja: Centro Verde Esperanza (Fundación Tau)³ y, Fundación El Roble⁴. Entre ambas entidades se trabaja con alrededor de 70 alumnos, de los cuales, por cuestiones de nivel de madurez para entender la temática a abordar, fueron invitados a la visita, alumnos mayores de 10 años.

3.1 Preparativos para la Actividad

Una vez coordinada la visita, miembros del equipo encargado de realizar los recorridos brindó material bibliográfico para que los docentes pudieran trabajar temáticas tan complejas, con alumnos del nivel primario y secundario. De esta forma, se pudo plantear y desarrollar, junto con las asociaciones civiles, un abordaje previo

³ <http://www.fundaciontau.org.ar/index.php/centro-verde-esperanza>

⁴ <https://es-la.facebook.com/elroble.asociacioncivil>

de la temática[7] [8]. Con dicha información y con recursos informáticos se preparó material para trabajar en las diferentes capacitaciones.

En el marco de la capacitación informática, se les comentó a los alumnos la visita que iban a realizar y, se les planteó la actividad posterior que iban a realizar. La misma iba a ser una nota periodística sobre la información recabada en la actividad.

Se destinaron dos clases para interiorizar a los alumnos sobre la historia del predio, donde se encuentra emplazada la entidad educativa a visitar. Para esto los docentes, brindaron información sobre el lugar, colaboraron con los alumnos en la búsqueda, a través de la Web, de información específica y, les plantearon la escritura de ciertos datos relacionados.

El trabajo previo tuvo como objetivo que los alumnos pudieran llegar a la visita con un conocimiento previo, aprovechando mejor la actividad y, además pudieran hacer uso de las recursos tecnológicas TICs (presentaciones, editores de textos, uso de navegadores, entre otros) aprendidos a lo largo de las capacitaciones dadas.

La actividad comenzó con la muestra de fotos de la Facultad que se iba a visitar donde, en una de las mismas mostraba el cartel del BIM 3, como se visualiza en la Fig.1. Eso fue un disparador para que los alumnos indagaran a través del navegador que significaban esas siglas y, pudieran llegar a decir que correspondían al "Batallón de Infantería Marina 3". En la búsqueda de información, los alumnos notaron la asociación del nombre del batallón con los centros clandestinos de detención. Eso llevó a que los mismos continuaran indagando y profundizando sobre lo que había sucedido en ese lugar.

Se finalizó la actividad escribiendo algunas dudas que quedaron de la investigación realizada.



Fig. 1. Diapositiva perteneciente al material realizado para comenzar con la historia del predio.

4 Desarrollo de la actividad

Un grupo de 40 alumnos de las dos asociaciones civiles Centro Verde y El Roble fueron los que realizaron la visita. La misma se diagramó en dos espacios: en la

Facultad de Informática y en la Facultad de Humanidades. El primer espacio estuvo pensado para enseñar a los alumnos lo que es el trabajo colaborativo y, preparar en forma conjunta una serie de preguntas a llevar a la visita de Humanidades. El segundo espacio para llevar a cabo la visita programada.

4.1 En la Facultad de Informática: Aprendiendo a trabajar colaborativamente

Para la enseñanza del concepto sobre el trabajo colaborativo, con recursos informáticos, se procedió a preparar una serie de veinte cuentas de correo con nombres de usuario que iban desde grup1.proyecto hasta grup20.proyecto. La creación de las cuentas, por parte del equipo de trabajo, se debió a que la mayoría de los alumnos eran menores de edad.

En la sala de PC de la Facultad, se les entregó, a los alumnos, información sobre la cuenta de correo que usarían para desarrollar la actividad y, se procedió a comenzar con la explicación sobre la temática. A continuación, en la Fig.2, se puede observar parte de la presentación que se realizó para presentar el concepto.

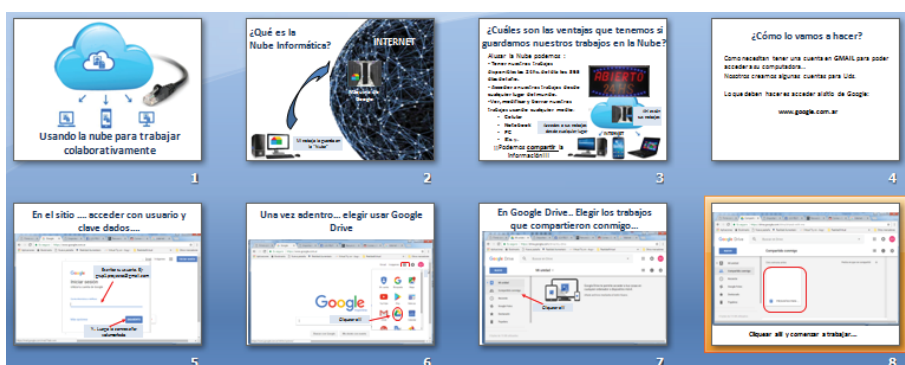


Fig. 2. Material explicativo de trabajo colaborativo realizado por el equipo de trabajo del proyecto "El barrio va a la Universidad"

Una vez explicado el concepto del trabajo colaborativo, se comenzó a trabajar sobre un documento que se había armado y compartido con todos los grupos. La idea fue escribir entre todos, las preguntas o dudas sobre la temática que se iba a abordar.

En un comienzo la tarea no estuvo muy organizada, dado que el entusiasmo de los alumnos fue tan grande que hacía que algunos sobrescribieran lo que otros habían escrito. Para dar solución a ese problema, los docentes agregaron, al comienzo de las líneas del documento, el número del grupo que debía escribir en ella. Esa decisión permitió que el trabajo pudiera continuar organizadamente. En la Fig.3 se puede observar el resultado final de la producción del documento.

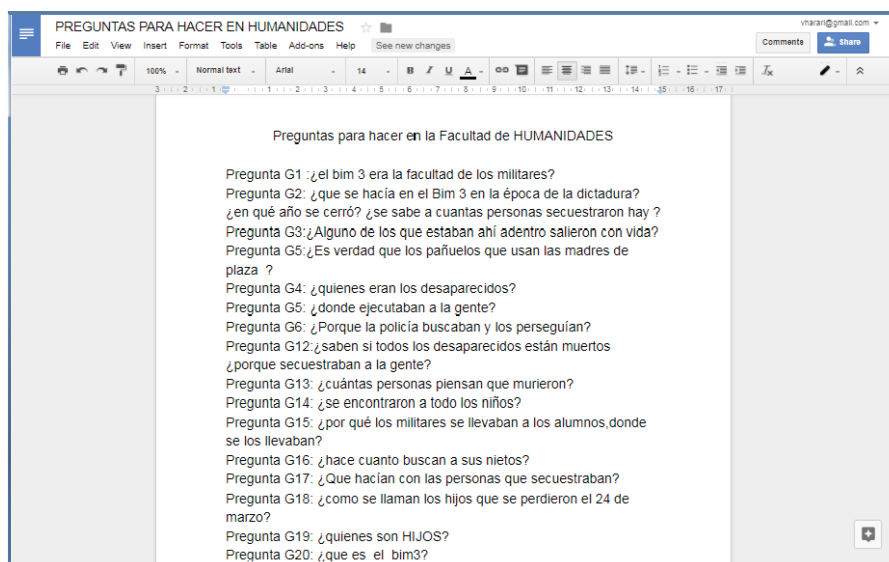


Fig. 3. Preguntas realizadas colaborativamente por 20 grupos de alumnos de entre 10 y 13 años

Se imprimieron unas copias del documento para que los alumnos los llevaran a la Facultad de Humanidades junto con un par de tabletas que iban a servir para documentar la visita. La misión final de los alumnos era la de escribir una nota periodística de la actividad realizada.

4.2 En la Facultad de Humanidades y Ciencias de la Educación: Aprendiendo sobre el pasado reciente (Recorrido)

En Enero del 2014, la Facultad de Humanidades y Ciencias de la Educación se mudó al predio en dónde funcionó el Batallón de Infantería número 3 de la Armada, hasta 1999. Como detallamos al comienzo, durante la última dictadura cívico militar (1976-1983) allí funcionó un Centro Clandestino de Detención, Tortura y Exterminio que mantuvo en cautiverio a centenares de trabajadores y trabajadoras, militantes y estudiantes detenidos por la Fuerzas de Tareas 5, dependiente de la Armada. Con el objetivo de reponer la historia y memorias del predio, el recorrido comenzó en el cartel de señalización que indica que allí se cometieron crímenes de lesa humanidad. En esa primer "parada", las docentes de la Facultad de Humanidades comenzaron reponiendo brevemente algunos puntos del pasado reciente para situar a los alumnos en la comprensión de los hechos. Con preguntas disparadoras, se los interpeló sobre lo sucedido durante la última dictadura, sus conocimientos a cerca de la existencia de centros de detención de personas y sobre porqué eran "clandestinos". Los alumnos se sentaron en ronda, con la intención de fomentar un diálogo horizontal entre sí y con las docentes.

Luego de este primer momento, en dónde surgieron debates interesantes en torno a la historia argentina y sus significados, la persecución con motivos políticos y el

cercenamiento de los derechos humanos a quienes militaban por una sociedad diferente, nos dispusimos a continuar la visita. La segunda “parada” se produjo frente a las “huellas” que quedaron del viejo edificio. Desde el lugar seleccionado, se pueden visibilizar muros permitrales, garitas, un portón original con su puesto de vigilancia y letras de gran tamaño con la sigla BIM. Las docentes consultaron qué les llamaba la atención y si habían notado esas “marcas” al ingresar. La intención de esta parada fue reponer al predio como un espacio de memoria y remarcar que aún cuando allí funcionen las facultades de Psicología y Humanidades, la idea de preservar huellas del pasado se vincula con una política orientada a reflexionar sobre el mismo. En este punto, se enfatizó en la importancia que el espacio sigue teniendo para la justicia, para los sobrevivientes y testigos y como sostén de la memoria colectiva. Las docentes interpelaron a los alumnos sobre si estaban de acuerdo con este tipo de políticas y si consideraban correcto que allí funcionen facultades. Muchos asintieron y resaltaron la importancia de que hoy se eduque en lugares en los que antes se asesinaba o torturaba.

El recorrido continuó por detrás de los edificios de la facultad de Psicología, para que los chicos pudieran apreciar el muro perimetral, las garitas y los alambres de púa y electricidad que continúan allí desde aquellos tiempos. De esta forma, volvimos a los edificios de Humanidades y nos detuvimos frente a una placa de mármol con los nombres de los detenidos-desaparecidos docentes, no docentes y estudiantes vinculados a ambas facultades. Allí, los alumnos sacaron fotos, tomaron nota de algunos nombres y escucharon sobre el valor simbólico de ese recordatorio, que fue una política de memoria construida por la Facultad de Humanidades y aún hoy sigue construyéndose, a medida que nuevos familiares y víctimas acercan nombres de personas que están desaparecidas y estuvieron detenidas en el BIM 3.

Finalmente, la última “parada” se dio en el Jardín de la Memoria, ubicado entre los tres edificios que componen a Humanidades. Los alumnos se sentaron en ronda, escucharon el relato de las docentes y realizaron preguntas, según el material que anteriormente habían elaborado. Las talleristas comentaron que ese espacio era un homenaje a los desaparecidos vinculados a la facultad y señalaron una plantación de árboles que florecen en Marzo, fecha simbólica respecto del terrorismo de Estado. También comentaron los murales que recorren las paredes de la facultad y señalaron que los muros son “territorios” en los que se manifiestan disputas de memoria. Con este fin, comentaron que las diferentes organizaciones políticas y estudiantiles, podían realizar sus pintadas en espacios designados, realizando homenajes políticos sobre el pasado, pero también sobre temas del presente. Finalmente, mencionaron la búsqueda de las Abuelas de Plaza de Mayo y enfatizaron sobre el caso del hijo/a de Norma Raggio (estudiante de Psicología que pasó por el BIM 3) y Raúl Alberto Balbuena, que debió haber nacido para noviembre de 1976. Resaltaron la necesidad de comprender que al trabajar esta temática, no estamos hablando sólo del pasado: hay más de 400 nietos y nietas que aún hoy están “desaparecidos/as” por no conocer su real identidad.

A modo de cierre, se produjo una conversación entre los alumnos y las docentes en la que aquellos pudieron indagar aún más sobre lo visto en la visita y comentaron sus impresiones.

5. Producción periodística

Con la información recabada en la Facultad de Humanidades, los alumnos debieron continuar con el trabajo en los diferentes comedores.

Los mismos debieron armar una nota periodística para ser publicada en medios como la página Web de la asociación a la cual pertenecen.

Para el armado de la nota, se contó con un material brindado por docentes y alumnos de la Facultad de Periodismo, que integran el Área de Comunicación (ACI) del Laboratorio LINTI de la Facultad de Informática. El mismo informaba "¿Cómo construir una nota?". A continuación, en la Fig.4, se puede observar parte del material brindado, que permitió realizar la tarea.

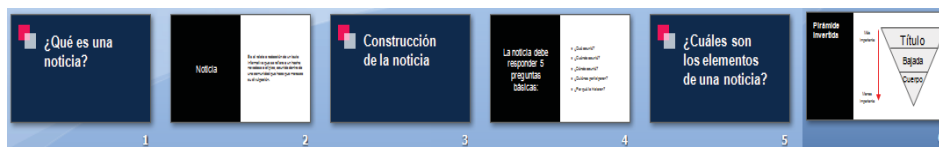


Fig.4. Material sobre cómo realizar una nota periodística.

En lo que respecta al cuerpo de la nota, los alumnos trabajaron colaborativamente sobre un documento que fueron armando entre todos. En la misma, se expuso información sobre la historia del predio con fotos alusivas y se describieron algunas de las carreras que hoy día se estudian en la entidad.

Una vez finalizada la producción, se compartió el documento con una de las integrantes del grupo de investigación de la Facultad de Humanidades que tuvo a cargo del recorrido, para que pueda visar lo que se había escrito. La misma realizó un comentario que los alumnos debieron corregir.

A continuación se puede observar la producción final, que se muestra en la Fig.5.



Fig.5. Producción final de la nota realizada por los alumnos de Centro Verde Esperanza.

6. Conclusiones

En este artículo se hizo un recorrido sobre las diferentes actividades de extensión realizadas en el marco del proyecto El Barrio va a la Universidad, en donde se enseña las tecnologías informáticas aplicadas a diferentes temáticas disciplinares.

Dado que las capacitaciones se mantienen a lo largo de los años, el utilizar las aplicaciones de uso general en forma continua, para reforzar las diferentes temáticas abordadas, permite que los alumnos vayan adquiriendo implícita y progresivamente habilidades relacionadas con el buen uso de las tecnologías. No solo el uso particular de la aplicación sino también el uso combinado de las mismas.

El acercamiento hacia las distintas facultades, el codeo con profesionales y docentes de distintas áreas de investigación, la conformación de propuestas interdisciplinares permitieron que los niños de los comedores barriales no solo afiancen sus conocimientos informáticos con producciones aplicadas a temáticas variadas y de suma importancia, sino que se familiaricen con el ámbito universitario que el día de mañana puede serle propio.

Puntualmente se describió un trabajo multidisciplinario sumamente importante y sensible como lo es la propuesta sobre la Memoria en BIM 3, donde los alumnos recorrieron sus instalaciones, investigaron sobre el tema, descubrieron marcas o evidencias de dicha historia en uno de los centros clandestinos de detención más importantes de la ciudad.

En este sentido se detalla cómo la extensión y docencia se articulan con trabajos de investigación en pos de mejorar cada vez más la calidad de la propuesta extensionista, profundizando un vínculo formal, comprometido y de excelencia, cada vez más estrecho entre los comedores barriales y la Universidad.

Referencias

1. Díaz, J., Banchoff, C., Harari, V., Harari, I. y Ambrosi, V.: Accesibilidad, Brecha Digital y Medio Ambiente: Líneas estratégicas en la Facultad de Informática de la UNLP. Memorias de Universidad 2012. ISSN 978-959-1614-34-6. Habana, Cuba (2012).
2. Díaz, J., Harari, V. y Harari, I.: University project: The marginalized neighborhood goes to University. ICEER Proceedings, International Conference on Engineering Education and Research. ISBN 9789954909126. Morocco, Marrakech (2013).
3. Díaz, J., Banchoff, C., Harari, V., Harari, I. y Ambrosi, V.: Articulating two social based informatics projects. Proceedings Invoice for ICIT-2012, International Conference on Informatics Technologies. ISSN 1553-9911. Saratov, Rusia (2012).
4. Harari, V. y Harari, I.: Trabajando la inclusión universitaria en sectores marginados de la sociedad. Memorias del VI Congreso Nacional de Extensión Universitaria, Universidad Nacional de Rosario. Rosario (2014).
5. Díaz, F., Harari, V. y Harari, I.: Realidad Aumentada en prácticas educativas de índole social. Actas del Congreso TE&ET. ISBN 978-987-3977-30-5. Buenos Aires (2016).
6. Sitio herramienta de programación Scratch. <https://scratch.mit.edu>
7. Inés Dussel. A 30 años del golpe: Repensar las políticas de la transmisión en la escuela. FLACSO. 2008. Argentina
8. Pensar la dictadura: Terrorismo de estado en la Argentina. Preguntas, respuestas y propuestas para su enseñanza. Ministerio de Educación de la Nación. 2010. Argentina

Formulación de un Examen Óptimo

Enrique E. Tarifa^{1,2}, Sergio L. Martínez¹, Samuel Franco Domínguez¹ y Jorgelina F. Argañaraz³

¹Facultad de Ingeniería, Universidad Nacional de Jujuy

²Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)

³Facultad de Humanidades y Ciencias Sociales, Universidad Nacional de Jujuy
4600 San Salvador de Jujuy, Argentina

{eetarifa, smartinez, sfrancodominguez}@fi.unju.edu.ar
jfarganaraz@hotmail.com

Resumen. El objetivo del presente trabajo es formular un examen óptimo para una materia dada. Para ello, primero se modela la probabilidad de que un estudiante apruebe el examen en función de la cantidad de unidades que evalúa el profesor. El modelo de simulación que se presenta en este trabajo se desarrolla realizando un análisis probabilístico. Luego, se define un examen óptimo como aquel que adjudica la nota que merece el estudiante, y que puede tomarse respetando las limitaciones de tiempo y esfuerzo que afectan al docente; en consecuencia, en un examen óptimo, aprueban los que merecen aprobar, y desaprueban los que no merecen aprobar. En base a esta definición, empleando el modelo de simulación, se formula un modelo de optimización del tipo INLP que determina la cantidad de unidades a evaluar que maximiza la probabilidad de tomar un examen óptimo.

Palabras clave: Evaluación académica, Análisis probabilístico, Optimización

1 Introducción

La evaluación es un tema crítico en cualquier institución, y particularmente en las educativas. Según Frola y Velásquez [1], el proceso de evaluación implica la obtención de información, la emisión de juicios una vez procesada la información, y la consecuente toma de decisiones tendientes a la mejora de los procesos y servicios. La evaluación puede ser cualitativa o cuantitativa, siendo la primera preferida para evaluar el aprendizaje, mientras que la segunda es elegida para medir los conocimientos que al final de un periodo el estudiante retiene.

A pesar de la importancia que tiene la evaluación, es una cuestión que no ha sido todavía adecuadamente resuelta en el ámbito educativo, con el consiguiente impacto negativo sobre la formación de los estudiantes. En efecto, Trillo Alonso y Porto Currás [2] analizaron la percepción que tenían los estudiantes sobre la evaluación en la Facultad de Ciencias de la Educación de la Universidad de Santiago de Compostela durante el curso 1997-1998, y concluyeron que, para los estudiantes, la evaluación no cumple con los objetivos que debería tener. Frente a este resultado, los autores plantean

que, si esto ocurre en la Facultad de Ciencias de la Educación, no es probable un escenario mejor en otras facultades. Por otra parte, los resultados son más preocupantes si se considera que dichos estudiantes serán futuros profesionales de la educación.

Las TIC, además de influir en otros aspectos de la educación, también influyen en la forma de evaluar, abriendo nuevas posibilidades al automatizar las correcciones, calcular índices estadísticos, realizar histogramas [3]. Estas funciones favorecen una mejor evaluación; pero no brindan una solución acabada; es más, abren nuevos temas. Por ejemplo, los cuestionarios de opciones múltiples, al poder ser contestados empleando el azar, plantean el problema de corregir la nota del cuestionario considerando la posibilidad de que el estudiante haya contestado probando suerte [4].

Huapaya *et al.* [5] plantean que, para realizar una evaluación justa, además de la nota de un examen, se deben considerar otros aspectos: promedios de notas del estudiante, promedio de la clase, evolución de las notas del estudiante. Estos datos son procesados por un sistema experto con lógica difusa para diagnosticar el nivel de conocimiento del estudiante. El empleo de un sistema experto tiene la ventaja de quitar subjetividad a la evaluación, produciendo evaluaciones uniformes; pero está limitado por el conocimiento de los expertos y el proceso de adquisición del conocimiento [6].

Un aspecto importante poco investigado de las evaluaciones es el rol del profesor. El profesor, al diseñar la evaluación, toma varias decisiones: la cantidad de preguntas a realizar, los temas cubiertos por las preguntas, la nota de aprobación (si no está fijada por la institución). Las decisiones que el profesor tome en esta etapa causan un profundo impacto en el resultado de la evaluación. Por ese motivo, el presente trabajo tiene como objetivo analizar los efectos que dichas decisiones tienen sobre la evaluación. Para ello, se define un examen óptimo como aquel que adjudica la nota que merece el estudiante, y que puede tomarse respetando las limitaciones de tiempo y esfuerzo que afectan al docente; en consecuencia, en un examen óptimo, aprueban los que merecen aprobar, y desaprueban los que no merecen aprobar. En base a esta definición, se formula un modelo de optimización que determina la cantidad de unidades a evaluar que maximiza la probabilidad de tomar un examen óptimo.

2 Modelo de Simulación

2.1 Formulación del Problema

El problema inicial que se aborda en este trabajo consiste en estimar la probabilidad que tiene un estudiante que rinde un examen de una materia compuesta por UM unidades de evaluación —en adelante llamadas unidades de la materia— cuando el estudiante estudia UE unidades y el profesor toma UT unidades. El examen se aprueba con una nota igual o superior a NA , con nota máxima igual a UM . Las unidades de evaluación de la materia representan el grado de detalle en el que el profesor desglosa la materia para realizar la evaluación. En grado de detalle creciente, las unidades de evaluación pueden ser unidades del programa de la materia, temas de la materia, subtemas de la materia. La respuesta de un estudiante a una unidad de evaluación puede ser correcta o incorrecta, no se consideran grados intermedios. Se supone que todas las

unidades de evaluación tienen el mismo grado de dificultad tanto para estudiarlas como para evaluarlas.

2.2 Solución Analítica

Dada la formulación planteada, para determinar la probabilidad de que el estudiante conteste correctamente una cantidad particular u de preguntas, se puede aplicar la distribución hipergeométrica [7], una distribución discreta relacionada con muestreos aleatorios sin reemplazo. Efectivamente, para el problema que se está analizando, se tiene una población de UM elementos que pertenecen a dos categorías: las unidades que el estudiante sabe y las unidades que no sabe; UE unidades pertenecen a la primera, y $UM-UE$ unidades pertenecen a la segunda. Definidas esas categorías, la distribución hipergeométrica permite calcular la probabilidad de obtener u elementos de la primera categoría en una muestra sin reemplazo de UT elementos de la población original; es decir, la probabilidad $P(U = u)$ de que el estudiante conteste bien u preguntas de las UT realizadas por el profesor:

$$Pu(u) = \frac{\binom{UE}{u} \binom{UM-UE}{UT-u}}{\binom{UM}{UT}} \quad (1)$$

La nota que obtendrá ese estudiante será:

$$N(u) = \text{Redondeo} \left(NM \frac{u}{UT} \right) \quad (2)$$

Se deduce también que el valor mínimo de U es:

$$U_{\min} = \max(0, UT - (UM - UE)) \quad (3)$$

mientras que el valor máximo que puede alcanzar U es:

$$U_{\max} = \min(UE, UT) \quad (4)$$

A partir de la distribución probabilística de U , se deduce la probabilidad PA de que el estudiante apruebe la evaluación al obtener una nota $N(u)$ igual o superior a NA :

$$PA = \sum_{u=U_{\min}}^{U_{\max}} H(N(u) - NA) Pu(u) \quad (5)$$

donde $H(\cdot)$ es la función escalón:

$$H(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases} \quad (6)$$

Por otra parte, la probabilidad de que el estudiante obtenga la nota n en el examen $P(N = n)$ se calcula de la siguiente manera:

$$PN(n) = \sum_{u=U_{\min}}^{U_{\max}} H(N(u)-n)H(n-N(u))Pu(u) \quad (7)$$

es decir, que la probabilidad de n se obtiene sumando las probabilidades de los u con $N(u) = n$.

Finalmente, se define el valor de la nota justa NJ , como la nota que merece el estudiante que estudió UE unidades:

$$NJ = \text{Redondeo} \left(NM \frac{UE}{UM} \right) \quad (8)$$

Si esta nota es igual o superior a NA , el estudiante merece aprobar; de lo contrario, no merece aprobar.

3 Caso de Estudio

Para el análisis que sigue, se adopta como caso de estudio una materia con $NM = 10$, $NA = 4$ y $UM = 10$. En este caso, NJ será igual a UE , por lo que la cantidad mínima que puede estudiar el estudiante para merecer aprobar es 4 unidades.

4 El Examen Óptimo

Como se planteó en la introducción del presente trabajo, se define como examen óptimo a aquel que adjudica la nota que merece el estudiante, y que puede tomarse respetando las limitaciones de tiempo y esfuerzo que afectan al docente (tiempo de realización del examen, tiempo de corrección, etc.); en consecuencia, en un examen óptimo, aprueban los que merecen aprobar, y desaprueban los que no merecen aprobar. Definido de esta forma el examen óptimo, para un estudiante que estudió UE unidades, se pueden plantear las siguientes posibles funciones objetivo particulares que deberán ser maximizadas:

- Probabilidad de la nota justa: El estudiante obtendrá la nota que merece. Esta es una medida difícil de satisfacer, ya que se debe asignar al estudiante una de las $NM+1$ categorías posibles (al asignarle una nota de 0 a NM); y es más difícil mientras mayor es NM :

$$f_1(UE) = P(NJ) \quad (9)$$

- Probabilidad de aprobación justa: El estudiante aprobará si merece aprobar, y desaprobará si no merece aprobar. Esta es una medida más relajada que la anterior debido a que considera sólo dos categorías, aprobado y desaprobado:

$$f_2(UE) = \begin{cases} 1 - PA & NJ < 4 \\ PA & NJ \geq 4 \end{cases} \quad (10)$$

Las funciones objetivo particulares planteadas miden qué tan óptimo es el examen tomado a un tipo particular de estudiante: el que estudió UE unidades. Para considerar a todos los estudiantes, se debe plantear una función objetivo global independiente de UE . Dos funciones objetivo globales posibles que cumplen con esa condición son las siguientes:

$$FO_1(UT) = \sum_{ue=0}^{UM} Pb(ue) f_1(ue) \quad (11)$$

$$FO_2(UT) = \sum_{ue=0}^{UM} Pb(ue) f_2(ue) \quad (12)$$

donde $Pb(ue)$ es la probabilidad de que un estudiante estudie ue unidades $P(UE = ue)$. Es decir, que las funciones objetivo globales planteadas son promedios probabilísticos de las funciones objetivo particulares analizadas anteriormente.

En este trabajo, se supone que $Pb(ue)$ obedece una distribución binomial:

$$Pb(ue) = \binom{UM}{ue} p^{ue} (1-p)^{UM-ue} \quad (13)$$

donde p es la probabilidad de que un estudiante estudie una unidad dada. Mientras más alta p , más dedicados son los estudiantes. Para el caso de estudio, se toma $p = 0.5$. La Fig. 1 presenta cómo varían las dos funciones objetivo globales para ese caso.

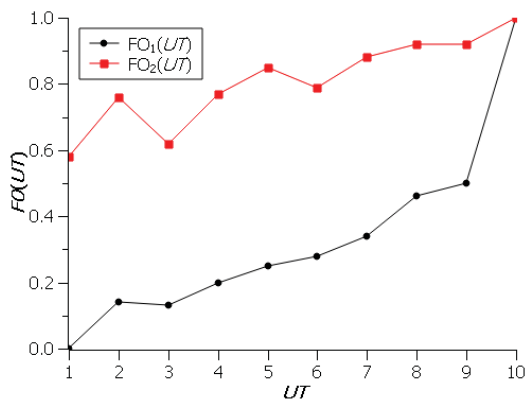


Fig. 1. Funciones objetivo globales el caso de estudio, con $NM = 10$, $NA = 4$, $UM = 10$ y $p = 0.5$.

Las funciones objetivo globales presentadas deben ser maximizadas para alcanzar un examen óptimo. Para ello, la única variable de decisión que dispone el profesor es UT . En la Fig. 1, se observa que los dos criterios planteados se cumplen cuando $UT = UM$. Sin embargo, esa solución tiene el mayor costo (duración del diseño del examen, duración del examen, duración de la corrección). Para que el costo de la evaluación no sea demasiado alto, es conveniente establecer la siguiente restricción:

$$UT \leq UT_{\max} \quad (14)$$

donde UT_{\max} es la cantidad máxima de unidades que el profesor puede o está dispuesto a evaluar. Para el caso de estudio, se toma $UT_{\max} = UM/2 = 5$.

En la Fig. 1, puede apreciarse que el valor de UT que cumple con la restricción planteada anteriormente y que maximiza ambas funciones objetivo es 5. Entonces, para el caso de estudio, lo recomendable es evaluar cinco unidades. En esa figura, también, puede apreciarse que al ser la segunda función objetivo más relajada que la primera, tiene valores más altos y es menos sensible a UT , lo que significa que es más probable que un examen sea óptimo si se realiza con la segunda función. Finalmente, es importante notar que, si bien ambas funciones objetivo presentan una tendencia creciente, no son monótonas crecientes; lo que hace posible que en algunos casos puedan existir soluciones no triviales al problema de optimización planteado.

4.1 Modelo de Optimización con Restricción UT

Para resolver el problema para cualquier UM y p , se planea el siguiente modelo de optimización INLP:

$$\begin{aligned}
 & \underset{UT}{\text{Max}} FO(UT) \\
 & \text{s. a:} \\
 & UT \leq UT_{\max} \\
 & UT \in \mathbb{N}
 \end{aligned} \tag{15}$$

Debido a que la región factible es pequeña, este modelo se resolvió por búsqueda exhaustiva empleando una planilla Microsoft Excel.

La solución producida por este modelo garantiza el máximo valor de $FO(UT)$, pero no el mínimo valor de UT . Para el caso de existir múltiples soluciones, la solución deseada es la de UT mínimo. Para encontrar este valor, se resuelve un segundo problema de optimización INLP:

$$\begin{aligned}
 & \underset{UT}{\text{Min}} UT \\
 & \text{s. a:} \\
 & FO(UT) \geq FO_{\text{opt}} \\
 & UT \in \mathbb{N}
 \end{aligned} \tag{16}$$

donde FO_{opt} es el máximo valor de la función objetivo reportado por el primer problema de optimización.

La Fig. 2 presenta los resultados correspondientes a $p = 0.5$ y $UT_{\max} = UM/2$. Los valores $UT1_{\text{opt}}$ fueron obtenidos con $FO(UT) = FO_1(UT)$; mientras que los $UT2_{\text{opt}}$, con $FO(UT) = FO_2(UT)$. En dicha figura, puede apreciarse que la cantidad UT no siempre alcanza el máximo valor permitido por la restricción, con lo que se obtiene el resultado inesperado que a veces menos preguntas produce una mejor evaluación (llegando al extremo de 3 unidades evaluadas menos que la cantidad máxima permitida). Adicionalmente, al ser la segunda función objetivo más relajada que la primera, los valores óptimos de $UT2_{\text{opt}}$ son menores o iguales a $UT1_{\text{opt}}$, produciendo una evaluación de menor costo.

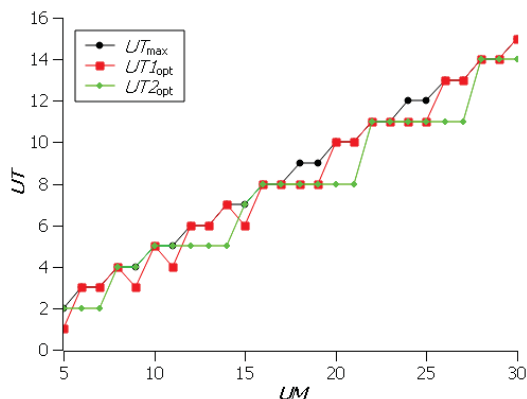


Fig. 2. Valores óptimos de UT para FO_1 y FO_2 con $NM = 10$, $NA = 4$ y $p = 0.5$.

4.2 Modelo de Optimización con Tolerancia

En la Fig. 1 se puede apreciar que no existe una gran diferencia entre $FO_1(2)$ y $FO_1(5)$, ni tampoco entre $FO_2(2)$ y $FO_2(5)$. Debido a esto, si el profesor tiene una cierta tolerancia, la nueva cantidad práctica de unidades a evaluar UT_{pra} puede ser menor a la UT_{opt} recomendada en la sección anterior.

Con este nuevo parámetro de tolerancia, el problema de optimización se descompone en dos problemas secuenciales. Primero, se debe resolver el primer problema de optimización planteado en la sección anterior, que determina el valor máximo de la función objetivo FO_{opt} . Luego, con ese valor, se debe resolver el siguiente problema de optimización INLP:

$$\begin{aligned}
 & \text{Min}_{UT} UT \\
 & \text{s. a:} \\
 & FO(UT) \geq FO_{opt} - Tol \\
 & UT \in \mathbb{N}
 \end{aligned} \tag{17}$$

donde $Tol \in [0, 1]$ es la tolerancia, o merma en la probabilidad de realizar un examen óptimo, que acepta el docente.

Aplicando este modelo al caso de estudio ($NM = 10$, $NA = 4$, $UM = 10$, $NJ = NE$, $p = 0.5$ y $UT_{max} = 5$) con $Tol = 0.1$, se obtiene que la cantidad práctica de unidades a evaluar es 2 en lugar de 5 para la segunda función objetivo global, lo que es una apreciable reducción del costo de la evaluación.

La Fig. 3 y la Fig. 4 presentan los resultados para valores de Tol igual a 0.1 y 0.2, respectivamente, con $p = 0.5$ y $UT_{max} = UM/2$. Los valores $UT1_{pra}$ fueron obtenidos con $FO(UT) = FO_1(UT)$; mientras que los $UT2_{pra}$, con $FO(UT) = FO_2(UT)$. En dichas figuras, puede apreciarse que la cantidad UT está muy por debajo de la cantidad

permitida y de la óptima, y esto se logra sin afectar demasiado la calidad de la evaluación.

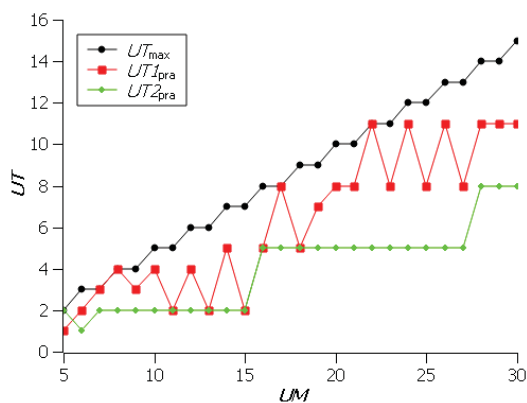


Fig. 3. Valores prácticos de UT para FO_1 y FO_2 con $NM = 10$, $NA = 4$, $p = 0.5$ y $Tol = 0.1$.

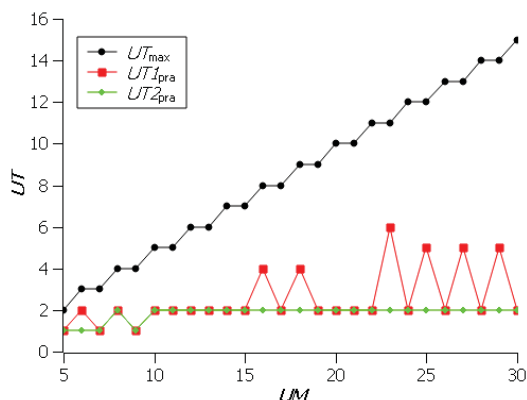


Fig. 4. Valores prácticos de UT para FO_1 y FO_2 con $NM = 10$, $NA = 4$, $p = 0.5$ y $Tol = 0.2$.

4.3 Efecto de la Nota de Aprobación

En el estudio realizado, se supuso que la nota de aprobación NA era igual a 4, que es el estándar de aprobación en algunas universidades. Sin embargo, otras universidades, o las cátedras, adoptan diferentes notas de aprobación de acuerdo a la instancia que están evaluando. Por ejemplo, una cátedra puede adoptar la nota 5 para aprobar un parcial, 7 para promocionar y 4 para aprobar el final. Teniendo en cuenta esta situación, se realizaron pruebas para analizar el efecto que tiene modificar NA . La modificación de la nota de aprobación modifica a $FO_2(UT)$ y, por lo tanto, también modifica a $UT_{2_{opt}}$ y $UT_{2_{pra}}$. No obstante, en las pruebas realizadas, si bien se observan cambios en $UT_{2_{pra}}$ para $Tol = 0.2$ cuando se varía NA de 4 a 7, la diferencia es a lo sumo una unidad de evaluación.

4.4 Efecto de la escala

Cuando se incrementa la nota máxima NM a 100 y NA a 40, se observa un marcado deterioro de $FO_1(UT)$, mientras que $FO_2(UT)$ permanece casi inalterada. Este deterioro se debe a que, al aumentar la escala, las notas posibles también aumentan, y es menos probable que la nota del examen concuerde con la nota justa NJ . Por lo planteado, tanto $UT1_{pra}$ como $UT2_{pra}$ adoptan valores menores o iguales a 2 con $UT_{max} = 5$ y $Tol = 0.2$.

5 Conclusiones

En este trabajo, se definió un examen óptimo, y en base a esta definición, se formuló un modelo de optimización que determina la cantidad de unidades a evaluar que maximiza la probabilidad de tomar un examen óptimo. Este modelo se resolvió para diferentes casos, y se encontró que la cantidad óptima de unidades a evaluar no siempre fue la máxima permitida. Luego, se planteó un segundo modelo de optimización que permitió resolver el problema con una cierta tolerancia. Al aumentar la tolerancia que acepta el profesor, la cantidad de unidades a evaluar se redujo notablemente. También, se analizaron los efectos de modificar la mínima nota que necesita el estudiante para aprobar, y se encontró que no tiene un efecto importante sobre la cantidad de unidades que deben evaluarse para lograr un examen óptimo. Por otra parte, modificar la nota máxima del examen reduce notablemente la probabilidad de asignar una nota justa al estudiante.

Finalmente, del estudio realizado, se concluye que a veces menos preguntas producen una mejor evaluación, y que es más probable realizar un examen óptimo cuando se califica como aprobado o desaprobado, en lugar de asignar una nota.

Referencias

1. Frola, P., Velásquez, J.: Competencias docentes para la evaluación cuantitativa del aprendizaje. Centro de Investigación Educativa y Capacitación Institucional, D. F. México (2011)
2. Trillo Alonso, F., Porto Currás, M.: La percepción de los estudiantes sobre su evaluación en la universidad. Un estudio en la facultad de ciencias de la educación. *Innovación Educativa*, 9, 55-75 (1999)
3. Álamo Serrano, J.: Nuevas posibilidades de evaluación usando las TIC's: un vistazo a cuatro casos. En: *La evaluación de los estudiantes en la Educación Superior*, Servei de Formació Permanent, 54-73. Universidad de Valencia, Valencia (2007)
4. González-Santander, J.L., Martín, G.: Análisis de la fórmula para la calificación de pruebas tipo test multi-respuesta. *Nereis. Revista Iberoamericana Interdisciplinaria de Métodos, Modelización y Simulación*, 3, 53-59 (2010)
5. Huapaya, C.R., Lizarralde, F.A., Arona, G.M.: Modelo basado en Lógica Difusa para el Diagnóstico Cognitivo del Estudiante. *Formación Universitaria*, 5(1), 13-20 (2012).
6. Bojadziev, G.: *Fuzzy Sets, Fuzzy Logic, Applications (Series on Advances in Mathematics for Applied Sciences)*. World Scientific Publishing Company, New York (1996)
7. Walck C.: *Hand-book on STATISTICAL DISTRIBUTIONS for experimentalists*, University of Stockholm, Stockholm (2007)

Guitarreando: Estudo sobre a utilização de modelagem¹ e tablatura no ensino de violão mediado por tecnologia

Ana Elisa Bonifácio Barros¹, Damián Keller¹, Leandro L. Costalonga²,

¹NAP, Universidade Federal do Acre. ² Universidade Federal do Espírito Santo
{alaudista@hotmail.com, dkeller@ccrma.stanford.edu, leandro.costalonga@googlemail.com}

Resumo. O presente projeto estuda a efetividade da modelagem e da tablatura no ensino do violão mediado por tecnologia, utilizando procedimentos experimentais com o objetivo de subsidiar a eleição, o design e o desenvolvimento de ambientes musicais educativos. Dentro deste contexto, 41 participantes realizaram exercícios musicais através de três meios digitais: vídeo, áudio e tablatura. Foram modeladas atividades de execução de acordes utilizando cordas soltas, de melodias com cordas soltas usando somente uma mão, e de melodias empregando cordas pisadas que demandam a coordenação motora de ambas mãos. Através do estudo, respondemos às seguintes perguntas: Quais modelos digitais – vídeo e áudio ou tablatura – são mais efetivos para o ensino inicial de violão? Quais materiais sonoros (acorde, melodia com cordas presas ou melodia com cordas soltas) são mais adequados? Em relação aos critérios número de acertos e tempo de execução, os resultados indicam que a tablatura é mais efetiva que os modelos digitais disponibilizados em formatos audiovisuais ou sonoros. Quanto aos materiais musicais, os resultados sugerem o uso de atividades empregando melodias com cordas soltas (arpejos) nas fases iniciais.

Abstract. We focus on the effectiveness of modeling and tablature as strategies for knowledge sharing in technologically based guitar-training techniques. We adopt experimental procedures with the purpose of subsidizing election, design and development of educational musical environments. Forty-one participants performed musical excerpts furnished through three digital media: video, audio and tablature. Activities featured chords using loose strings, melodies with

¹ Do original em inglês *modeling*, que significa: a “apresentação (ao vivo ou gravada) de algo a ser imitado posteriormente pelo observador” (Madsen et al. 1975, p. 271). A instrução por meio da modelagem consiste segundo Dickey (1992) na alternância de demonstrações e processos imitativos entre professores e estudantes. O professor utiliza um instrumento musical, voz ou meios eletrônicos para proporcionar o modelo, e os estudantes respondem com seus instrumentos ou vozes. Uma mínima comunicação verbal é necessária neste processo, mas a instrução é predominantemente não verbal.

loose strings, and melodies using pressed strings that require coordination between the hands. The study proposes the following questions: What digital models – video and audio or tablature – are most effective for initial guitar methods? What sonic materials (chord, melody with pressed strings or melody with loose strings) are more suitable? Regarding the number of errors and the execution time, the results indicate that tablature is more effective than audiovisual and sonic models. Concerning musical materials, the results suggest the use of activities using loose strings during the initial training phases.

Keywords: Education and technology, Music and technology, Music education, Instrumental education.

1 Introdução

Guitarreando é um projeto que busca conhecer o impacto da modelagem e da tablatura no ensino de violão (*guitarra* em espanhol) mediado por tecnologia. Investigamos qual suporte para transmissão do conhecimento musical – modelo digitalizado (vídeo e áudio) ou de representação simbólica (tablatura) - é mais efetivo, utilizando para tanto os parâmetros – número de acertos e tempo de execução – como critério para a aferição da sua efetividade.

A descoberta do Sistema de Neurônios Espelho (SNE) reforça a proposta de uso da imitação como ferramenta no processo de ensino instrumental musical. No campo da música, vários autores vêm realizando estudos sobre a teoria das células espelho aplicada a execução instrumental (Hallett et al. 1994; Sirigu et al. 1995; Stephan et al. 1995; Lotze et al. 1999; Gerardin et al. 2000; Grezes and Decety 2001; Jeannerod 2001; Kimberley et al. 2006). Os experimentos indicam que as áreas do cérebro ativadas durante a performance também são ativadas durante a visualização motriz (Viegas, 2012). Como o cérebro exercita a ação por observação, se deduz que é possível praticar música através da observação de exemplos musicais. As pesquisas de Haslinger (2005) e Lahav et al. (2007) sobre o SNE – ambas com execução musical com pianistas – chegaram a conclusões muito importantes para o ensino instrumental: 1 – existia uma ativação significativa da zona temporal pré-frontal quando os sujeitos observavam outros pianistas tocando. 2 – Ao observar vídeos sem áudio de execução de pianistas, se ativam as áreas do cérebro responsáveis pela audição.

As pesquisas de Lahav (2007), Helding, (2010) e Overy e Molnar-Szakacs (2009) indicam que a imitação e a sincronização dentro da experiência musical promovem a aprendizagem em música. Diversos estudos sugerem que a modelagem é uma estratégia efetiva de ensino musical (Davidson, 1989; Dickey, 1991; Rosenthal, 1984; Rosenthal, 1988; Sang, 1987; Davidson, 1989; Dickey, 1991; Linklater, 1997; Lisboa et al. 2005; Frewen, 2010; Madsen, 2000). A modelagem ativa o SNE ajudando os estudantes de música com o treino da execução instrumental inclusive quando não estão executando música.

O presente trabalho apresenta os resultados e a análise da efetividade do uso de diversos meios digitais (vídeo, áudio e tablatura) e materiais musicais (acorde,

melodia com cordas soltas e melodia com cordas presas) utilizando os critérios tempo de execução e número de acertos durante as atividades. Os resultados servem para embasar a escolha de técnicas de suporte no design instrucional, acrescentado base empírica para um campo em rápida expansão.

2 Metodologia:

2.1 Perfil dos sujeitos:

Os sujeitos de Guitarreando tem duas características principais: 1. São maiores de 18 anos. Focamos no público adulto, já que a maioria dos cursos de violão mediados pela tecnologia contempla esse público. 2. Não possuem nenhum conhecimento de música ou de execução de violão. O conhecimento violonístico anterior do sujeito poderia demandar atividades muito diversas com difícil mensuração dos resultados. Não ter conhecimento foi a estratégia adotada para controlar essa variável.

Os sujeitos foram distribuídos em três grupos e cada grupo foi vinculado a um protótipo. O grupo 1 trata da aprendizagem por vídeo, o grupo 2 da aprendizagem por áudio e o grupo 3 da aprendizagem por tablatura. Cada grupo tem três tarefas: acorde, melodia com cordas soltas e melodia com cordas presas ou pisadas (tabela 1). Cada sujeito realiza as atividades em apenas uma sessão, utilizando somente um protótipo.

Quantidade de sujeitos por grupo			
Vídeo	Áudio	Tablatura	Total
15	12	14	41

Idade:

A idade dos sujeitos varia de 18 até 58 anos. Com uma média de 28,7 anos, e um desvio padrão de 9,71.

Gênero:

O sexo dos sujeitos é predominantemente feminino em todos os protótipos. A média total do experimento foi de 63,24% de participação feminina e 36,76% masculina.

Grau de instrução:

A média geral dos sujeitos é: grau superior incompleto (49,30%), ensino médio completo (31,3%), superior completo (14,6%), e pós-graduação (4,8%). Sendo que parte do experimento foi realizada dentro da universidade, predominam os sujeitos com grau superior incompleto.

Ocupação:

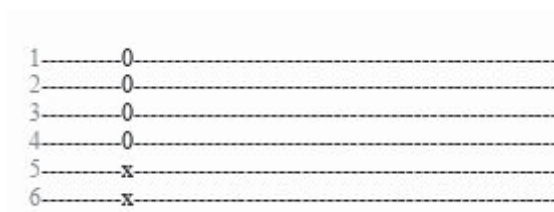
Os sujeitos do experimento apresentam ocupações diversas, sendo a maioria estudantes universitários.

2.2 Protótipos e materiais musicais

Definimos um protótipo como o conjunto de materiais musicais disponibilizados através de um modelo (vídeo / áudio) ou de notação em formato digital (tablatura). Cada protótipo inclui três tarefas: execução de um acorde, de uma melodia tocada com cordas soltas (usando somente uma mão – direita ou esquerda dependendo se o sujeito é destro ou canhoto) e de uma melodia tocada com cordas presas (usando ambas as mãos). A notação tradicional (partitura) não foi trabalhada neste projeto pois seriam necessárias aulas prévias de teoria musical, o que não foi possível por motivos de tempo, espaço e orçamento. Para padronizar as tarefas, todos os exercícios têm um total de quatro notas, o mesmo número de notas em todas as atividades.

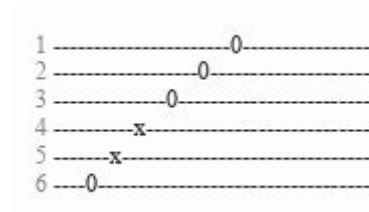
Acorde: Optou-se pela elaboração de acordes de fácil execução porém com suficiente complexidade como para aproximar a demanda cognitiva da tarefa aos outros exercícios propostos. Foi utilizado o acorde de mi menor com sétima no baixo (Em/D). O acorde foi gravado em vídeo e áudio e foi escrito em tablatura.

Figura 1. Acorde em tablatura.



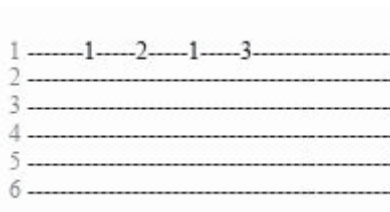
Melodia – Cordas soltas: Elaborou-se uma melodia simples usando as notas da tonalidade de mi menor (Em). A figura 2 apresenta a atividade em tablatura.

Figura 2. Melodia cordas soltas.



Melodia – Cordas presas (pisadas): Desenhou-se uma melodia com quatro notas, todas tocadas na primeira corda (para facilitar a execução), utilizando somente as três primeiras casas do violão, sendo duas notas repetidas.

Figura 3. Melodia cordas presas.



2.3 A eleição dos formatos

Foram escolhidos os formatos AVI (siglas em inglês de Audio Video Interleave) e WAVE (Waveform Audio File Format). Esses formatos são compatíveis com diversas plataformas e não possuem compressão, garantindo a qualidade visual e sonora. Para a representação simbólica da informação musical (tablatura) foi adotado o formato PDF (Portable Document Format) pela facilidade de uso em qualquer plataforma sem que se modifique o aspecto ou a estrutura do documento original (Adobe, 2017).

2.4 Procedimentos

Foram coletados dados no período de fevereiro até abril de 2017. Os lugares de coleta foram: Universidade Federal do Acre, Escola Acriana de Música, e em poucos casos a residência dos sujeitos (todos no Brasil). Participaram quarenta e um (41) **sujeitos divididos em três grupos. 15 que receberam as informações por vídeo, 12 por áudio e 14 por tablatura.**

- **Entrega, leitura e assinatura do TCLE (Termo de Consentimento Livre e Esclarecido) exigido pela pró-reitoria de pesquisa da UFAC.**
- **Orientações sobre: procedimentos de estudo, recursos tecnológicos, e no caso do protótipo 3 também foram fornecidas orientações resumidas sobre leitura da tablatura.**
- **Coleta de dados do perfil dos participantes por meio de formulários.**
- **Execução das tarefas.**
- **Registro do tempo cronometrando durante a execução.**
- **Gravação em vídeo do produto final da execução musical.**
- **Preenchimento dos formulários com respostas sobre níveis de dificuldade de cada tarefa e sobre as preferências dos sujeitos.**

Duração do experimento: 1 (um) encontro com duração variável para cada sujeito, dependendo do tempo necessário para executar as tarefas. Cada sujeito pôde utilizar o material quantas vezes quisesse.

2.5 Avaliação dos resultados

Avaliamos o suporte para transmissão de conhecimentos musicais (vídeo, áudio e tablatura) e os materiais musicais (acorde, melodia cordas soltas e presas) através do número de acertos e do tempo de execução. O procedimento de cálculo foi realizado em cinco etapas.

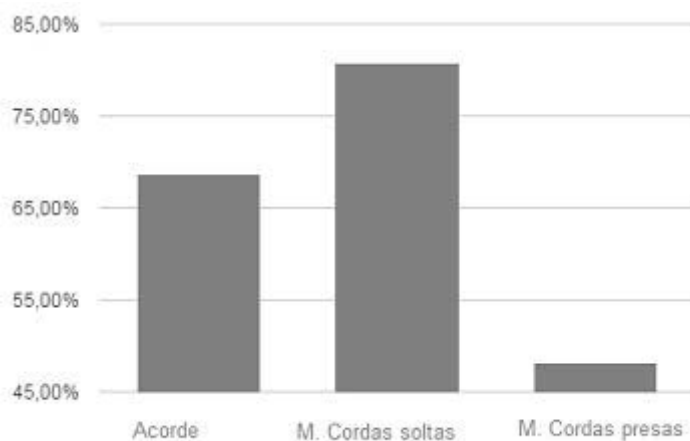
- Avaliação dos resultados para cada material musical (acorde, melodia com cordas soltas e melodia com cordas presas).
- Cálculo de acertos e duração da tarefa para cada material.
- Soma dos resultados para os três materiais contidos em um protótipo, configurando assim a média do protótipo.
- Comparação entre as médias de cada protótipo.
- Análise do resultado final.

3 Resultados:

3.1 Materiais – Número de acertos:

O resultado geral mostra que a melodia com cordas soltas é o material com maior número de acertos. Na análise de todos os protótipos, a média foi de 81% de acertos. Acorde é o segundo melhor colocado com 69%. Melodia com cordas presas é o pior avaliado com 48%.

Gráfico 1. Média de acertos – Materiais



3.2 Materiais – Tempo de execução:

Na análise geral dos protótipos os resultados mostram diferenças no tempo investido nos protótipos.

- 1) Melodia com cordas soltas com 0,84 minuto.
- 2) Acorde com 1,09 minuto.
- 3)

Melodia com cordas presas com 1,4 minuto.

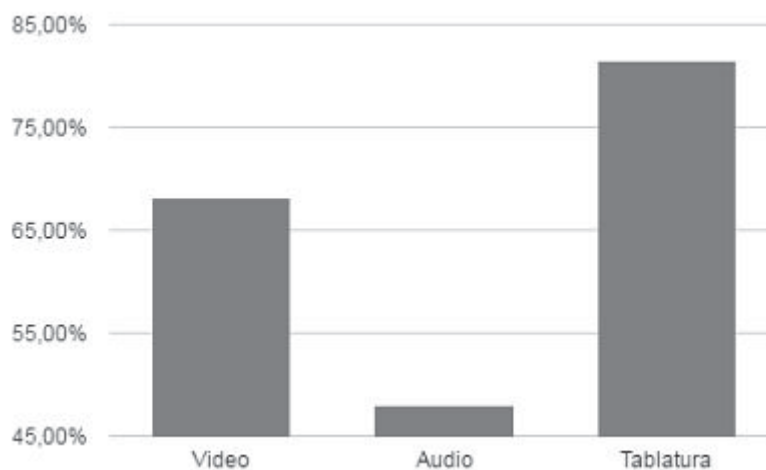
3.3 Análise do número de acerto e do tempo de execução dos materiais musicais:

Em todos os protótipos, cordas soltas foi o material com maior número de acertos (81%) e com menor tempo de execução da tarefa (0,84 minuto). O segundo material melhor avaliado foi o acorde, com a média de 69% de acertos, e tempo médio de execução de 1,09 minuto. O material pior avaliado foi a melodia com cordas presas com uma média de 48% de acertos e de 1,4 minuto de duração. Tal resultado dá suporte para os métodos de ensino de violão clássico para iniciantes que começam suas atividades por notas individuais de cordas soltas (Filho, 1980; Pinto, 1978; Sageras, 1936). Os piores resultados – para a melodia com cordas presas – podem estar relacionados ao grau de dificuldade do material. É necessária a coordenação motora de ambas as mãos para a execução, enquanto que o restante dos materiais utilizados necessitam apenas do uso de uma mão.

3.4 Meio de transmissão – Número de acertos:

Os dados mostram que o protótipo 3 teve o maior número de acertos (81,53%) contra 68,22% do protótipo 2 e 47,91% do protótipo 1. Em relação aos materiais musicais estudados (acordes, melodia com cordas soltas e melodia com cordas pisadas), a tablatura foi o meio de transmissão mais efetivo para o ensino de violão com mediação tecnológica para o público sem conhecimento musical (gráfico 2).

Gráfico 2. Meio de transmissão – Média de acertos



3.5 Meio de transmissão – Tempo de execução:

A menor média de tempo de execução corresponde à tablatura (0,73 minuto). A execução dos outros protótipos demandou 1,73 minuto para o vídeo (protótipo 1) e 0,86 minuto para o áudio (protótipo 2).

3.6 Análise do número de acertos e do tempo de execução dos meios de transmissão:

A tablatura (protótipo 1) é o melhor suporte em relação ao número de acertos e ao tempo necessário para a realização da tarefa. O vídeo é o segundo melhor em porcentagem de acertos 43% contra 35% do áudio. O áudio demandou menos tempo para execução da tarefa (0,99 minuto) do que o vídeo (2,4 minutos). O tempo de execução reduzido pode ser resultante de dois fatores: 1. A rápida decodificação da tarefa, ou 2. a extrema dificuldade em decodificá-la, resultando na desistência e portanto diminuindo o tempo de execução. Temos como hipótese em relação à segunda colocação do protótipo 2 (áudio) que não foi motivada pela velocidade da decodificação da informação, mas porque o nível de dificuldade do protótipo resultou numa baixa disposição para completar a tarefa.

4. Conclusão

Guitarreando se constitui em um experimento exploratório que objetiva contribuir com a área de ensino instrumental mediado por tecnologia e na compreensão sobre a utilização de modelos por esta área. O presente projeto teve como objetivo avaliar a eficiência da transmissão de conhecimento musical através de modelos em suporte digital (vídeo e áudio, e tablatura) dentro do contexto de ensino de violão mediado pela tecnologia.

Limitações: Guitarreando é um projeto de tarefas introdutórias e pontuais de violão. Todas as respostas foram obtidas imediatamente após a execução. Portanto não é possível saber quais respostas teríamos em outro contexto de desenho instrucional, seja envolvendo um curso extenso ou um nível mais avançado de ensino de violão.

Nas pesquisas anteriores sobre modelagem no ensino instrumental que comparam os modelos transmitidos por vídeo, áudio e grupos de controle (sem uso de modelos) observou-se que a estratégia de ensino por vídeo com áudio foi a mais efetiva (Zurcher 1975; Rosenthal, 1984; Linklater, 1997). Os resultados obtidos em Guitarreando mostram porém que a tablatura é mais efetiva do que os outros modelos nos critérios número de acerto e tempo de execução. É necessário portanto realizar novos experimentos que comparem o uso da tablatura com o uso de vídeo e áudio de forma integrada. O resultado indica que o emprego da tablatura em cursos de violão mediados por tecnologia tem grande potencial. Atualmente a tablatura é pouco usada e em alguns casos é menosprezada. Outro fator a levar em consideração é que a tablatura, por ter menor quantidade de símbolos permite uma ação mais rápida e acertada na execução do instrumento (Barros, 2006). No entanto, ela não é a melhor

opção para a compreensão teórica da música, por isso não deveria ser usada como elemento único de ensino mas como parte integrada de materiais diversos disponibilizados para os estudantes.

Outro resultado obtido no projeto Guitarreando é que a melodia com cordas soltas teve melhor rendimento em relação aos outros materiais (acorde e melodia com cordas presas) em todas as formas de transmissão (vídeo, áudio, e tablatura). Isso dá suporte para os métodos de iniciação de violão clássico que começam suas lições com exercícios de cordas soltas. Por outro lado indica que iniciar com acordes - como ocorre no ensino de violão popular - talvez acrescente dificuldades desnecessárias nas primeiras lições.

References

1. Adobe 2017. O que é PDF? Disponível em: <https://acrobat.adobe.com/pt/pt/why-adobe/about-adobe-pdf.html>. Acesso em: 1 de julho de 2017.
2. Davidson, L. 1989. Observing a Yang-Chin Lesson, Learning by Modeling and Metaphor. *Journal of Aesthetic Education* 23 (1): 85-99.
3. Dickey, M. R. (1991). A comparison of verbal instruction and nonverbal teacher-student modeling in instrumental ensembles. *Journal of Research in Music Education*, 39(2), 132-142
4. Dickey, M. R. (1992). A review of research on modeling in music teaching and learning. *Bulletin of the Council for Research in Music Education*, 27-40.
5. Filho, O. G. R. (1980) *Minhas Primeiras Notas ao Violão*. Rio de Janeiro: Irmãos Vitale.
6. Frewen, K. G. 2010. Effects of Familiarity with a Melody Prior to Instruction on Children's Piano Performance Accuracy. *Journal of Research in Music Education* 57 (4): 320-333.
7. Grezes, J. y J. Decety. 2001. Functional anatomy of execution, mental simulation, observation, and verb generation of actions: A meta-analysis. *Human Brain Mapping* 12 (1): 1-19.
8. Gerardin, E., A. Sirigu, S. Lehericy, J. B. Poline, B. Gaymard, C. Marsault, Y. Agid, e D. Le Bihan. 2000. Partially overlapping neural networks for real and imagined hand movements. *Cerebral Cortex* 10 (11): 1093-1104.
9. Hallett, M., J. Fieldman, L. G. Cohen, N. Sadato, y A. Pascual-Leone. 1994. Involvement of Primary Motor Cortex in Motor Imagery and Mental Practice. *Behavioral and Brain Sciences* 17 (2): 210-210.
10. Haslinger, B., Erhard, P., Altenmüller, E., Schroeder, U., Boecker, H., y Ceballos-Baumann, A. O. (2005). Transmodal sensorimotor networks during action observation in professional pianists. *Journal of cognitive neuroscience*, 17(2), 282-293.
11. Jeannerod, M. 2001. Neural simulation of action: A unifying mechanism for motor cognition. *Neuroimage* 14 (1): 103-109.
12. Kimberley, T. J., G. Khandekar, L. L. Skraba, J. A. Spencer, E. A. Van Gorp, e S. R. Walker. 2006. Neural substrates for motor imagery in severe hemiparesis. *Neurorehabilitation and Neural Repair* 20 (2): 268-277.

13. Lahav, A., E. Saltzman, e G. Schlaug. 2007. Action representation of sound: Audiomotor recognition network while listening to newly acquired actions. *Journal of Neuroscience* 27 (2): 308-314.
14. Linklater, F. (1997). Effects of audio-and videotape models on performance achievement of beginning clarinetists. *Journal of Research in Music Education*, 45(3), 402-414.
15. Lisboa, T.; Williamon, A.; Zicari, M. y Eiholzer, H. (2005). Mastery through imitation: A preliminary study. *Musicae Scientiae*, 9(1), 75-110.
16. Lotze, M., P. Montoya, M. Erb, E. Hulsmann, H. Flor, U. Klose, N. Birbaumer, e W. Grodd. 1999. Activation of cortical and cerebellar motor areas during executed and imagined hand movements: An fMRI study. *Journal of Cognitive Neuroscience* 11 (5): 491-501.
17. Madsen, C., Greer, R. y Madsen, C. H. 1975. *Research in music behavior: Modifying music behavior in the classroom*. New York: Teachers College Press.
18. Madsen, C. K. 2000. Research in music teaching and learning. *Revista de la Lista Electrónica Europea de Música en la Educación* (5): 1-7.
19. Overy, K., y I. Molnar-Szakacs. 2009. Being Together in Time: Musical Experience and the Mirror Neuron System. *Music Perception* 26 (5): 489-504.
20. Pinto, H. (1978). *Iniciação ao violão*, Vol. 1. São Paulo: Ed. Ricordi.
21. Rosenthal, R. K. 1984. The Relative Effects of Guided Model, Model Only, Guide Only, and Practice Only Treatments on the Accuracy of Advanced Instrumentalists Musical Performance. *Journal of Research in Music Education* 32 (4): 265-273.
22. Rosenthal, R. K., M. Wilson, M. Evans, e L. Greenwalt. 1988. Effects of Different Practice Conditions on Advanced Instrumentalists Performance Accuracy. *Journal of Research in Music Education* 36 (4): 250-257.
23. Sagreras, J. S. (1936). *Las primeras lecciones de guitarra*. Ricordi Americana.
24. Sang, R. C. (1982). Modified path analysis of a skills-based instrumental effectiveness model for beginning teachers in instrumental music education. Tesis de doctorado. University of Michigan.
25. Sirigu, A., L. Cohen, J. R. Duhamel, B. Pillon, B. Dubois, Y. Agid, y C. Pierrot de Seilligny. (1995). Congruent Unilateral Impairments for Real and Imagined Hand Movements. *Neuroreport* 6 (7): 997-1001.
26. Stephan, K. M., G. R. Fink, R. E. Passingham, D. Silbersweig, A. O. Ceballos-Baumann, C. D. Frith, e R. S. J. Frackowiak. 1995. Functional-Anatomy of the Mental Representation of Upper Extremity Movements in Healthy-Subjects. *Journal of Neurophysiology* 73 (1): 373-386.
27. Zurcher, W. (1975). The Effect of Model-Supportive Practice on Beginning Brass Instrumentalists. C. K. Madsen; R. D. Greer; C. H. Madsen Jr. (Edt). *Research in Music Behavior: Modifying Music Behavior in the Classroom*. 131-138. New York: Teachers College Press.

Impact of information and communication technologies on teaching-learning processes in children with special needs autism spectrum disorder.

Mónica. R. Romero¹, Javier Diaz², Ivana Harari²

¹ Posgrado Tecnología de Información de la Comunicación, Universidad Internacional SEK, Campus Miguel de Cervantes. Alberto Einstein, s/n. Quito - Ecuador

² LINTI. Facultad de Informática, Universidad Nacional de la Plata
Calle 50 y 120, 1900 La Plata. Buenos Aires,

¹ monica.romero@uisek.edu.ec, ² jdiaz@info.unlp.edu.ar, ² iharari@info.unlp.edu.ar

Abstract. This research determines the impact of the use of information and communication technologies (ICTs) in students with special educational needs (SEN) in the specific case of children diagnosed with autism spectrum disorder (ASD) in their academic performance. Its objectives were to determine the educational benefits of ICTs and to evaluate their use. The previous premises were approached as a starting point by collecting bibliographical material to conceptualize the variables to be treated. Subsequently, a sample was taken for convenience, with a descriptive, analytical - synthetic tendency; surveys were developed and applied to teachers and students. After the application of the instruments was evidenced as a problem, the lack of adequate teaching practices of ICT use, especially in children with ASD, was able to confirm the hypothesis, proving that more than sixty percent of the students improve their academic performance through teaching-learning processes that add.

Key words: Autism, Education, Teaching, NEE, Processes, ICTs, SEN.

1 Introduction

Students with special educational needs have certain conditions that make it difficult to abstract the concepts analyzed in class and generally require greater accommodations in the teaching and learning methods used. It is where the information and communication technologies play a key role in teaching as a differentiating element in the construction of innovative processes.

During the present academic year 2016-2017 in the SEK Educational Unit in the primary section it works with students with special educational needs diagnosed with mild autism spectrum disorder trying to know:

- The impact of the use in classes of the technologies of the information and the communication.
- The IT tools used in the teaching-learning process.
- Academic progress achieved in elementary students with special educational needs TEA children using ICT S as a tool in strengthening school activities

As an essential objective, this work seeks to analyze the use of ICTs, since it is considered that the use of TICs can be a primordial strand in the assimilation of knowledge.

In addition, it will be investigated in the determination of the incidence of the use of TICs in the school performance of students and in the identification of the educational benefits they provide, primarily in the students mentioned above.

The research work is organized in four sections; through the study of art will define the special needs of education, briefly explained is autism spectrum disorder and what are its characteristics, a tour is made to explain how information and communication technologies essentially in recent years have contributed to innovative processes of education and inclusion, through experimental research can demonstrate the established hypothesis, at the end of the research show the main conclusions, results achieved, and also state possible future lines of work.

2 Development

Since the definition of Kanner in 1943 and Hans Asperger in 1944, Autism Spectrum Disorder (TEA), has presented itself as a distant, strange world full of enigmas [3]

On the other hand, biological theory maintains that in the SEN there are one or several abnormalities in the brain and that these are produced by one or several biological factors such as: genes, complications during pregnancy or childbirth and viral infections. It is also stated that the medical disorders found in children and young people with autism, although they are diverse, share the trait of being associated with damages or cerebral dysfunctions[5]

ASD is a complex neurological disorder that usually lasts a lifetime. It is part of a group of disorders known as Autism Spectrum Disorders. Currently, 1 in 68 individuals and 1 in 42 boys are diagnosed with autism.[6]

On the other hand, living in this age also has its benefits, as many Information and Communication Technologies (ICTs) have been produced and developed which have had a very positive impact on the lives of people with this disorder. In fact, as some authors have commented, in recent years there has been an increased interest in learning how the teaching-learning process develops in people with ASD[7].

Within the framework of an optimistic position, computer technology and communication, plays an invaluable role, since it can provide the subject with the physical supports necessary for the development of their communicative, cognitive and affective partner potentialities; through it can achieve an approximation to three major objectives: autonomy, independence and inclusion, [8]

Also, there are technological resources based on the use of ICT specifically designed to enable the accessibility of people with disabilities. We will focus on these resources for their relevance in the subject and for the importance of disseminating them, and, above all, to boost the use of technology and make it a useful tool to improve learning through innovative pedagogical methods and techniques. [9]

The definition of special educational needs that we are promulgating in his book highlights students with SEN as those who present greater difficulties in accessing the

common learning at their age in relation to the rest of their group, noting that in most of these SEN are not only differences inherent in the student (s) who present it and cannot always be resolved by the classroom teacher without support from outside professions[20].



Fig 1: Teaching learning processes using ITCs .Source: Ministerio Educación

Author: Hospital classroom project

Despite the fact that current government policies in Ecuador promote an inclusive education that favors and encourages the inclusion of children and young people with different capacities [16]. The reality that the Children with ASD is different because ultimately those who have to take care of them within the classroom are the teachers, and they do not have the training and knowledge that this work demands.

It is of great importance to dwell on the benefits that the use of ICTs could have in acquiring the knowledge of ASD students, who have difficulty in internalizing them in various ways. This reality is evidenced as a problem for two specific conditions in our country, the first to mention is the limited access to technology that has the population, as evidenced by the website of the Guayaquil Chamber of Commerce in1:Figura1.

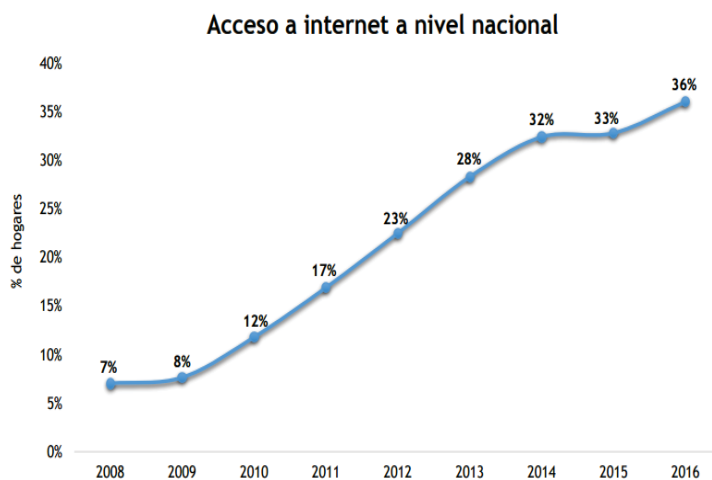


Fig. 2. Access to the internet at the national level. Source: Webpage Guayaquil Chamber of Commerce. Author: Guayaquil Chamber of Commerce

And the second and perhaps more worrisome is that there are certain teachers who are unaware of the positive impacts that can be caused by the use of ICTs, especially in students who have ASD, in order to better understand classroom content and the internalization of concepts Basic and primordial for the resolution of diverse problems. The educational institutions are in a constant process of change, they look for to offer better services to the community, to reach the excellence. The SEK Educational Unit located in the city of Quito during the academic year 2016-2017 has had students at the primary level, for this research we are in the third of basic where there are 3 students diagnosed with AS.

It is worth mentioning that the classrooms have technologies such as digital whiteboards, iPad's, internet, web page, Moodle platform, for student use; However in spite of having the mentioned technology, in certain cases it is not given an adequate use, taking advantage of the advantages that they can offer.



Fig 3: Students making use of TIC'S devices. Source: SEK Educational Unit. Author: Mónica Romero

The qualifications of students with special educational needs (SEN) are relatively low relative to those of their peers because they have certain conditions that make it difficult to abstract the concepts analyzed in class and require greater accommodations in the methods used in classes and in evaluations.

As an essential objective, this paper seeks to analyze the use of ICTs in students with ASD, because it is considered that the use of ICTs can be a primordial strand in the assimilation of knowledge. Additionally, it will be investigated in the determination of the incidence of the use of ICTs in the school performance of students with ASD and in the identification of the educational benefits of ICTs.

3 Methodology

3.1 Population

The population was chosen from the International School SEK in the same we located in the elementary students of third of basic respectively this study is formed by 56 people, divided into: 3 authorities, 10 teachers and 40 students of third of basic of the latter 3 are previously diagnosed with ASD. Five visits were made to the study center.

Table1: Study Population Source: SEK Educational Unit. Author: Mónica Romero

Fellows	Quantity
Autorities	3
Teachers of students with ASD in primary	10
Third-grade students from basic	40
SEN Students Diagnosed	3
Total	56

For this work, the sample will be used for convenience, which is a non-probabilistic sample, the participants are chosen for convenience of accessibility or proximity of the researcher, specifically the three students with special educational needs are chosen from the institution diagnosed with ASD. They are chosen because they are participants in the educational process and because they have the necessary tools to carry out the necessary implementations to carry out the use of ICTs in classrooms.

2.2 Research Design

The investigation was carried out using the procedure described below:

1. Construction of instruments.
2. Categorization of information.
3. Compilation of data.
4. Collection of information.
5. Selection of information.
6. Analysis of the data obtained.
7. Presentation of the data obtained.

General hypothesis

More than 60% of children with special educational needs improve their academic performance with the use of ICTs handled in class in the academic period 2016 - 2017.

Specific hypotheses.

1. A high percentage of elementary students with special educational needs benefit from the appropriate use of ICTs as tools in the teaching - learning process.
2. There are at least two academic progress achieved by students with special educational needs when using ICTs as a tool in strengthening school activities

Table2: Operationalization of the general hypothesis

Operationalization of the general hypothesis			
Variable	Conceptual definition	Dimensions	Indicators of measurement criteria
Use of ICTs in class	ICTs are used to support a curricular discipline or content.	Infrastructure Frequency	Connectivity Hardware IPad's Frequency of use of ICTs
Academy Performance	Qualitative and quantitative values through which there is an approximation to the evidence knowledge, aptitudes and values developed by the student in the teaching process through the ICTs	Quantitative Qualitative	Qualifications average Compliance Report

The investigative process was started with the construction of the instruments starting from the operationalization matrices of the general hypothesis, afterwards the categorization of the information was performed according to the variables.

The data collection phase revolves on the correct choice of instruments according to the previously performed categorizations, in order to carry out the collection of the information that is executed directly in the educational unit, in a personal way. The selection of the information is based on the reliability, the accuracy and the validity of the same, taking into consideration its origin.

In order to proceed with the analysis of the data obtained in a statistical manner following the analytical - synthetic method, concluding with the presentation of the data obtained to make a interrelation between them, extracting valid criteria for the verification of the hypothesis raised.

3.2 Data Collection Instruments and Techniques

As research techniques, the survey has been chosen for students with special educational needs diagnosed of the educational institution because they are considered descriptive research instruments that need to identify questions a priori To perform the selected people in a representative sample of the population, to specify the answers and to determine the method used to collect the information that is obtained

Direct observation in the classroom based on the collection of non-obstructive information regarding behaviors and processes.

In the student survey the following topics will be handled: Knowledge of the management of the technologies that the institution possesses, knowledge of special educational needs, benefit of the use of ICTs in students, frequency of use of ICTs as Tool to enhance education, academic performance, teacher training in technological

aspects and the relationship between their behavior and the use of ICTs in class. Teachers manage a survey based on the following topics: Institutional infrastructure, knowledge of the management of available technologies, students' SEN, training on ICTs and SEN, ICTs and the strengthening of school activities, frequency of use of ICTs and academic performance, benefit of ICTs in academic performance.

2.4 Data Analysis Plan

Excel software is used to make the results empty, to design the tables and graphical representations.

2.5 Procedure

The research will be descriptive, analytical - synthetic and propositional, according to because data will be gathered from students, teachers and managers about different aspects of the use of ICT in the Institution, then performing the analysis of the information collected and the measurement thereof.

It will use the quantitative methodologies to emit criteria in numerical terms that are evidenced in percentages to optimize their comprehension and qualitative because it has a social approach, where it is necessary to analyze behavioral and emotional factors of the students diagnosed with ASD, that influence in the educational process. To be used are:

Descriptive. - Data representation and characteristics of the research population, data are collected objectively and accurately, can be used averages, frequencies and other statistical information.

Analytical - synthetic. - The analytic seeks to separate the constituent elements of the phenomenon to investigate and the synthetic unites them to achieve a global understanding of the situation, seen from various perspectives.

4 Conclusions and Future Work

The information collected by the various instruments applied in the research give foundations to consider that the educational center under study has the technological infrastructure necessary to give students an avant-garde education.

Teachers have a professional training that guarantees their ideal level to perform as educators; however, there is evidence of mismatches between theoretical knowledge and teaching practice, at the time of merging special educational needs children with ASD and information technologies and communication. This requires new and precise methodological strategies to fulfill the premises of curricular adaptation.

The academic and strengthening benefits in school activities are undoubtedly enhanced by the use of technologies within the educational field. It is important to consider that within the educational community all contribute in the construction of knowledge and assimilation of the same, with more emphasis in the case of having students with special educational needs within the working group.

The academic progress is relative in each student, since having improved significantly, often due to the difficulties it presents, it moves away from the excellence in the qualifications, with the support of the authorities, tutors and teachers to generate spaces, where it recognizes these progresses and motivate the students, thus demonstrating that their efforts are valid and important for themselves and for the educational process to which they submit.

As a result of the information collected, the hypothesis is confirmed that more than 60% of students with SEN diagnosed with ASD improve their academic performance with the use of ICTs, as students, teachers and managers agree that performance was optimized 100% with the implementation of technologies in class in the academic period 2016 – 2017.

The research made, allows concluding in the following points:

- 1.- Information and communication technologies significantly benefit students with special educational needs in this case, especially children diagnosed with ASD, as it provides various adaptations required for the mastery of the academic curriculum and for the integral development of the Students, delivering to society fully capable and fully realized citizens.
- 2.- The use of ICTs in the teaching-learning process is constant in the investigated institution. Its teachers and managers are at the forefront both in infrastructure and in the training of the appropriate management of technologies in the educational field.
3. Educational progress is evident not only in the quantitative aspect of qualifications, as evidenced by the hypothesis that more than fifty percent of students improve their academic performance, but also in the qualitative aspect, in order to observe to the students certain of their capacities and connoisseurs of the uses that must give to the ICTs to meet the educational needs that they present.

As possible future lines of work we can indicate the following:

- Initiate new research in order to determine the level of self-esteem of students who are favored by the use of ICTs as tools to meet their special educational needs.
- Conduct a study proposing a methodological guide for teacher training in ICTs for children with autism spectrum with various adaptations required for the mastery of the academic curriculum, where teachers can practice using the proposed technologies and create new pedagogical strategies based on their teaching actions.
- It is important to add that studies need to be further focused on the parents or guardians of children with SEN in order to provide guides, processes, and tools to reinforce the teaching-learning processes taught in the educational units from the homes

References

1. Alonso, R. y Alonso, I. (2014) «Investigaciones recientes sobre el autismo». Madrid: Psylicom. Autismo Diario (2013).
2. Ausín, V., Abella, V., Delgado, V., & Hortigüela, D. (2016). Aprendizaje Basado en Proyectos a través de las TIC: Una Experiencia de Innovación Docente desde las Aulas Universitarias. *Formación universitaria*, 9(3), 31-38.
3. Bartak, L. y Pickering, G. (1982). *Objetivos y métodos de enseñanza*. L. Wing (Ed.), Autismo infantil.
4. Area-Moreira, M., Hernández-Rivero, V., & Sosa-Alonso, J. J. (2016). Modelos de integración didáctica de las TIC en el aula. *Comunicar: Revista Científica de Comunicación y Educación*, 24(47), 79-87.
5. Bustos, A., & Román, M. (2016). La importancia de evaluar la incorporación y el uso de las TIC en educación. *Revista Iberoamericana de evaluación educativa*, 4(2).
6. Cabero, J.; Fernández, J.M. y Córdoba, M. (2007). “Las TIC como elementos en la atención a la diversidad”, Sevilla, Eduforma, 15-35.
7. Cabero, J., Salinas, J., Duarte, A., & Domingo, J. (2000). *Nuevas tecnologías aplicadas a la educación*. Madrid, Síntesis.
8. Cabero-Almenara, J., Fernández-Batanero, J. M., & Córdoba-Pérez, M. (2016). Conocimiento de las TIC aplicadas a las personas con discapacidades. Construcción de un instrumento de diagnóstico. *Magis*, 8(17), 157.
9. Claro, M. (2011). El papel de las tecnologías de la información y las comunicaciones en la educación inclusiva.
10. Cintado, J. L. G., Bueno, C. M., Rodríguez, M. B., Bueno, E. M., de la Fuente Martínez, M. L., & Muñoz, L. G. (2016). Una “inclusiva” mirada de la tecnología en nuestro colegio. *Trabajando con una mirada especial: TEA y TIC en nuestro colegio*. *Padres y Maestros/Journal of Parents and Teachers*, (365), 41-48.
11. Coll Salvador, C., Rochera Villach, M. J., Mayordomo Saíz, R. M., & Naranjo Llanos, M. (2007). Evaluación continua y ayuda al aprendizaje. Análisis de una experiencia de innovación en educación superior con apoyo de las TIC. *Electronic Journal of Research in Educational Psychology*, 5(13).
12. Cornago, A. (2013) Los mitos sobre el autismo y los usos peyorativos que debemos evitar. Disponible en: <http://www.crianzanatural.com/art/art197.html> Echenique, L. [en línea]; Disponible en: <https://tecnoytea.com/tag/aplicaciones/> [Visitado el 15/03/2017].
13. del Moral Pérez, M. E., Martínez, L. V., & Piñeiro, M. D. R. N. (2014). Oportunidades de las TIC para la innovación educativa en las escuelas rurales de Asturias. *Aula abierta*, 42(1), 61-67.
14. Fernández, J.M. (2007). Las TICs para la igualdad. *Nuevas tecnologías y atención a la diversidad*. Sevilla, Eduforma. 225-276.
15. Fernández Sanz, L., & Rueda Bernao, M. J. (2014). Certificaciones profesionales TIC como base de la innovación educativa en dos proyectos europeos.

16. García Peñarrocha, L., & Marín Suelves, D. (2016). TIC y Educación Especial. *Didáctica, innovación y multimedia*, (34), 0001-9.
17. Haller, M., Billinghamurst, M. & Thomas, B. (2012). *Emerging technologies of augmented reality: interfaces and design*. USA: Idea Group Inc.
18. Heredero, E. S., & Carralero, A. O. (2014). Experiencias y recursos con las TIC para la atención al alumnado con necesidades educativas especiales. *Acta Scientiarum. Education*, 36(2).
19. Martínez, R. S. (2011). Percepción del profesorado sobre su capacitación en el uso de las TICs como instrumento de apoyo para la integración del alumnado con discapacidad. *Profesorado. Revista de Currículum y Formación de Profesorado*, 15(2), 299-314.
20. Martínez, J. L., Pagán, F. J. B., García, S. A., & Máiquez, M. C. C. (2016). Las tecnologías de la información y comunicación (TIC) en el proceso de enseñanza y aprendizaje del alumnado con trastorno del espectro autista (TEA). *Revista Fuentes*, (14), 193-208.
21. Moya, D. P. R., Grijalva, J. S., & Pérez, Y. G. (2016). Sistema De Comunicación Con Intercambio De Imágenes Pecs, A Través De Una Interfaz Nui Como Parte De La Terapia De Lenguaje En El Trastorno De Espectro Autista. *Revista Órbita Pedagógica*. ISSN 2409-0131, 3(1), 31-48.
22. Moreira, M. A. (2011). LOS EFECTOS DEL MODELO 1: 1 EN EL CAMBIO EDUCATIVO EN LAS ESCUELAS. EVIDENCIAS Y DESAFÍOS PARA LAS POLÍTICAS IBEROAMERICANAS1. *Revista Ibero-Americana de Educação*, (56), 49-74.
23. Nee, L. E., Caine, E. D., Polinsky, R. J., Eldridge, R., & Ebert, M. H. (1980). Gilles de la Tourette syndrome: clinical and family study of 50 cases. *Annals of Neurology*, 7(1), 41-49.
24. Palomo, R. (2013). Trastorno del espectro de autismo DSM5 criterios diagnósticos. Disponible en: <http://espectroautista.info/criteriosdiagnósticos/DSM-5> [Visitado el 15/03/17]
25. Pérez, L. (2002). Aplicaciones informáticas para personas con trastornos del espectro autista. Ponencia presentada en: I Congreso Regional "Las TIC Y EL APRENDIZAJE EN ALUMNOS CON TEA RUIZ ADAME, Necesidades Educativas Especiales: Situación actual y retos de futuro". Mérida, 16 y 30 de noviembre de 2002.
26. Prendes, C. (2015). Realidad Aumentada y educación: análisis de experiencias prácticas. *Pixel- Bit. Revista de Medios y educación*, 46, 187-2013. Recuperado, a fecha 10 /04/2017 <http://acdc.sav.us.es/pixelbit/images/stories/p46/12.pdf>
27. Ramírez, E., Domínguez, A. B., & Clemente, M. (2007). Cómo valoran y usan las Tecnologías de la Información y la Comunicación (TIC) los profesores de alumnos con Necesidades Educativas Especiales (NEE). *DIM: Didáctica, Innovación y Multimedia*, (9).
28. Salinas, J. (2004). Innovación docente y uso de las TIC en la enseñanza universitaria. *RUSC. Universities and Knowledge Society Journal*, 1(1).
29. Samaniego, P., Laitamo, S. M., Valerio, E., & Francisco, C. (2012). Informe sobre el Uso de las Tecnologías de Información y Comunicación (TIC) en la

Educación para Personas con Discapacidad. Quito: UNESCO y TRUST For The Americas.

30. Víquez, I. (2014). Las TIC en la educación inclusiva e influencia en el aprendizaje de preescolares. In Congreso Iberoamericano de Ciencia, Tecnología, Innovación y Educación (Vol. 12, No. 13, p. y14).

Material educativo hipermedial con Realidad Aumentada para la enseñanza y aprendizaje de triángulos a alumnos hipoacúsicos

SilviaValeiras¹, Natalí Salazar Mesía^{1,2,3}

¹Maestría en Tecnología Informática Aplicada en Educación. Facultad de Informática – Universidad Nacional de La Plata

² Becaria Tipo A – Universidad Nacional de La Plata

³Instituto de Investigación en Informática LIDI. Facultad de Informática – Universidad Nacional de La Plata

silviaev09@yahoo.com.ar, nsalazar@lidi.info.unlp.edu.ar

Abstract. Se presenta el material educativo hipermedial para la enseñanza de triángulos a alumnos hipoacúsicos. Este material incluye actividades con Realidad Aumentada que favorecen el aprendizaje a los alumnos con estas necesidades especiales. Actualmente la estrategia pedagógica utilizada suele ser la misma para todos los alumnos y surge la necesidad de incluir este tipo de materiales para mejorar los procesos de enseñar y aprender. Este trabajo presenta una mejora del material educativo para el Seminario “Entornos de Aprendizaje de Hipermedia. Desarrollo de Material Educativo” de la Maestría Tecnología Informática Aplicada en Educación de la Facultad de Informática de la Universidad Nacional de La Plata. Ha sido evaluado y aprobado por docentes de este seminario y en el corto plazo, se realizarán pruebas con alumnos de nivel medio para medir tanto el impacto en el aprendizaje, como también aspectos generales de diseño y usabilidad.

Keywords: Material Educativo Hipermedial, Alumnos Hipoacúsicos, Realidad Aumentada, enseñanza en nivel medio.

1 Introducción

La comunicación es la barrera principal de los alumnos hipoacúsicos y las estrategias didácticas que se usan en la actualidad no alcanzan a cubrir todas las necesidades que se presentan. Por este motivo, la utilización de materiales educativos digitales hipermediales permite abordar otras formas de comunicación, y así ayudar a mejorar los procesos de enseñanza y aprendizaje.

Del mismo modo, se incorpora Realidad Aumentada (RA) como una alternativa de solución a esta problemática. La Realidad Aumentada es una tecnología que complementa la percepción e interacción con el mundo real y permite a la persona vivenciar un entorno real aumentado, con información adicional generada por la computadora. Además, posibilita el desarrollo de aplicaciones interactivas que combinan la realidad con información sintética, tal como imágenes 3D, sonidos,

videos, textos, sensaciones táctiles, en tiempo real, y de acuerdo al punto de vista de quien está observando la escena [1].

Existen varios tipos de materiales educativos, en este trabajo se presenta el material educativo hipermedial para la enseñanza de triángulos, que incluye una actividad con Realidad Aumentada. Con este material se presenta una mejoría en la comunicación, donde los alumnos pueden alcanzar otro nivel de abstracción, y así, reforzar su sistema cognitivo y emotivo, lo que redundará en un aumento de su autoestima.

Permitirá también que puedan contextualizar una idea utilizando actividades creativas sin necesidad de que sean complejas o de muy alta elaboración.

“Las personas hipoacúsicas utilizan básicamente la visión para captar la realidad, puesto que estas personas necesitan más de la visión para su comunicación”; como lo afirma Skliar “la experiencia visual de las personas sordas incluye todo tipo de significación, representaciones y/o producciones en el campo intelectual, lingüístico, ético, estético, artístico, cognitivo, etc.”, es decir, permite el desarrollo de un lenguaje que define la importancia de participar como miembro de una comunidad.”, este material utiliza fuertemente la visualización como recurso [2][3].

Este material pretende, como expresa Brunshvieg, “Enriquecer al aprendiz con ciertos hábitos de cálculo, tratar de captar lo que puedan tener en común ciertas situaciones así como el entrelazamiento de las componentes de cada una de ellas; ejercicio de abstracción de análisis y de síntesis posiblemente más enriquecedor cuanto que no debe consistir en recitar un texto sino en construir activamente una experiencia” [4].

Como afirma Gómez “a la posibilidad de manejar los sistemas de representación se agrega el aspecto dinámico de los sistemas que le permite al sujeto manipular los objetos matemáticos y sus relaciones, construyendo una experiencia matemática difícil de vivir de otra manera” [5].

Este trabajo se enmarca dentro del Modelo instruccional de Gagné, ya que tiene como objetivo activar el proceso de aprendizaje de los alumnos, estimulando sus receptores para permitirles seleccionar y captar la información a través de la interacción con el material multimedial [6].

Se pretende realizar la descripción de este material educativo incluyendo un análisis detallado de las ventajas que presenta para el alumnado hipoacúsico.

Este artículo se organiza de la siguiente manera, de aquí en adelante: en la sección 2 se presentan antecedentes sobre los materiales educativos hipermediales y experiencias de uso de los mismo, en la sección 3 se describe la mejora realizada al material educativo hipermedial, en la sección 4 se realiza la descripción de la actividad de Realidad Aumentada propuesta, y finalmente en la sección 6 se presentan conclusiones y trabajos futuros.

2 Antecedentes

Se tomaron en cuenta varios trabajos de otros autores a la hora de diseñar este material educativo digital. A continuación se describen dos materiales educativos y las experiencias que realizaron a modo de antecedente del trabajo que aquí se presenta.

En [7][8] se presenta una experiencia para la enseñanza de programación utilizando Realidad Aumentada (RA) con el material educativo digital EPRA cuyo tema principal son las estructuras de control. Se describe una experiencia que se evalúa desde tres perspectivas: Incidencia en el aprendizaje, motivación y nivel de satisfacción. Los resultados demuestran una mejoría en la comprensión del tema en los alumnos que han utilizado EPRA. Respecto al nivel de satisfacción se considera que EPRA alcanza niveles altos en varios aspectos en cuanto a su diseño y a las actividades de RA que dispone. Cabe destacar que esta experiencia se llevó a cabo con estudiantes de primer año de la carrera Ingeniería en Computación de la Facultad de Informática de la UNLP.

En [9] se presentan aplicaciones informáticas educativas aptas para alumnos hipoacúsicos. Entre ellos hay aplicaciones para estimular el desarrollo del lenguaje, la lecto-escritura, el lenguaje de señas.

Estos antecedentes son la base de las mejoras introducidas en este trabajo así como también parte del material dispuesto por el seminario.

3 Material Educativo Hipermedial

El material educativo digital “La enseñanza de Triángulos” está destinado especialmente a alumnos hipoacúsicos. Se diseñó y desarrolló por una de las autoras¹ como trabajo final para el seminario “Tecnología Informática. Evolución y aplicaciones” de la Maestría Tecnología Informática Aplicada en Educación (TIAE) de la Facultad de Informática de la UNLP, en la Fig. 1 se puede observar la primera versión de este material [10].



Fig. 1. Pantalla principal de la primera versión de este material educativo digital

El material educativo hipermedial se realizó utilizando la herramienta de autor Ardora. Esta aplicación informática para docentes permite crear contenidos propios de un modo muy sencillo, sin tener conocimientos técnicos de diseño o programación web. Se pueden crear más de 35 tipos distintos de actividades, crucigramas, sopas de

¹ Profesora C.C. Silvia Esther Valeiras

letras, completar, etc., así como más de 10 tipos distintos de páginas multimedia: galerías, panorámicas o zooms de imágenes, reproductores mp3 o mp4, etc., así como las "páginas para servidor", anotaciones y álbum colectivo, líneas de tiempo, póster, chat, poster, sistema de comentarios y gestor de archivos, pensadas fundamentalmente para el trabajo colaborativo entre el alumnado.²

Para el trabajo final del seminario "Entornos de Aprendizaje de Hipermedia. Desarrollo de Material Educativo" de la misma Maestría, se incorporaron mejoras para el tema "Teorema de Pitágoras". Estas mejoras incluyen cambios en el diseño, navegabilidad, así como también la incorporación de una actividad de RA y un avatar. En la Fig. 2 se puede observar el cambio en la portada.



Fig. 2. Pantalla mejorada de la portada de este material educativo digital

Se destaca también la utilización de Geogebra, un tipo de software geométrico dinámico, que permite la comprensión, entendimiento, discernimiento e interpretación de los conceptos abordados.

La incorporación de un personaje (avatar) tiene como objetivo motivar al alumno, y enriquecer su interacción emocional con este personaje virtual. Se ha demostrado que la incorporación de personajes resulta atractiva para los alumnos que interactúan con el material. En el caso de este material educativo, se busca que el alumno identifique a Pitágoras como un guía para poder recorrer todo el material teórico y una ayuda en la resolución de actividades. En la Fig. 3 se visualiza este avatar.

Este personaje realiza principalmente las siguientes funciones didácticas:

- Dar un mensaje de bienvenida para todos los usuarios de este material educativo hipermedial.
- Brindar información general sobre este material educativo y su forma de recorrerlo.
- Sugerir un plan de trabajo y una secuencia de actividades para realizar en el recorrido de este material.
- Proporcionar información acerca del desarrollo de las actividades y ayudar con pistas más o menos útiles que le permitan avanzar con el desarrollo de las actividades.

² Ardora: <http://webardora.net/>



Fig. 3. Avatar incorporado en la mejora de este material

Por otro lado, en la Fig. 4 se muestra un mapa de navegación del material educativo digital, solamente la parte que se mejoró.



Fig. 4. mapa de navegación del tema a mejorar

Este material brinda retroalimentación inmediata para que el alumno pueda descubrir sus errores, analizarlos y corregirlos. Está destinado para alumnos hipoacúsicos (con sordera parcial) del nivel secundario de 2° y 3° Año. Se distribuirá entre los alumnos con el fin de profundizar, visualizar, experimentar y aplicar a situaciones problemáticas de la vida cotidiana.

La mejora del material inicial consistió en: a) utilizar una fuente y un tamaño que mejore la legibilidad, b) incorporar un diseño de fondo acorde a la temática. A su vez, las mejoras en cuanto a usabilidad constaron de: c) incluir un menú lateral izquierdo, que se puede minimizar y maximizar, con toda la funcionalidad necesaria para el recorrido de las distintas secciones (navegabilidad), d) incorporar botones de acción, e) insertar burbujas de diálogos, f) modificar el mensaje de bienvenida y agregar un mensaje final que permita volver recorrer el material. Ejemplos de estas mejoras se muestran en la Fig. 5.



Fig. 5. Mejoras incorporadas al material educativo digital.

4 Actividades de Realidad Aumentada

La Realidad Aumentada (RA) es una tecnología innovadora que ha ido evolucionando a lo largo de los años y se ha incorporado en diferentes campos de acción. En particular, en el ámbito educativo existen diferentes aplicaciones de RA que se aplican en diferentes escenarios educativos para la enseñanza de diferentes temas tanto teóricos como prácticos, en particular permite alcanzar un mayor entendimiento de conceptos abstractos como en este caso pueden resultar conceptos de Matemática como Triángulos geométricos.

Como una de las mejoras de este material educativo se describe la incorporación de esta tecnología en el desarrollo de las actividades con el fin de atender la necesidad educativa antes planteada.

Es necesario que el alumno disponga de una cámara web para realizar la actividad. En la misma se plantea una situación problemática en la que el alumno debe responder eligiendo la opción correcta entre dos opciones que se le presentan, en este caso son dos triángulos sombreados. La respuesta se proporciona a través de un marcador que se dispone dentro del mismo material educativo como se observa en la Fig. 6.



Fig. 6. Portada de la actividad de Realidad Aumentada sobre el Teorema de Pitágoras

Esta actividad utiliza la tecnología de reconocimiento del marcador para indicar la respuesta del alumno, así como también reconocimiento facial para la ubicación del

objeto triángulo en una imagen 2D. La misma se muestra como una imagen que marca los elementos de un triángulo rectángulo, como se observa en la Fig.7.

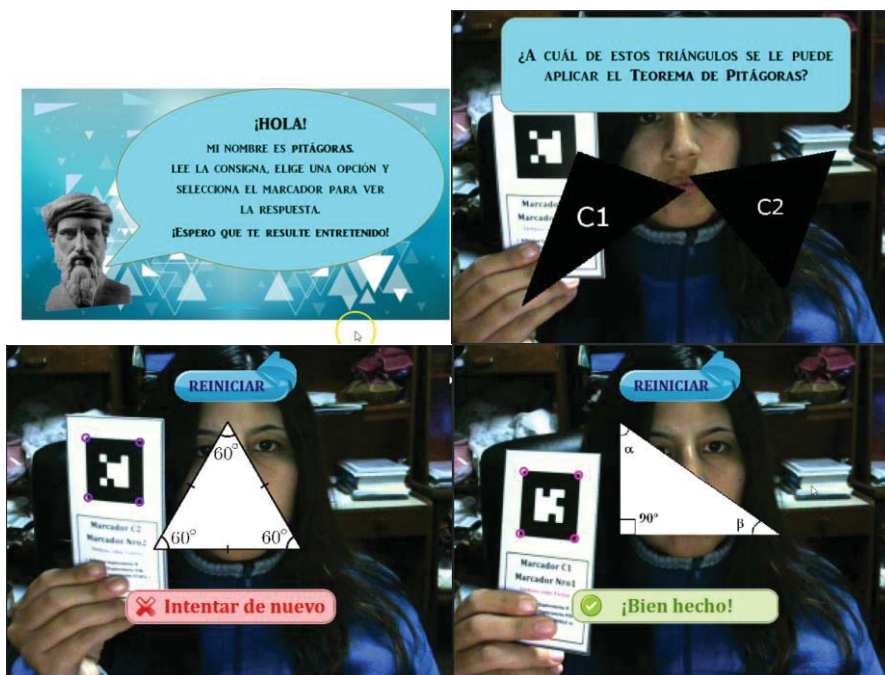


Fig. 7. Actividad de Realidad Aumentada propuesta para este material educativo. Pantalla inicial, consigna y *feedback* recibido de acuerdo a la elección del alumno

Este tipo de actividades interactivas permite que el alumno tenga un rol activo en la solución de problemas. Se clasifica como una actividad reflexiva donde el alumno debe aplicar lo aprendido en otras partes del material o en las clases teóricas.

5 Conclusiones y Trabajos Futuros

El desarrollo de este material educativo hipermedial se realizó como parte del trabajo final para el seminario “Tecnología Informática. Evolución y aplicaciones” de la carrera de postgrado Magister en Tecnología Informática Aplicada en Educación de la Facultad de Informática de la UNLP. Ha sido desarrollado por una de las autoras de este trabajo.

La mejora que se introdujo tanto en diseño, navegabilidad y usabilidad, así como también la incorporación de una actividad de RA, se realizó en el marco de otro seminario de esta maestría “Entornos de Aprendizaje de Hipermedia. Desarrollo de Material Educativo”. Este material educativo hipermedial ha sido evaluado y aprobado por los docentes de este seminario.

Con este material se busca atender la necesidad educativa que tienen los alumnos hipoacúsicos en la enseñanza de Matemática, particularmente en el tema triángulos. El uso de los colores y refuerzos en la navegabilidad del material mejoraron la usabilidad del mismo.

La incorporación de RA en una actividad para visualizar características de los triángulos rectángulos en la explicación del Teorema de Pitágoras ayuda al alumno a asimilar el concepto de una forma interactiva y permite observar desde diferentes puntos de vista los componentes del mismo permitiendo bajar el nivel de abstracción que presenta.

Se propone como trabajo futuro la inclusión de imágenes 3D en actividades de RA para la continua mejora del material así como también la planificación de la experiencia con alumnos de nivel medio para medir el impacto en el aprendizaje, la usabilidad del material y la motivación que provoca en los usuarios del mismo.

Este material educativo está actualmente disponible sólo para el uso de PC, el siguiente paso será adaptarlo para poder ser utilizado en dispositivos móviles en sus diferentes sistemas operativos.

Referencias

1. Milgram Kishino, P.; Takemura, H.; Utsumi, A.; y Kishino, F. (1994). Augmented Reality: A class of displays on the reality-virtuality continuum. *Telemanipulator and Telepresence Technologies*, 2351, 282-292.
2. Skliar, C. (1999). *Atualidade da educação bilíngüe para surdos*. Editora Mediação.
3. Balacheff, N. (2000). "Entornos informáticos para la enseñanza de las matemáticas: complejidad didáctica y expectativas". En Gregorio, N., Deulofeu, J. y Bis-hop, A. (Eds.). *Matemáticas y Educación: Retos y cambios desde una perspectiva internacional* (pp.70-88). Barcelona: Graó ICE-UB.
4. Brunschvicg. (1977). *Boletín de Matemáticas*. Vol.XI pág. 205. Universidad Nacional de Colombia. Bogotá. Año 1977.
5. Gómez P. (1997). "Tecnología y educación Matemática". *Informática Educativa* Vol 10, No. 1, UNIANDES - LIDIE, Colombia pp. 93-111, Año 1997.
6. Belloch, C. (2013). *Diseño instruccional*. Universidad de Valencia. <http://www.uv.es/~belloch/pedagogia/EVA4.pdf>, Enero 2013.
7. Salazar Mesía N., Sanz C., Gorga G. (2016) Experiencia de enseñanza de Programación con Realidad Aumentada. *Jornadas de la Enseñanza Universitaria de la Informática JENUI 2016*. Almería, España. Julio de 2016. Publicado en las Actas de las Jornadas ISBN: 978-84-16642-30-4 - Pp. 213 a 220.
8. Salazar Mesía N., Sanz C., Gorga G. (2016) Augmented Reality for Programming Teaching. Student Satisfaction Analysis. *Proceedings of the 2016 International Conference on Collaboration Technologies and Systems*. En cooperación con IEEE, ACM and IFIP. Noviembre 2016, Orlando, Florida, USA – ISBN: 978-1-5090-2300-4/16 IEEE DOI 10.1109/CTS.2016.43 – Pp. 165-171.
9. Vilches Vilela, M. Josefa.(2006). *Aplicaciones informáticas para personas con discapacidad auditiva*. CURSO: Enseñanza y comunicación del alumnado con déficit auditivo: necesidades educativas Baza, marzo de 2006. *Boletín CEAPAT* n° 51 y 52 Disponible en http://www.ceapat.org/6_centro/centro04.jsp (6/2/2006).
10. Valeiras S., Libano M. (2016). La enseñanza de triángulos y teorema de Pitágoras a alumnos hipoacúsicos. *Jornadas de Enseñanza Media Universitaria Pedes in terra ad sidera*

visus: revisando prácticas docentes: trabajos completos y resúmenes extendidos / Juan Pablo Gómez. - 1a ed. - San Miguel de Tucumán: Juan Pablo Gómez, 2016. CD-ROM, PDF ISBN 978-987-42-2294-7 – Pp. 239 a 241.

Método para generar recomendaciones personalizadas para integrar grupos de aprendizaje ubicuo y colaborativo

Margarita Alvarez¹, Silvina Únzaga¹, Elena Durán¹

¹ Instituto de Investigaciones en Informática y Sistemas de Información,
Facultad de Ciencias Exáctas y Tecnologías,
Universidad Nacional de Santiago del Estero, Argentina
{alvarez, sunzaga, eduran}@unse.edu.ar

Resumen. La formación de grupos de aprendizaje eficaces representa uno de los principales factores de éxito en el Aprendizaje Colaborativo. Sin embargo, las características distintivas de los alumnos hacen que la operación de formar grupos de aprendizaje adecuados sea una tarea difícil. El advenimiento del m-learning y la evolución de las comunicaciones inalámbricas, ha llevado a buscar soluciones que, además, contemplen la movilidad para el agrupamiento de los estudiantes. Con el surgimiento del aprendizaje ubicuo, se suma la necesidad de considerar las condiciones contextuales que pueden influir en el éxito de un aprendizaje colaborativo y situado. En este trabajo proponemos un método para generar recomendaciones de compañeros con los que un estudiante puede conformar un grupo de trabajo para desarrollar una tarea colaborativamente en un ambiente de aprendizaje ubicuo. El método considera condiciones temporales, académicas, disponibilidad horaria, y cercanía física. Se presenta, además, un ejemplo demostrativo de la aplicación del método.

Palabras Clave: Aprendizaje colaborativo, Aprendizaje Ubicuo, Personalización, Integración de grupos.

1 Introducción

El Aprendizaje Colaborativo (AC) representa un enfoque educativo que Dillenbourg [1] define como “una situación, en la cual dos o más personas aprenden o buscan aprender algo juntas”. Con el surgimiento del e-learning, se ha comenzado a buscar cómo mejorar las pedagogías del AC. Como resultado, desde fines de los noventa, se ha creado una nueva rama del AC llamada Aprendizaje Colaborativo Soportado por Computadora (ACSC). Además, el rápido desarrollo de la comunicación inalámbrica y las tecnologías móviles contribuyó con la aparición de una nueva forma de aprendizaje denominada aprendizaje móvil (m-learning). Este permite a los estudiantes a través del uso de dispositivos móviles aprender en cualquier momento y en cualquier lugar, de manera formal o informal. Como resultado, AC se ha hecho posible en situaciones móviles y entornos del mundo real, dando lugar al Aprendizaje Colaborativo Soportado por Dispositivos Móviles (M-ACSC) [2].

El surgimiento de la computación ubicua y su influencia en el aprendizaje ha permitido un cambio desde el m-learning (basado en la comunicación por redes inalámbricas con dispositivos móviles) al u-learning (basado en la comunicación por redes inalámbricas con dispositivos móviles y tecnología de sensores) [3].

Por otro lado, la formación de grupos de aprendizaje eficaces representa uno de los factores importantes que determinan la eficiencia del AC [4]. De acuerdo con [5], los estudios demuestran que se requieren tres condiciones clave para cualquier AC exitoso: las características de la tarea, los medios de comunicación y la composición del grupo. Sin embargo, las diversidades sociales, culturales, psicológicas y cognitivas de los alumnos hacen que la operación de formar grupos de aprendizaje adecuados sea una tarea difícil y que consume mucho tiempo.

Durante este siglo XXI, coincidiendo con el advenimiento del aprendizaje a distancia, muchos investigadores están intentando ofrecer soluciones (modelos matemáticos, algoritmos de agrupamiento de los estudiantes, sistemas inteligentes, etc.), para el M-ACSC. Sin embargo, la mayoría de estas soluciones no contemplan la movilidad de los estudiantes. Sólo algunos trabajos [6, 7, 8] se centran en grupos dinámicos en función de la posición del alumno.

En el caso del aprendizaje ubicuo, a todo esto, se suma la necesidad de considerar las condiciones contextuales que pueden influir en el éxito de un aprendizaje colaborativo y situado. Por ejemplo, condiciones ambientales, horarios, objetos de interés del contexto, lugar del aprendizaje, etc. En consecuencia, en este trabajo proponemos un método para generar recomendaciones personalizadas a un estudiante, sobre compañeros con los que puede desarrollar una tarea colaborativamente en un ambiente de aprendizaje ubicuo.

Este artículo se organiza en cinco partes. Primero se define y caracteriza el aprendizaje ubicuo. En la segunda se muestran antecedentes de trabajos que combinan el aprendizaje colaborativo y el aprendizaje ubicuo. En la tercera parte se describe el método de formación de grupos propuesto. En la cuarta parte se describe un ejemplo de aplicación del método a un caso concreto; y, por último, se presentan conclusiones y trabajos futuros.

2 Aprendizaje Ubicuo

Desde principios de los años 2000, las nuevas tecnologías móviles introdujeron dispositivos con sensores adicionales que proporcionan nuevas direcciones para el aprendizaje asistido por tecnología. Esto ha llevado a una conciencia omnipresente del contexto [9], que permite a los usuarios interactuar y aprender con sensores e identificadores de radiofrecuencia embebidos en los objetos de su entorno. Esta tecnología de computación ubicua ha comenzado a influir en el aprendizaje en diversos campos y disciplinas. Surge así el aprendizaje ubicuo como un paradigma emergente en el campo del aprendizaje soportado por computadora [10], que significa un cambio desde el e-learning convencional (basado tecnológicamente en las redes de computadoras) al m-learning (basado en la comunicación por redes inalámbricas con dispositivos móviles); y otro desde el m-learning al u-learning (basado en la comunicación por redes inalámbricas con dispositivos móviles con tecnología de

sensores) [3]. Burbules [11] destaca que el aprendizaje ubicuo crea un recurso para apoyar el aprendizaje situado y remarca que la colaboración es una ventaja para este tipo de aprendizaje y, de hecho, para muchos propósitos, es indispensable. La potencialidad de un recurso ubicuo a menudo reside en las contribuciones que hacen varias personas.

En [12] identifican cinco características principales del u-learning, a saber: permanencia, accesibilidad, inmediatez, interactividad y conciencia del contexto. Dey y Abowd [13] definen la conciencia del contexto como "la capacidad de un programa o dispositivo para detectar diversos estados de su entorno o de sí mismo". Según ellos, ubicación, identidad, tiempo y ambiente son los principales tipos de contexto para caracterizar la situación de una entidad en particular.

En [3] se definen parámetros de situación que se deben considerar para un aprendizaje ubicuo. Estos son: situación personal detectada por el sistema; situación ambiental detectada por el sistema; feedback del sensor del dispositivo móvil; datos personales recuperados de las bases de datos; y datos ambientales recuperados de las bases de datos. A partir de estos parámetros situacionales, se proponen doce modelos, que pueden usarse para orientar actividades de aprendizaje ubicuo. En este trabajo se han considerado los modelos 11 y 12. *Modelo 11*: Recolección cooperativa de datos y *Modelo 12*: Resolución cooperativa de problemas.

3 Antecedentes

En esta sección se presentan algunos trabajos que combinan el aprendizaje colaborativo con el aprendizaje ubicuo.

En [14] se ofrece un conjunto de servicios basados en agentes, que permiten la construcción de grupos de colaboración mediante la captura de datos, como el gesto de un alumno y la voz, para ser usados en el proceso de razonamiento del agente. Además, se brinda al estudiante conciencia de aquellos miembros con los cuales congenia, según el perfil del alumno, sus intereses e información histórica. En base a esta información contextual, se construyen dinámicamente varios grupos.

En [9] se propone un entorno de aprendizaje omnipresente y contextual que consta de tres sistemas: el sistema de acceso y adaptación de contenido peer-to-peer, el sistema de gestión de anotaciones personalizado y el sistema de discusión en grupo en tiempo real multimedia. Para este último, se ha diseñado un servicio de mensajes, que permite la discusión; mensajería instantánea; el intercambio, filtrado, extracción y la sincronización de mensajes dentro de un grupo. En el diseño de grupo colaborativo que se propone, cada par es libre de iniciar un grupo de interés especial (SIG) y libre de unirse a cualquier SIG iniciado por otros pares en la red peer-to-peer.

Arroyo y otros [15] presentan un sistema personalizable para desarrollar un juego colaborativo de resolución de problemas multiusuario. Aborda la creciente demanda de apelar experiencias de aprendizaje informal en lugares parecidos a museos. El sistema facilita la colaboración remota al permitir que los grupos de estudiantes se comuniquen a través de un sistema de videoconferencia e interactúen simultáneamente a través de una superficie interactiva multitáctil compartida.

Bravo-Torres y otros [16] presentan la plataforma OPPIA (OPPortunistic Intelligent Ambient learning"), un entorno de aprendizaje omnipresente e inteligente,

que despliega redes de aprendizaje esporádicas (SLNs) entre personas (estudiantes, profesores y expertos) que se encuentran en un lugar común o en un lugar remoto pero conectadas a la plataforma a través de Internet. La idea es establecer redes de aprendizaje dinámicas que animen a sus miembros a trabajar colaborativamente y crear un ambiente de aprendizaje con recursos y actividades adecuadas para satisfacer sus necesidades de aprendizaje. Sin embargo, en el artículo no se describe cómo se construyen o establecen tales redes de aprendizaje.

Por último, en [17] se presenta un modelo de enseñanza-aprendizaje colaborativo y ubicuo, que permite no sólo aprender sino adquirir habilidades comunicativas, de cooperación y trabajo grupal. Además, se presenta el diseño de una plataforma que soporta el modelo y facilita a los docentes la creación de cursos de forma colaborativa y ubicua, y a los estudiantes les permite participar en actividades colaborativas. El modelo propuesto no contempla la conformación de los grupos colaborativos.

4 Descripción del Método

Para el diseño del método que permite generar recomendaciones de pares para realizar una actividad colaborativa y ubicua se ha tomado como referencia la estrategia de personalización para aplicaciones de aprendizaje ubicuo presentada en [18]. En la misma se consideran los siguientes tipos de adaptaciones: *Didáctica, de Interfaz y Colaborativa*. Esta última consiste en adaptar las recomendaciones para las actividades colaborativas del estudiante acorde a las características personales y contextuales propias de los mismos y de sus posibles colaboradores (*pares* o *expertos*). La estrategia propone un enfoque de personalización híbrido que combina la *personalización basada en perfiles de usuario* con la *personalización sensible al contexto*. Mediante la personalización basada en perfiles de usuario se ofrece el material educativo de acuerdo con el estilo de aprendizaje y considerando los conocimientos previos del estudiante. Empleando una personalización sensible al contexto se muestra el material educativo de acuerdo con las características del dispositivo que usa el estudiante en el momento del aprendizaje, de la red de comunicación empleada, su ubicación con respecto al punto de interés (PI), su ubicación con respecto a los expertos o compañeros con los que trabaja, su tiempo disponible y el de sus compañeros. Se utiliza además un enfoque semántico mediante el uso de ontologías para el modelado tanto de los datos del perfil de usuario como del contexto. En [19] se presentó el diseño de las ontologías. A continuación, se describen únicamente las que son necesarias para recomendar grupos de estudiantes para realizar actividades ubicuas y colaborativas:

- *Ontología del Modelo del Estudiante*: describe los datos personales, conocimientos previos, estilo de aprendizaje y nivel de interactividad. Este último dato es fundamental para conformar los grupos colaborativos. Para determinar el nivel de interactividad se consideró el trabajo de Hsieh et. al [8] donde se propone analizar los datos de las intervenciones de los estudiantes y se clasifica a cada alumno con nivel “Alto”: son aquellos estudiantes que trabajan fácilmente en equipo y nivel “Bajo”: son los estudiantes que están aislados y no les gusta construir conexión con otros.

- *Ontología del Modelo del Dominio*: representa los temas, tareas y las clases que permiten la organización y representación de los OA sobre temas del dominio de la aplicación.
- *Ontología de PI*: describe los objetos o lugares sobre los cuales el estudiante puede realizar un aprendizaje.

El método que aquí se presenta para generar recomendaciones de pares que puedan realizar una actividad colaborativa y ubicua, ha sido diseñado en base a la estrategia de personalización descripta y a los parámetros situacionales definidos en [3] y presentados en la sección 2. Los mismos se particularizan para este método de la siguiente manera: a) *Situación personal detectada por el sistema*: ubicación del estudiante, b) *Situación ambiental detectada por el sistema*: ubicación de los compañeros, c) *Feedback del sensor del dispositivo móvil*: Código QR del PI sensado. d) *Datos personales recuperados de las Ontologías*: Nivel de conocimiento de los estudiantes, OA de cada PI, datos del curso, metadatos de los OA (pre-requisitos, tiempo mínimo y máximo para realizar la actividad), e) *Datos ambientales recuperados de las Ontologías*: restricciones horarias de los PI. A estos parámetros situacionales se agregan los *Datos ingresados por el estudiante* a través de su dispositivo móvil, tales como: su identificación, curso y actividad seleccionada, y su tiempo disponible para realizar la tarea.

Asumiendo que el estudiante ha seleccionado una actividad colaborativa para realizar en un determinado entorno, el método personalizado para la recomendación de grupo de estudiantes comprende las siguientes etapas (figura 1):

- Verificación de condiciones temporales*: a partir de la actividad seleccionada, se filtran los PI involucrados en la tarea. Por cada PI se debe verificar si el horario actual se encuentra en el rango de las restricciones horarias del PI. Esta información se infiere de las ontologías del Modelo de Dominio y de PI.
- Selección de pares basada en el perfil de estudiante*: se obtiene una lista de compañeros de grupo candidatos que deben cumplir con las condiciones: tienen que realizar la misma actividad y deben haber concluido con los temas y/o actividades anteriores que son pre-requisitos para realizar la actividad ubicua y colaborativa actual. Para realizar esta etapa, se requiere determinar el tema de la actividad y los pre-requisitos necesarios para realizarla, infiriendo este conocimiento de la ontología del Modelo de Dominio. Luego se infiere desde la ontología del Modelo del Estudiante, aquellos alumnos que cumplen con las condiciones académicas descriptas.
- Selección de pares basada en tiempo disponible*: a partir de la lista obtenida anteriormente, se infieren los compañeros cuyos tiempos disponibles (dato ingresado por los estudiantes a través de sus dispositivos móviles) sea mayor o igual a lo consignado como tiempo necesario para realizar la actividad (obtenido de la Ontología del Modelo de Dominio).
- Selección de pares basada en el perfil de interacción de los estudiantes*: En este método se propone que a partir de la lista de compañeros candidatos anterior se infiera a los estudiantes que tiene un alto nivel de interactividad y a los que poseen un bajo nivel de interactividad de manera tal de obtener grupos heterogéneos. Los datos del nivel de interactividad se obtienen de la Ontología del Modelo del Estudiante. Como salida de esta etapa se obtienen dos listas de estudiantes, una con los de alto nivel de interactividad y la otra con los de bajo nivel.

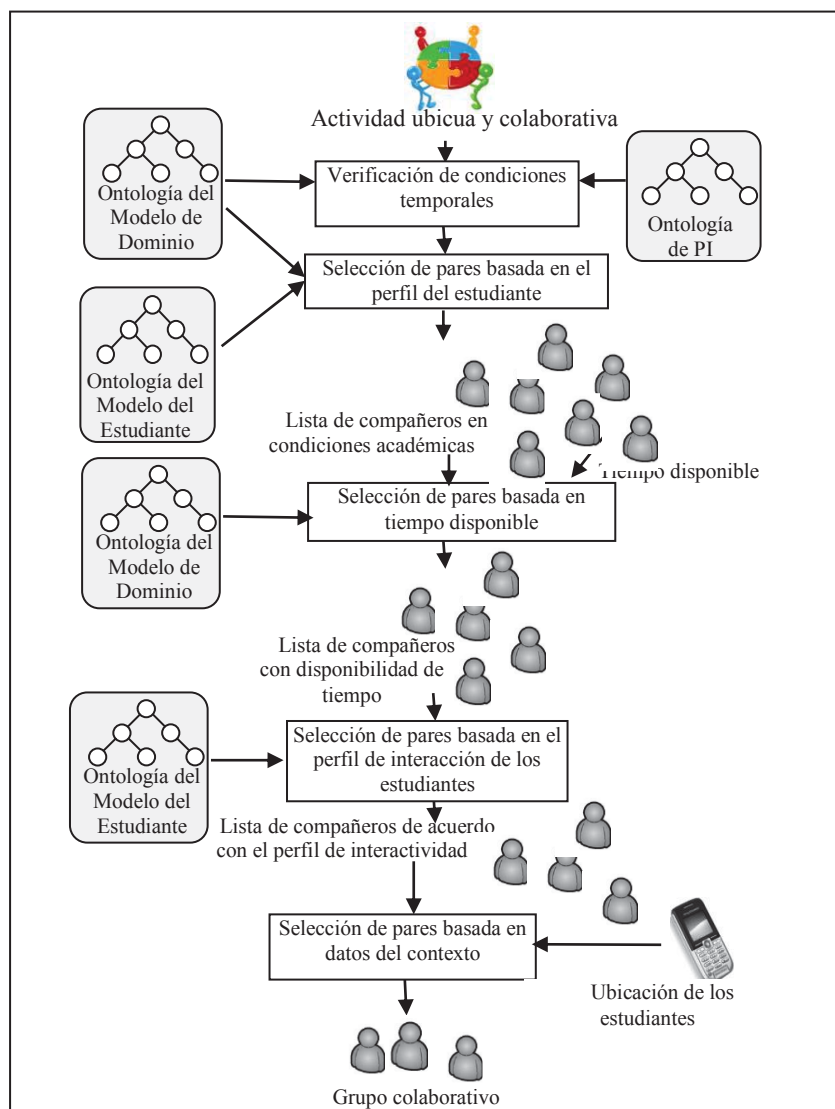


Figura 1: Método de Personalización para aplicaciones de aprendizaje ubicuo y colaborativo

- e) *Selección de pares basada en datos del contexto:* a partir de las listas generadas en la etapa anterior, se infieren aquellos compañeros que se encuentren físicamente en una ubicación cercana. En este caso para la formación del grupo se considera la propuesta de [7] que utiliza el algoritmo de Dijkstra. El mismo calcula el camino más corto desde un nodo origen hacia los demás nodos del grafo. En este método, el nodo origen es la ubicación geográfica del alumno que requiere realizar la actividad ubicua y colaborativa y los restantes nodos del grafo son la posición (latitud y longitud) de los otros estudiantes que integran las listas obtenidas en el

paso anterior. La información de las coordenadas geográficas de los estudiantes se obtiene del GPS de sus dispositivos. El grafo tendrá dos tipos de nodos que diferencian a los estudiantes con alto y bajo nivel de interactividad. Previo a la aplicación del algoritmo se calculan las distancias entre los nodos, valor que se asigna a las aristas del grafo. Bajo la premisa de conformar grupos pequeños (4 integrantes) y heterogéneos (con al menos uno de cada nivel de interactividad) fue necesario modificar el algoritmo de Dijkstra. El algoritmo deberá encontrar el camino más corto desde el nodo origen hasta completar un recorrido de 4 nodos con estudiantes heterogéneos. Ya que el algoritmo presenta un orden de complejidad de $O(|V|^2)$, sin utilizar cola de prioridad, con $|V|$ = cantidad de vértices, se considera adecuado realizar los filtrados presentados en los ítems para reducir los posibles candidatos, y por lo tanto disminuir la cantidad de vértices.

En resumen, los criterios utilizados en el método se muestran en la tabla 1.

Tabla 1: Criterios de agrupamiento empleados en el método

CRITERIOS DE AGRUPAMIENTO	OBJETIVO	DATOS NECESARIOS	FUENTE
Condiciones temporales	Verificar si la actividad se puede realizar en ese momento, acurdo a la disponibilidad de los PI	Hora actual	Reloj
		Restricción horaria de PI	Ontología de PI.
Condiciones académicas	Lograr homogeneidad en los conocimientos de los integrantes del grupo, reuniendo los que deban realizar la misma actividad y tengan los conocimientos previos.	Conocimiento previo del estudiante.	Ontología del Modelo del Estudiante.
		Tema de la actividad Pre-requisitos	Ontología del Modelo de Dominio
Tiempo disponible	Lograr que los estudiantes que integren el grupo tengan tiempo disponible (en ese momento) para realizar la actividad.	Tiempo disponible de los estudiantes.	Ingresado por los estudiantes.
		Tiempo necesario para realizar la actividad.	Ontología del Modelo de Dominio.
Interacción	Formar grupos de estudiantes heterogéneos que mejoren la interacción real (con alto y bajo nivel de interactividad).	Nivel de interactividad.	Ontología del Modelo del Estudiante.
Ubicación	Lograr grupos de aprendizaje con estudiantes que se encuentran cerca.	Ubicación de cada estudiante.	GPS de los dispositivos.

5 Ejemplo de aplicación del método

Teniendo en cuenta el método descrito se propone su aplicación para recomendar a los estudiantes del curso de ingreso a la universidad, integrantes de grupos con quienes realizar una actividad colaborativa y ubicua. Esta actividad permitirá aprender sobre los aspectos académicos y administrativos de la vida universitaria: inscripciones, becas, horarios de clases, etc. Para el diseño de la actividad se tienen en cuenta uno de los modelos de aprendizaje ubicuo propuestos por [3] y presentados en la sección 2: *Modelo 11: Recolección cooperativa de datos*. Se le pide a un grupo de estudiantes recolectar colectivamente los datos del mundo real y discutir sus hallazgos con otros a través de dispositivos móviles. Para la actividad propuesta se considera además un aprendizaje colaborativo (construcción conjunta). En este marco, se

plantea la siguiente actividad identificada como ACT23, que tiene como pre-requisitos el haber completado las actividades ACT10 y ACT8 y requiere 2 hs para su desarrollo.

Objetivo de la actividad: construir colaborativamente un mapa o esquema ilustrado que identificar y visualizar los lugares de la universidad o PI a los que debe dirigirse un estudiante para realizar diferentes trámites académicos y administrativos.

Descripción de la actividad: Dado los siguientes trámites: inscripción por materias correspondientes al primer año, gestionar becas de estudio y solicitar un libro en la biblioteca. Se pide: 1) Identificar los PI que se encuentren involucradas en los mismos. 2) Obtener imágenes e información de cada PI. 3) Realizar el mapa incluyendo las imágenes y los instructivos de cada trámite.

Situación: El estudiante A1 quiere hacer la tarea descripta, se asume que son las 11 hs y se encuentra en la Universidad. En la tabla 2 se muestran los datos de los otros estudiantes.

Tabla 2: Datos de los alumnos

Alumnos	Actividad Realizada	Tiempo Disponible	Nivel de Interactividad
A1	ACT10, ACT8	2 hora	Alto
A2	ACT8, ACT10	1 hora	Bajo
A3	ACT8, ACT10	3 horas	Alto
A4	ACT2, ACT10, ACT8	2 horas	Bajo
A5	ACT6, ACT23	3 horas	Bajo
A6	ACT23, ACT18	1 hora	Alto
A7	ACT8, ACT7, ACT10	2 horas	Alto
A8	ACT3, ACT9	3 horas	Bajo
A9	ACT8, ACT10	3 horas	Bajo
A10	ACT23	2 horas	Bajo
A11	ACT2, ACT8, ACT10	2 horas	Bajo
A12	ACT8, ACT23	4 horas	Bajo
A13	ACT3, ACT8, ACT10	2 horas	Alto

Aplicación del método: de la aplicación del paso a) *Verificación de condiciones temporales* surge que los PI involucrados en esta actividad son: PI1 Dpto. Alumnos, PI2 Dpto. Becas y PI3 Biblioteca. El horario en que se encuentran disponibles es de 8 a 20 hs. Como son las 11 hs, es posible acceder a los tres PI. Aplicando el paso b) se descarta a A5, A6, A10 y A12 porque ya han realizado la ACT23, y a A8 porque aún no ha realizado los requisitos previos ACT8 y ACT10. Aplicando el paso c) se descarta a A2 porque su tiempo disponible es inferior a las 2 hs necesarias para hacer la tarea. Aplicando el paso d) se obtienen dos listas de alumnos: lista 1 con nivel de interactividad alta formada por A1, A3, A7, A13 y la lista 2 con nivel de interactividad baja formada por A4, A9, A11. A partir de estas lista se aplica el paso e) y se construye el grafo, donde los vértices identifican a los alumnos, los que están sombreados son alumnos con nivel de interactividad "Alto" y los sin sombreados "Bajo" y en las aristas se informan las distancias entre ellos. Se aplica el algoritmo de Dijkstra y se obtiene el grupo final de estudiantes más cercanos a recomendar para realiza la tarea colaborativa y ubicada (figura 2), cumpliendo con las restricciones de armar un grupo pequeño y con al menos un integrante con nivel de interactividad distinto al del nodo origen. En consecuencia la lista a recomendar es: A1, A4, A11, A13.

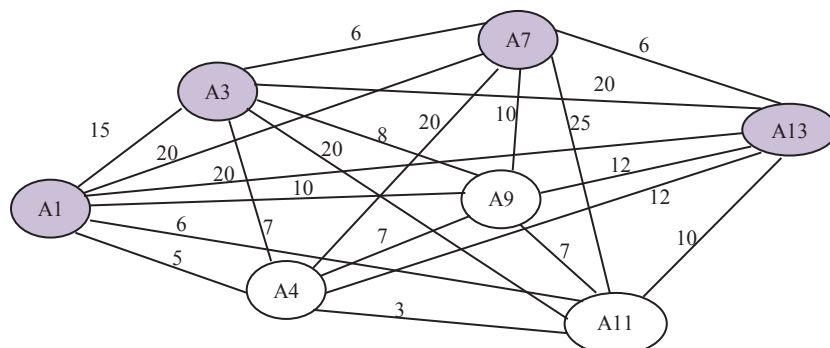


Figura 2: Grafo con los alumnos candidatos a formar el grupo colaborativo

6 Conclusiones y Trabajos Futuros

El objetivo de este trabajo fue proponer una solución al agrupamiento de estudiantes cuando tienen que realizar un aprendizaje situado y colaborativo apoyado por tecnología de computación ubicua. El método propuesto tiene como fortaleza que contempla en el agrupamiento los parámetros situacionales que son relevantes en un aprendizaje ubicuo: la ubicación del estudiante y de sus compañeros, nivel de conocimiento de los estudiantes, tiempo disponible para realizar la tarea y nivel de interactividad de cada estudiante. Esto permite obtener grupos homogéneos en cuanto al nivel de conocimiento de sus integrantes y heterogéneos en cuanto al nivel de interactividad. Otra fortaleza del método es considerar la cercanía de los integrantes para el agrupamiento. Esto se logra aplicando el Algoritmo de Dijkstra. Pero como este algoritmo presenta un orden de complejidad creciente según el número de vértices del grafo, se consideró adecuado realizar filtrados de candidatos por otros criterios para reducir la cantidad de vértices. Además, se impone la restricción de recomendar un grupo que cumpla con la condición de ser pequeño (no más de cuatro integrantes) para eficientizar la tarea, y heterogéneo en cuanto al nivel de interactividad. Para ello, se propuso la modificación del Algoritmo de Dijkstra, considerando en el grafo dos tipos de nodos y armando el camino de manera de lograr como objetivo el camino más corto constituido por 4 nodos, donde al menos uno debe ser de un tipo diferente al nodo origen. Como trabajo futuro planeamos implementar el método y evaluar su impacto en aplicaciones de u-learning concretas.

Referencias

1. Dillenbourg, P.: What do you mean by collaborative learning? In: P. Dillenbourg (Ed.). Collaborative-learning: Cognitive and Computational Approaches, 1-19. NY:Elsevier (1999)
2. Zurita, G. and Nussbaum M.: Computer supported collaborative learning using wirelessly interconnected handheld computers. *Computer & Education*. 42(3), 289-314 (2004)

3. Hwang, G. J., Tsai, C. C. and Yang, S. J. H.: Criteria, strategies and research issues of context-aware ubiquitous learning. *Educational Technology & Society*, 11, 2, 81–91 (2008)
4. Amara S., Macedo J., Bendella F. and Santos A.: Group Formation in Mobile Computer Supported Collaborative Learning Contexts: A systematic literature review, *Journal of Educational Technology & Society*, vol. 19, pp. 258-273 (2016)
5. Dillenbourg, P.: Over-scripting CSCL: The risks of blending collaborative learning with instructional design. In P. A. Kirschner (Ed.), *Three worlds of CSCL. Can we support CSCL?* (pp. 61–91). Heerlen: Open Universiteit Nederland (2002)
6. Wassa Mballo M. H.: *Forming groups of mobile learners that promote collaborative learning supported by mobile devices*. Springer (2017)
7. Messeguer, R., Medina, E., Royo, D., Navarro, L., and Juárez, J. P.: Group prediction in collaborative learning. In Callaghan, V., & Kameas, A. (Eds.), *The 6th IEEE International Conference on Intelligent Environments* (pp.350-355). IEEE Computer Society (2010)
8. Hsieh, J. C., Chen, C. M. and Lin, H. F.: Social interaction mining based on wireless sensor networks for promoting cooperative learning performance in classroom learningenvironment. In: U. Hoppe et al. (Eds.), *The 6th IEEE international conference on wireless, mobile, and ubiquitous technologies in education* (pp.219-221). Los Alamitos, CA: IEEE Computer Society (2010)
9. Yang, S. J. H.: Context Aware Ubiquitous Learning Environments for Peer-to-Peer Collaborative Learning. *Educational Technology & Society*, 9 (1), 188-201 (2006)
10. Okamoto, T. and Tseng, S. S.: Context-aware and ubiquitous learning (Guest editorial). *Educational Technology & Society*, 11, 2, 1–2 (2008)
11. Burbules N.: L'apprentissage omniprésent et l'avenir de l'enseignement. *Rencontres on Education*, Vol. 13, 3 – 14 (2012)
12. Yahya S., Ahmad E. A. and Jalil K. A.: The definition and characteristics of ubiquitous learning: A discussion. *International Journal of Education and Development using Information and Communication Technology*. Vol. 6, Issue 1, pp. 117-127 (2010)
13. Dey, A.K. and Abowd, G.D.: *Towards a Better Understanding of Context and Context-Awareness*. CHI 2000 Workshop on the What, Who, Where, When, and How of Context-Awareness (2000)
14. Zhang, G., Jin, Q., and Lin, M: A Framework of Social Interaction Support for Ubiquitous Learning. Paper presented at the 19th International Conference of Advanced Information Networking and Applications, March 28-30, 2005, Tamkang University, Taiwan (2005).
15. Arroyo E., Righil V., Tarrago R., Santos P., Hernandez-Leo D. and Blat J.: Remote Collaborative Multi-User Informal Learning Experiences: Design and Evaluation. *Towards Ubiquitous Learning*, 43-56 (2011).
16. Bravo Torres, J. F., Ordoñez Morales, E. F., López Nores, M. and Blanco Fernández, Y.: OPPIA: A context-aware ubiquitous learning platform to exploit short-lived student networks for collaborative learning. In 8th International Conference on Computer Supported Education (CSEDU). Rome, Italy: INSTICC. (2016)
17. Coto M., Collazos C.A. and Rivera S. M.: Collaborative and Ubiquitous Model to Support Teaching and Learning Processes in Iberoamerica. *Revista de Educación a Distancia*. Núm. 48. Art. 10 (2016)
18. Alvarez M. M.; Duran E. B. y Unzaga, S. I.: *Recomendaciones Personalizadas para Aplicaciones de Soporte al Aprendizaje Ubicuo*. XXI Congreso Argentino de Ciencias de la Computación (CACIC 2015), Buenos Aires. Argentina. (2015).
19. Durán E. B., Álvarez, M. M. and Unzaga S. I.: "Ontological model for the personalization of u-learning applications". EATIS 2016. Cartagena de Indias. Colombia. Publicado en IEEE Xplore. Disponible en: <http://ieeexplore.ieee.org/document/7520154/>. (2016)

Nuevas Tecnologías Aplicadas a la Enseñanza de la Ingeniería. Análisis Comparativo de Ambientes Virtuales 3D

Francisco Lizarralde, Constanza Huapaya

Grupo de Investigación en Inteligencia Artificial Aplicada a Ingeniería,
Departamento de Matemática, Facultad de Ingeniería, U.N.M.d.P.,
Av. Juan B. Justo 4302, Mar del Plata, Argentina
{flizarra, huapaya}@fi.mdp.edu.ar

Resumen El objetivo del presente trabajo es presentar un análisis comparativo de las capacidades que ofrecen los entornos virtuales tridimensionales para su utilización dentro del ámbito educativo, como una herramienta que permita reducir la brecha cognitiva presente entre conceptos teóricos y su aplicación práctica, en particular, en la enseñanza de ciertos contenidos que conforman la currícula básica de diversas especialidades de Ingeniería. Para ello se estableció un marco de evaluación basado en ciertos aspectos considerados relevantes para la elaboración de contenidos y para su utilización como recurso didáctico, en especial, en lo que se refiere a la implementación de técnicas de simulación numérica. Posteriormente se aplicó este marco de evaluación a un conjunto de ambientes virtuales 3D, elegidos de acuerdo a cierta cualidad representativa, con la finalidad de evaluar comparativamente su desempeño.

Keywords: Ambientes Virtuales 3D, Enseñanza Ingeniería, Simulaciones Educativas

1. Introducción

En la enseñanza de la Ingeniería se suele analizar el comportamiento dinámico de diversos sistemas, así como también, se pone énfasis en el estudio de los modelos matemáticos que los representan. Sin embargo, a menudo los estudiantes suelen “perdersse” en tales representaciones, lo que dificulta el aprendizaje de los conceptos fundamentales. En este sentido, los medios tradicionales como el pizarrón o los libros de texto, resultan insuficientes para expresar la naturaleza dinámica de estos sistemas.

En este contexto, los entornos virtuales 3D emergen como un medio adecuado para la experimentación y la representación, tanto dinámica como espacial, de diversos fenómenos físicos, pues brindan la posibilidad de simular, experimentar y analizar el comportamiento de dichos sistemas dinámicos, bajo diferentes condiciones.

Además, existen fenómenos que por su escala, ya sea microscópica o astronómica, resultan imposibles de reproducir en el ámbito de un laboratorio. Sin

embargo, es posible estudiar el comportamiento de tales fenómenos, simulando numéricamente las leyes naturales que los gobiernan, al dotar de propiedades físicas a los objetos del ambiente virtual 3D y estudiar su comportamiento bajo la influencia de campos gravitatorios, eléctricos, magnéticos, etc.

De esta forma, las simulaciones actúan como mediadores entre el modelo matemático (*conocimiento abstracto*) y la forma en que el estudiante percibe su funcionamiento (*conocimiento concreto*). Lo cual permite abordar problemas tan complejos como el estudio de la mecánica de fluidos o las trayectorias de los cuerpos celestes y al mismo tiempo aumentar el interés y la motivación en el estudiante, mejorando claramente su aprendizaje.

En este sentido, los ambientes virtuales 3D operan como una extraordinaria herramienta para la adquisición de competencias y habilidades en un contexto simulado, permitiendo a los estudiantes aplicar los conocimientos teóricos adquiridos en el aula, en ambientes “*cuasi-reales*” representados en la computadora. Esto permite la generación de un modelo informático al que es posible someter a pruebas exhaustivas con la finalidad de comprobar su exactitud, sin el costo o el riesgo que implicaría realizar en el mundo real aquello que se afirma desde el punto de vista teórico. [1].

2. Metodología

Si bien los términos *Mundo Virtual 3D* y *Ambiente Virtual 3D* se suelen utilizar indistintamente, es conveniente aclarar, que dichos términos no son sinónimos. En este sentido, un *Mundo Virtual 3D* puede ofrecer facilidades de navegación, comunicación y juegos, pero si carece de facilidades para la creación de contenidos, personalización de espacios y programación del comportamiento de su entorno, no puede ser considerado como un *Ambiente Virtual 3D*.

Teniendo en cuenta estos aspectos y en concordancia con el alcance y los objetivos propuestos en el presente trabajo, se optó por tomar como objeto de estudio a los Ambientes Virtuales 3D, *Second Life*, *OpenSimulator* y *OpenCobalt*, debido a que cada uno presenta alguna característica relevante que lo distingue de los demás.

En este sentido, *Second Life* fue elegido por ser uno de los ambientes virtuales 3D más reconocidos, de mayor vigencia y cantidad de usuarios, contando en la actualidad con más de un millón de usuarios regulares. Este entorno posee, sin lugar a dudas, una clara finalidad comercial y de entretenimiento, aunque también es utilizado por diversas instituciones académicas para la creación de contenidos educativos, y principalmente para tener una presencia institucional en el mundo virtual.

El segundo ambiente virtual 3D considerado fue *OpenSimulator*, un proyecto surgido a partir de la liberación del código fuente del programa cliente de *Second Life*, a comienzos de 2007, lo que lo convirtió en una especie de alternativa de código abierto de *Second Life*. Esto representa un aspecto clave, ya que posibilita la instalación de servidores (*Grids*), totalmente independientes de la empresa

Linden Labs, e incluso abre la posibilidad de vincularlo con otros proyectos educativos de Software Libre, como por ejemplo *Moodle*.

Y por último, se seleccionó a *Open Cobalt*, un entorno muy poco conocido, íntegramente desarrollado en una versión moderna de *Smalltalk* llamada *Squeak* [2]. Este ambiente difiere totalmente del modelo tradicional que utilizan la mayoría de los ambientes virtuales 3D, los cuales administran todas las comunicaciones e interacciones a través de un servidor central. En este sentido, *Open Cobalt* utiliza un esquema descentralizado de comunicaciones del tipo P2P (*Peer to Peer*).

3. Marco de Análisis Comparativo

Establecer un marco para realizar un análisis comparativo entre diferentes ambientes virtuales 3D requiere de la consideración de un conjunto de características comunes y representativas, para su evaluación.

La FAS (*Federation of American Scientist*) ha establecido un conjunto de características para evaluar, desde un punto de vista general, a los entornos virtuales 3D [3] y según se desprende del análisis de diversos trabajos de investigación sobre esta temática [4], [5], estas propuestas varían según el criterio establecido por cada autor.

Sin embargo, la elección de este conjunto de características está fuertemente relacionada con el objetivo que se desea alcanzar a través de la utilización de dichos entornos. En este sentido, el ámbito de utilización propuesto es un ambiente universitario, de carreras de Ingeniería y sus destinatarios son estudiantes que se encuentran promediando su ciclo de estudios. Y desde el punto de vista curricular, el objetivo es simular sistemas dinámicos, con la intención que los estudiantes adquieran un conocimiento más profundo tanto del fenómeno en estudio como de los métodos numéricos involucrados en su resolución.

En este contexto se ha establecido un marco comparativo, a partir de la evaluación del siguiente conjunto de características:

- **Fidelidad de la Representación**

La fidelidad de la representación es indispensable para lograr una experiencia “*espacial*” creíble, es decir, para que el usuario tenga la sensación de estar dentro de ese espacio, interactuando con todos los elementos que allí se encuentran. Las tarjetas de video actuales poseen una gran potencia de procesamiento, lo que les permite manejar representaciones gráficas de gran calidad, texturas y efectos de iluminación con gran velocidad y en forma prácticamente autónoma, lo cual resulta imprescindible para lograr una representación dinámica convincente.

A su vez, los objetos del entorno, así como las representaciones de los usuarios (avatars) necesitan comportarse consistentemente. Es decir, el usuario espera que dentro del entorno se cumplan ciertas “*leyes físicas*”, como por ejemplo, que la intensidad de las fuentes sonoras presentes, varíe de acuerdo a la posición y la distancia que lo separa de las mismas (*sonido espacial*).

- **Navegabilidad e Interacción con el Entorno**

Por navegabilidad e interacción con el entorno, nos referimos a los mecanismos utilizados para recorrer los espacios y para modificar ciertos atributos, que afectan tanto al aspecto visual como a su comportamiento, así como también, la creación y manipulación de los objetos contenidos en el mismo.

Además, la sensación de presencia en primera persona, exclusiva de los ambientes virtuales permite generar nuevos aprendizajes, pues amplía las formas de adquirir conocimiento, en la interacción y transformación de ideas abstractas en representaciones perceptibles a través de objetos virtuales [6].

- **Comunicación entre Usuarios**

La comunicación oral puede lograrse por medio de sistemas de voz a través de Internet, también conocidos como VoIP (*Voice over Internet Protocol*); y la comunicación escrita por medio de herramientas de mensajería instantánea o IRC (*Internet Relay Chat*) integradas en la misma plataforma. Sin embargo, una característica propia de los mundos virtuales es la incorporación de la comunicación gestual, la cual se logra dotando a las representaciones visuales de los usuarios (*avatars*) de un cierto conjunto básico de movimientos o gestos.

- **Modelos 3D y contenidos Multimedia**

Los modelos 3D y los contenidos multimedia son recursos que producen una singular sinergia cuando se incorporan a los mundos virtuales. En el primer caso nos referimos a figuras, vehículos, edificios, etc.; y en el segundo caso, a elementos multimediales tales como fotografías, archivos de audio y video. En todos los casos para poder integrarlos como recursos pedagógicos, se requiere que el entorno tenga la capacidad de cargarlos y reproducirlos adecuadamente.

- **Integración de Aplicaciones 2D**

Para que un entorno virtual 3D no sea una aplicación aislada es conveniente que permita integrar aplicaciones externas dentro del mismo o al menos contar con la posibilidad de ejecutarlas aunque sea de forma remota. Esto permitiría a los estudiantes la utilización de navegadores de Internet, planillas de cálculo, editores de texto y otras aplicaciones 2D, sin necesidad de salir del “*espacio virtual*”.

- **Entorno de Programación**

Los procesos de aprendizaje y experimentación, como las simulaciones, son más efectivos y motivadores si los mismos pueden ser creados colaborativamente y compartidos con otros estudiantes. Para lograr este cometido, es necesario contar con un lenguaje dinámico y un entorno de programación integrado que permita realizar modificaciones en forma interactiva.

4. Análisis de las características

Para disminuir la granularidad del análisis se han considerado tres rasgos distintivos para cada una de las características expuestas en la sección anterior. Las figuras 1 y 2 representan un ejemplo. En cuanto a la evaluación, se ha establecido un rango de valores entre 0 y 5, lo que permite una clara representación de los atributos correspondientes a cada ambiente virtual 3D.

4.1. Fidelidad de la Representación

La Fidelidad de la Representación involucra aspectos tan diversos como la calidad de la representación gráfica de los elementos que componen el ambiente, así como también del comportamiento de los objetos móviles generado por el motor de física. Por último un aspecto no menos importante, son las herramientas que permiten editar y personalizar a la representación del usuario, es decir, de su Avatar.

En *Second Life* los simuladores son los encargados de ejecutar el motor de física (*Physics Engine*), detectar colisiones, y mantener la información de estado de todos los objetos de la región con la finalidad de enviar dicha información al visualizador, quien se encarga de mostrar la escena desde el punto de vista del usuario.

Las simulaciones físicas de los objetos móviles son manejadas a través de la interfaz de programación de aplicaciones (API) del motor de física Havok (*Havok Physics Engine*). Estas simulaciones se encuentran acotadas al ámbito de la mecánica clásica [7]. Es decir, el comportamiento de los objetos es descrito principalmente en función de la influencia de las fuerzas externas aplicadas y de la interacción con otros objetos de la región.

En el caso de *OpenSimulator* no se aprecia a primera vista una gran diferencia con el aspecto que presenta *Second Life*, en lo que se refiere al nivel de detalle y estética gráfica en general, así como en lo referente a la personalización de los *Avatars* y las herramientas que permiten construir objetos a partir de figuras primitivas. Sin embargo, la incorporación de un nuevo motor de física como *Bullet™* mejoró notablemente su nivel de realismo.

A diferencia de los ambientes 3D anteriormente analizados, en *OpenCobalt* los “mundos virtuales” no ocupan regiones rectangulares contiguas, sino que cada uno representa un espacio completamente independiente, incluso aquellos que se encuentran dentro de la misma computadora. *OpenCobalt* cuenta con los elementos básicos para crear espacios virtuales completamente diferentes, de forma muy rápida y sencilla, pero es evidente que en este aspecto, no posee el grado de sofisticación de otros ambientes como *OpenSimulator* o *Second Life*, los cuáles exhiben un nivel estético y de detalle claramente superior.

Por otra parte, si bien en el año 2008 se desarrolló un plugin para ODE (*Open Dynamics Engine*) en *Squeak*, con la intención de incorporar un motor de física en *OpenCobalt*, el proyecto fue abandonado, fundamentalmente por la imposibilidad de integrarlo junto con el esquema descentralizado de comunicaciones de *OpenCobalt*.

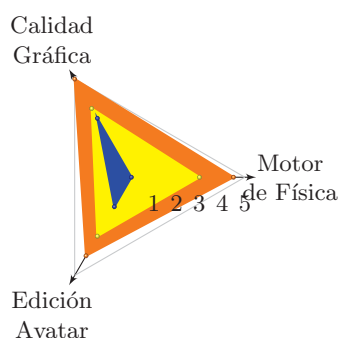


Fig. 1. Fidelidad de la Representación

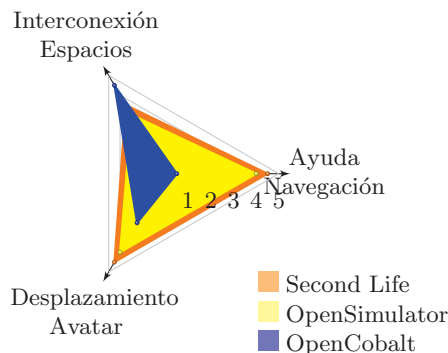


Fig. 2. Interacción con el Entorno

4.2. Navegabilidad e Interacción con el Entorno

En el esquema Cliente-Servidor que posee *Second Life*, el programa Cliente provisto por *Linden Labs*, denominado *SL Viewer*, es el que le permite al usuario conectarse con el Servidor y navegar por las diferentes regiones del entorno. En 2007, *Linden Labs* liberó gran parte del *código fuente* de este programa cliente, lo que dió lugar a que otros desarrolladores crearan sus propias versiones, como *Firestorm*, *Hippo*, *Imprudence*, *Singularity*, etc.

La tarea principal de los visualizadores es realizar las operaciones gráficas (*rendering*) de la escena. Los visualizadores se comunican con el servidor para obtener información sobre la ubicación y velocidad de los objetos que se encuentran en la región, aunque no son capaces de detectar colisiones entre ellos, ya que esta tarea es responsabilidad del motor de física y del simulador.

En *Second Life* el Avatar del usuario puede caminar, correr e incluso volar, aunque una de las formas más elementales de desplazarse por una región, es utilizando las teclas de desplazamiento *arriba*, *abajo*, *izquierda* y *derecha*. Sin embargo, para realizar desplazamientos con cierta precisión, lo usual es pasar al modo *Mouselook*, por medio de la rueda del mouse, en este modo la cámara muestra una vista en primera persona y el *Avatar* se desplazará hacia donde apunte el cursor del mouse.

La navegación dentro de una región de *OpenSimulator* es en apariencia, muy similar a la navegación en *Second Life*. Sin embargo, debido a que *OpenSimulator* puede ejecutarse en diferentes modos: *Standalone*, *Grid* y *Hypergrid*, se presentan algunas diferencias.

Una vez instalado el servidor de *OpenSimulator*, el mismo se ejecutará por defecto en el modo *Standalone*, es decir, ejecutando un mundo virtual completamente aislado. Sin embargo, si se ejecuta en modo *Hypergrid*, se puede vincular esa región con otros espacios virtuales, y una vez establecidos los enlaces, será posible *teleportarse*, simplemente utilizando el hipervínculo establecido, aunque con la ventaja de conservar los contenidos y objetos 3D propios, dentro de la misma computadora.

En el caso de *OpenCobalt* la navegación dentro de un espacio virtual 3D, se realiza en forma similar a los *juegos en primera persona*, en los que el usuario, representado por su *Avatar*, se desplaza utilizando ciertas teclas para avanzar, retroceder o girar o haciendo uso de los *botones de navegación* que se encuentran en la parte inferior derecha de la pantalla.

Por otra parte, como los espacios virtuales en *OpenCobalt* no ocupan regiones contiguas, para pasar de un espacio a otro, es necesario utilizar un *portal*. Esto incorpora un concepto interesante, ya que si bien cada espacio posee un tamaño determinado y por lo tanto es posible recorrer una cierta distancia dentro del mismo, cuando se atraviesa un portal, la distancia entre ambos espacios virtuales es prácticamente nula. En este sentido, un *portal*, no es simplemente una ventana que muestra, en tiempo real, lo que está sucediendo en el otro espacio, sino que al atravesarlo, se ingresa instantáneamente en el otro espacio. Por este motivo, como es posible tener varios portales abiertos simultáneamente, y cualquier usuario puede pasar de un espacio virtual a otro, simplemente atravesando el portal correspondiente; es probable que luego de atravesar varios portales se experimente una cierta sensación de desorientación, ya que en *OpenCobalt* no existen mapas de navegación que le indiquen al usuario donde se encuentra el portal que pueda conducirlo a su espacio de origen.

4.3. Comunicación entre Usuarios

Second Life posee diversos mecanismos de comunicación. La comunicación escrita se efectúa por medio de *Text Chat* e *Instant Messaging*, la oral por *VoIP* y por último la gestual. Debido a que *Second Life* fue concebido fundamentalmente como un lugar de encuentro y participación social, los mecanismos de comunicación juegan un papel preponderante. La comunicación a través de *VoIP* (Voice over Internet Protocol) no sólo permite la comunicación entre usuarios sino que es una herramienta clave para la realización de diversos eventos, desde foros y congresos hasta representaciones teatrales.

Por otra parte, la comunicación gestual es un aspecto muy interesante de los ambientes virtuales, ya que en el mundo real, este tipo de comunicación es tan o más importante que la comunicación oral o escrita. En este sentido *Second Life* cuenta con un menú mediante el cual se puede elegir el gesto más apropiado para dar a entender a los demás usuarios cuáles son nuestras intenciones. El repertorio de gestos es variado, y si se desea, se pueden crear nuevos gestos y agregarlos al menú.

Los mecanismos de comunicación oral y textual de *OpenSimulator* son prácticamente idénticos a los de *Second Life*. Y si bien, los usuarios utilizan principalmente el Chat y los mensajes instantáneos, la comunicación oral, a través de *VoIP*, junto con los gestos personalizables y sonidos asociados, configuran un medio de comunicación adicional, que reafirma la sensación de presencia.

OpenCobalt proporciona dos opciones de comunicación textual. Una local, para establecer comunicaciones dentro del mismo espacio, y otra externa, por medio del protocolo XMPP (*eXtensible Messaging and Presence Protocol*). Por otra parte, para la comunicación oral entre usuarios, utiliza *VoIP* (Voice over

Internet Protocol), aprovechando además, las ventajas del sonido espacial que provee *OpenAL*, una biblioteca de funciones de audio multiplataforma, desarrolladas por *Creative Labs*TM para la ejecución de audio posicional y multicanal en tres dimensiones.

4.4. Modelos 3D y contenidos Multimedia

En *Second Life* es posible generar todo tipo de objetos 3D a partir de figuras geométricas sencillas denominadas *prims*, las cuales pueden ser agrupadas para generar figuras más complejas. Además de los *prims*, existen otras estructuras denominadas *sculpted prims* o *sculpties*, y los modelos 3D de malla (*Mesh*).

Los mecanismos de edición de objetos en *OpenSimulator* a partir de figuras primitivas (*prims*), también son muy similares a los existentes en *Second Life*, ya que esta funcionalidad es aportada mayormente por el visualizador, que es precisamente el elemento común en ambos entornos. En cuanto a la importación de modelos, *OpenSimulator* permite importar modelos de malla (*Mesh*) en formato *.dae* (*Digital Asset Exchange*) y en referencia a otros contenidos multimedia *OpenSimulator* reproduce los formatos de audio o video más comunes como *.wav* y *.mpeg*.

OpenCobalt soporta la importación de modelos 3D en diversos formatos estándar, tales como *.ase*, *.kmz*, *.obj* y *.vrmf*, los cuales pueden crearse previamente con programas externos específicos como *Blender* o *Trimble SketchUp*. También pueden incorporarse modelos completos obtenidos de sitios como *Google 3D Warehouse*, sin embargo, la herramienta *3D EditBox* integrada en *OpenCobalt*, sólo es capaz de trasladar, escalar y rotar los modelos.

4.5. Integración de Aplicaciones 2D

A partir de la nueva versión del programa cliente de *Second Life* (Viewer 2.0), está disponible una nueva característica denominada *Shared Media*, mediante la cual es posible visualizar la información de un navegador web o un video de *Youtube*, simplemente arrastrando y soltando la correspondiente URL (Uniform Resource Locator) sobre la superficie de una figura primitiva (*prim*), accediendo de esta forma a contenidos de streaming de audio y video.

Una forma sencilla de integrar aplicaciones 2D en *OpenSimulator* es por medio de una pequeña aplicación, gratuita y de código abierto, denominada *Gua-camole*, que se instala junto con el servidor VNC (*Virtual Network Computing*). Esta aplicación se comunica con el servidor VNC y realiza el rendering de su salida en formato HTML5, lo que permite que los usuarios utilicen aplicaciones de escritorio dentro del ambiente, a través del visualizador de *OpenSimulator*.

En *OpenCobalt* es posible ejecutar colaborativamente aplicaciones 2D nativas de *Squeak* junto con aplicaciones 2D mono-usuario standard como navegadores web, procesadores de texto, planillas de cálculo, etc., dentro de un entorno multi-usuario colaborativo, por medio del cliente VNC que tiene incorporado.

4.6. Entorno de Programación

En cuanto a las facilidades de programación, *Second Life* posee un lenguaje de programación basado en eventos y estados, denominado **LSL** (*Linden Scripting Language*). Este lenguaje, que posee una sintaxis similar al lenguaje C, permite responder a una serie de eventos. Los eventos pueden ser generados por el mismo sistema, como los timers, o por otros usuarios, como los eventos de presencia, de chat o email, y también pueden ser producidos otros objetos 3D, por ejemplo, en el caso que se produzca una colisión.

La interfaz de este lenguaje de *scripting* proporciona más de 30 tipos de eventos y más de 400 funciones, a partir de las cuales se escriben pequeños fragmentos de código denominados *scripts*, los cuáles son interpretados y ejecutados por el servidor. En este sentido, el modelo de programación consiste en que cada objeto contiene sus propios scripts, que son los encargados de establecer el comportamiento del objeto.

Una de las principales críticas que se ha hecho a este modelo es que al no existir el concepto de clase, ni de objeto, como se entiende dentro del paradigma de programación orientada objetos, cada elemento es un prototipo que responde reactivamente a ciertos *eventos*. Por lo tanto, sus limitaciones son evidentes, ya que no existen bibliotecas de scripts que permitan su reuso, ni control de versiones, y resulta muy difícil desacoplar a los scripts de los objetos 3D. [8]

Los scripts en *OpenSimulator* están basados en **LSL** (*Linden Scripting Language*) con la intención de mantener cierto grado de compatibilidad. Sin embargo, no todas las funciones presentes en *Second Life* han sido implementadas, y por otro lado, se han incorporado algunas funciones nuevas.

El hecho que *OpenCobalt* esté íntegramente programado en *Squeak* le permite al usuario acceder todas las herramientas de programación y depuración propias de dicho lenguaje desde el mismo ambiente. *Squeak*, es un lenguaje de programación orientado a objetos, basado en clases, con capacidades de programación reflexiva y metaprogramación, que se ejecuta sobre una *Máquina Virtual* específica para cada plataforma, lo que hace que el código desarrollado sea totalmente portable inclusive entre diferentes sistemas operativos.

4.7. Resultados obtenidos

Luego de la aplicación del marco comparativo a los ambientes virtuales 3D seleccionados, referidas a cada rasgo en particular, se integraron los resultados de forma tal de obtener una representación gráfica comparativa (Spider Chart) de las seis características evaluadas, como se aprecia en la Figura 3.

5. Conclusiones

A partir de los resultados obtenidos puede afirmarse que *Second Life* se destaca por los aspectos estéticos y de comunicación entre usuarios. *OpenSimulator* aventaja al anterior por ser una aplicación de código abierto y brindar a los usuarios la posibilidad de instalar sus propios servidores. Finalmente *OpenCobalt* a

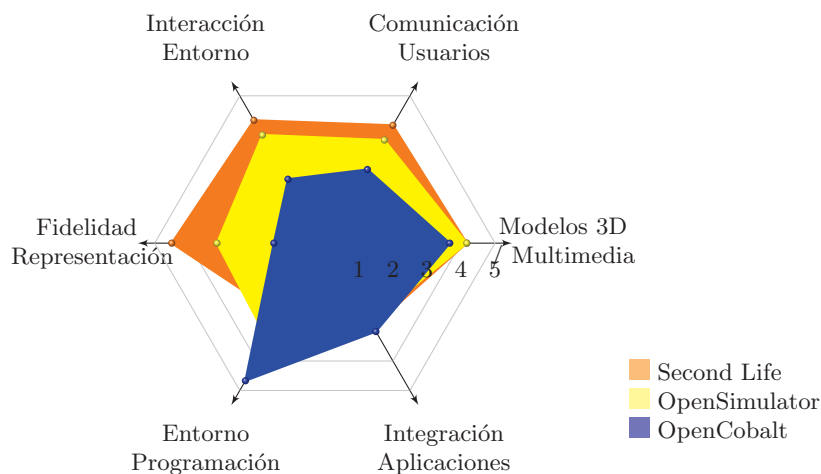


Fig. 3. Integración de las características analizadas

pesar de ser más limitado en el aspecto estético, supera a los anteriores fundamentalmente en lo que se refiere al lenguaje y las herramientas de programación, un aspecto clave para la implementación de simulaciones.

Referencias

1. Quinche, Juan, González, Franci. Entornos Virtuales 3D, Alternativa Pedagógica para el Fomento del Aprendizaje Colaborativo y Gestión del Conocimiento en Uniminuto. *Revista Formación Universitaria*. 4, pp.45-54. (2011)
2. Black, Andrew, Ducasse, Stéphane, Nierstrasz, Oscar, Pollet, Damien, Cassou, Damien, Denker, Markus : Squeak by Example. Square Bracket Associates, Switzerland (2009)
3. De Freitas, Sara: Serious Virtual Worlds. A scoping study. *JISC Virtual Worlds & Serious Gaming Resources*, pp.4-49 (2008)
4. Blas, Nicoletta, Garzotto, Franca, Poggi, Caterina. Web Engineering at the Frontier of the Web 2.0: Design Patterns for Online 3D Shared Spaces. *World Wide Web*. 4. 12. pp.345-379. (2009)
5. Tsiatsos, Thrasyvoulos, Konstantinidis, Andreas, Pomportsis, Andreas S. Evaluation Framework for Collaborative Educational Virtual Environments. *Educational Technology & Society*. 2. 13. pp.65-77. (2013)
6. Fredes, Claudio, Hernández, Juan, Díaz, Daniel. Potencial y Problemas de la Simulación en Ambientes Virtuales para el Aprendizaje. *Revista Formación Universitaria*. 1. 5. pp.45-56. (2012)
7. Dos Santos, Renato: Second Life as a Platform for Physics Simulations and Microworlds: An Evaluation. *Proceedings of the CBLIS 2012*, pp.173-180, Centre for Research in Science and Mathematics, Barcelona (2012)
8. Lopes, Cristina: The Worst Language Ever Designed : The Case for Better Programming Languages for 3D Environments. In: *POPL 2014: 41st ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages*, California (2014)

Propuesta de un Framework para la creación de Objetos de Aprendizaje.

María Lucía Violini^{1,2}, Cecilia Verónica Sanz^{1,3}, Patricia Mabel Pesado¹

¹ Instituto de Investigación en Informática LIDI,
Facultad de Informática, Universidad Nacional de La Plata.

² Becaria TIPO A de la Universidad Nacional de La Plata.

³ Investigadora Asociada de la Comisión de
Investigaciones Científicas de la Provincia de Buenos Aires.
La Plata, Buenos Aires, Argentina.

{lviolini, csanz, ppesado}@lidi.info.unlp.edu.ar

Resumen. En este artículo se presenta el diseño inicial de un *Framework para la creación de Objetos de Aprendizaje* (OAs), basado en una metodología específica. Previamente, se hace una revisión sobre *Metodologías de diseño de OAs* y se destaca la falta de *Herramientas para la creación de OAs* que consideren metodologías de diseño. En busca de atender esta necesidad, se propone el *framework* (denominado *MarCOA*) que integra las etapas de análisis, diseño, desarrollo, publicación y evaluación de OAs, establecidas por la *Metodología CROA*. Como primera evaluación del *framework* propuesto, se realiza una muestra del diseño inicial a expertos en la temática, lo que arroja resultados preliminares para analizar y continuar con su evolución.

Palabras clave: Objetos de Aprendizaje, Metodología CROA, Framework.

1 Introducción

La incorporación de las *Tecnologías de la Información y la Comunicación* (TICs) en Educación ha permitido enriquecer los escenarios donde se producen los procesos de enseñanza y aprendizaje [1]. Para aprovechar las TICs en los contextos educativos se hace necesario considerar los aciertos pedagógicos y tecnológicos, y tener en cuenta la interdisciplinariedad [2]. Al mismo tiempo, para evaluar las posibilidades que ofrecen los recursos tecnológicos en apoyo al aprendizaje, se requiere de la participación colectiva de diversas disciplinas [3]. Un ejemplo que denota la integración del contexto educativo y tecnológico actual es la utilización de *Objetos de Aprendizaje* (OAs).

El origen del término *Objeto de Aprendizaje* (OA) se atribuye a Wayne Hodgins en 1992 [4]. Desde ese entonces, y hasta ahora, no se ha establecido una definición formal de OA. Los autores especializados en el tema suelen coincidir en ciertas características que un OA debería tener para ser considerado como tal, se acuerda en que debe ser un material: digital, reutilizable y diseñado con una intencionalidad pedagógica [5]; sin embargo, la mayoría de los autores propone su propia

conceptualización del término. Para este trabajo se adopta el siguiente concepto: “*Un OA es un tipo de material educativo digital, que se caracteriza, desde el punto de vista pedagógico, por orientarse a un objetivo específico de aprendizaje, y presentar mínimamente: una serie de contenidos con el fin de abordar la temática relacionada con el objetivo, actividades que permitan al alumno poner en práctica o problematizar el contenido presentado, y una autoevaluación que posibilite conocer al alumno, si ha podido comprender esos contenidos vinculados al objetivo. Desde el punto de vista tecnológico, se caracteriza por contener un conjunto de metadatos estandarizados para su búsqueda, y recuperación, y estar integrado, utilizando un modelo de empaquetamiento que respete estándares, y de esta manera, permita su diálogo con diferentes entornos tecnológicos.*” [6] [7].

La falta de una definición consensuada de OA ha dado lugar al surgimiento de una serie de *Metodologías de diseño de OAs*, creadas bajo distintos puntos de vista conceptuales. En la siguiente sección se retoma y profundiza esta idea.

El núcleo de este trabajo consiste en presentar un primer diseño de un *Framework para la creación de OAs* basado en una *Metodología de diseño de OAs* en particular. El *framework* busca contemplar todos y cada uno de los aspectos que hacen al diseño y la creación de OAs, de acuerdo a la metodología específica. Intenta ser un medio para que cualquier docente, interesado en crear sus propios OAs, pueda hacerlo de manera completa e intuitiva, integrando en una herramienta las etapas de: análisis, diseño, desarrollo, publicación y evaluación. Además, se describe la primera instancia de evaluación realizada, ya que, se pretende construir el *framework* de manera evolutiva, con un diseño participativo, haciendo los ajustes que sean necesarios acorde a resultados que se vayan obteniendo en las validaciones con potenciales usuarios y expertos en la temática.

A partir de acá, el artículo se organiza de la siguiente manera. En la sección 2, se realiza una breve revisión sobre *Metodologías de diseño de OAs*, se describen las características principales que éstas poseen y se hace foco en *Herramientas para la creación de OAs* basadas en metodologías de diseño. En la sección 3, se describe la propuesta del *Framework para la creación de OAs*, se detallan sus características generales y la manera en que éste acompaña a la *Metodología CROA*, además se presenta su diseño inicial. En la sección 4, se presenta la primera evaluación del diseño inicial del *framework* propuesto, realizada a través de un juicio de expertos, y se exponen y analizan los resultados preliminares obtenidos. Finalmente, en la sección 5, se detallan las conclusiones a las cuales se ha arribado y se proponen líneas de trabajo futuro.

2 Metodologías de diseño de Objetos de Aprendizaje

En general, las *Metodologías de diseño de OAs* tienen como finalidad guiar a los docentes en las tareas que involucra diseñar OAs. Cada metodología adopta, como punto de partida, un concepto de OA. Las metodologías suelen definir un conjunto de etapas que generalmente se organizan en secuencia, aunque algunas metodologías establecen otros modos de organización. Si bien las metodologías se centran en proveer guías de diseño de OAs, también suelen proporcionar guías de planificación

y/o de análisis previas a la etapa de diseño. Además, aunque en menor cantidad, hay metodologías que proporcionan guías de desarrollo, publicación y evaluación.

Debido a que los OAs involucran tanto aspectos tecnológicos como aspectos pedagógicos [8] [9] [10], una buena metodología debería contemplar ambas dimensiones [11]. Se distinguen distintas clases de *Metodologías de diseño de OAs* de acuerdo al tipo de aspectos que predominan en ellas.

A continuación se establece una clasificación de *Metodologías de diseño de OA* y, para cada clase establecida, se describe una de sus metodologías más representativas:

- 1) **METODOLOGÍAS CON ENFOQUE TECNOLÓGICO:** Consideran mayormente aspectos tecnológicos, están fuertemente ligadas al desarrollo de *software*. Este es el caso de la *Metodología MESOVA (Metodología de desarrollo de Software para Objetos Virtuales de Aprendizaje)*. Esta metodología asume que la construcción de OAs debe enfocarse como un trabajo interdisciplinario pero orientado desde la *Ingeniería de Software*, pues considera que básicamente se trata de producir tecnología para aplicar en procesos de aprendizaje. La misma está estructurada en cinco fases: (1) concepción del objeto, (2) diseño y desarrollo modular evolutivo, (3) integración y despliegue del objeto, (4) pruebas de aprendizaje, (5) consolidación; cada etapa con sus respectivas actividades y especificación de resultados. Considera que un módulo es una parte temática o metodológica del OA [12] [13].
- 2) **METODOLOGÍAS CON ENFOQUE PEDAGÓGICO:** Destacan aspectos pedagógicos, están más vinculadas al diseño instruccional. Como ser la *Metodología ISDMELO (Instructional System Development – Conjunto de procedimientos para diseñar y desarrollar sistemáticamente instrucción)*. Esta metodología tiene soporte en *Teorías de Aprendizaje*. La misma permite diseñar y desarrollar contenido educativo a ser distribuido por la *web*. Usa un enfoque *top-down*, con dimensión pedagógica en varios niveles. Considera un OA como una pieza pequeña de contenido de aprendizaje que se concentra en un objetivo específico. Consta de cinco fases: (1) análisis, (2) diseño, (3) desarrollo, (4) implementación, (5) evaluación [14] [13].
- 3) **METODOLOGÍAS CON ENFOQUE HÍBRIDO:** Combinan, de manera equilibrada, aspectos tecnológicos y pedagógicos. Acá se encuentra la *Metodología CROA (Metodología para la Creación de Objetos de Aprendizaje)*. Esta metodología busca: (1) dar el mayor nivel de detalle posible para que se pueda alcanzar la etapa de desarrollo del OA, (2) permitir que sea el mismo docente quien pueda construir sus propios OAs, (3) trabajar con una granularidad que posibilite reutilización y ensamblaje de OAs. Propone el recorrido de una serie de etapas que orientan al autor del OA tanto en su diseño instruccional como en su diseño tecnológico, consta de un documento guía y anexos que profundizan la información de cada etapa [6] [7] [13]. En Tabla 1 se describe brevemente cada una de las etapas que componen la *Metodología CROA*, que serán la base del *framework*. Entre las etapas de diseño y desarrollo, *CROA* propone realizar un análisis de coherencia interna del OA. Este análisis invita al autor del OA a reflexionar sobre el diseño realizado antes de abordar su desarrollo.

Tabla 1. Etapas de la Metodología CROA.

ETAPA DE CROA	DESCRIPCIÓN
1. ANÁLISIS	Permite definir las necesidades educativas que dan origen al OA.
2. DISEÑO	Permite abordar el diseño instruccional del OA, definir la estructura que tendrá el OA y abordar el diseño multimedial del OA.
3. DESARROLLO	Comprende: selección de recursos para incorporar en el OA, uso/desarrollo de las plantillas previamente elegidas, integración de recursos y plantillas para conformar el OA, carga de metadatos que describen al OA y su empaquetado final.
4. PUBLICACIÓN	Propone la publicación del OA en entornos tecnológicos que respeten estándares de empaquetamiento.
5. EVALUACIÓN	Permite analizar las posibilidades y barreras encontradas en el uso del OA en contextos específicos.

2.1 Herramientas para la creación de Objetos de Aprendizaje basadas en Metodologías de diseño de Objetos de Aprendizaje

Las *Herramientas para la creación de OAs* basadas en *Metodologías de diseño de OAs* permiten a los docentes generar OAs, a través de la posibilidad de implementación y/o integración de sus distintos componentes, de acuerdo a una metodología de diseño específica. Estas herramientas contemplan etapas de: planificación, análisis, diseño, desarrollo, entre otras; consideran aquellas etapas que establece la metodología en cuestión.

Investigaciones previas de los autores de este trabajo han evidenciado la falta de *Herramientas para la creación de OAs* basadas en *Metodologías de diseño de OAs* específicas. Existen pocas herramientas de este tipo en comparación con las herramientas que no consideran una metodología (como ser: *eXeLearning*¹, *Ardora*², *REALOD Editor*³, sólo por mencionar algunas), a pesar de los importantes beneficios que traería para los docentes contar con ellas en sus procesos de creación de OAs [15].

En la próxima sección se presenta el núcleo de este trabajo, que se vincula con la propuesta de un *Framework para la creación de OAs* basado en la *Metodología CROA*.

3 Framework propuesto para la creación de Objetos de Aprendizaje

En esta sección se presenta el *framework* propuesto, al que se denomina *MarCOA* (*Marco para la Creación de Objetos de Aprendizaje*). En primer lugar, se expone la motivación por la cual se decide crear esta herramienta. Luego, se detallan sus

¹ *eXeLearning*: <http://exelearning.net/>.

² *Ardora*: <http://webardora.net/>.

³ *RELOAD Editor*: <http://www.reload.ac.uk/editor.html>.

características generales y se plantea cómo acompaña las etapas de *CROA*. Finalmente, se aborda su diseño inicial.

3.1 Motivación

El proceso de creación de OAs no es sencillo, involucra cuestiones pedagógicas y tecnológicas no triviales. Cuán importante sería para los docentes, al momento de crear OAs, disponer de herramientas de manejo intuitivo que asistan desde un comienzo, que proporcionen guías de: planificación, análisis, diseño y desarrollo y que acompañen también en etapas posteriores como: la publicación del OA y la evaluación del mismo. Sin embargo, se vislumbra la falta de *Herramientas para la creación de OAs* que se basen en *Metodologías de diseño OAs* actuales y efectivas [15].

Intentando atender la necesidad detectada, surge la propuesta del *framework* que pretende: (1) ser una herramienta de manejo intuitivo para cualquier docente al momento de crear OAs y (2) funcionar como guía/acompañamiento en el proceso completo que se requiere para crear un OA a partir de una metodología concreta.

3.2 Características generales

El *framework* que se propone es un “marco” para crear OAs, guía los pasos requeridos para crear un OA completamente. Considera todas las etapas que hacen a la creación de OAs de acuerdo a la *Metodología CROA*, incluye: análisis, diseño (instruccional, de estructura, multimedial) y desarrollo (implementación y carga de metadatos), y propone sugerencias para: publicación y evaluación. Incluye, además, un análisis de coherencia interna del OA.

El *framework* intenta no dejar de acompañar al docente en ningún momento durante el proceso de creación de OAs, se proporcionan guías y ayuda para cada una de las tareas. Se pretende que el *framework* pueda ser utilizado por cualquier docente, indistintamente del nivel de conocimientos que posea sobre manejo de herramientas de *software* y/o sobre programación, por lo que intenta proveer un manejo lo más intuitivo posible.

Además de permitir la generación del OA, el *framework* permite generar la documentación asociada a cada etapa previa al desarrollo. Por lo tanto, como resultado del proceso de creación, no sólo es posible exportar el paquete *SCORM*⁴ (que incluye los metadatos) correspondiente al OA sino que también se pueden exportar los documentos correspondientes a las etapas de análisis y diseño del OA.

Con respecto a las dos últimas etapas de la *Metodología CROA*, publicación y evaluación, el *framework* brinda sugerencias en cada una, para que el docente pueda llevarlas a cabo. Así, el *framework* abarca el proceso completo que *CROA* propone para crear OAs. Adicionalmente, permite configurar cuestiones referidas a la exportación del paquete *SCORM* del OA y la documentación asociada.

⁴ *SCORM*: <https://www.adlnet.gov/adl-research/scorm/>.

3.3 Seguimiento de CROA

La elección de *CROA* como metodología de base del *framework* se fundamenta principalmente por: (1) el enfoque híbrido que la metodología posee y (2) lo completa que resulta ser la metodología para diseñar y crear OAs. Cabe destacar que la elección de la metodología de base es un punto clave, ya que ésta determinará la potencialidad del *framework* resultante. Por un lado, contar con una metodología de base que complemente equitativamente aspectos tecnológicos con aspectos pedagógicos representa un buen escenario de partida para el *framework*. Por otro lado, contar con una metodología que abarque desde la etapa de análisis hasta la etapa de evaluación, pasando por todas las etapas y pasos intermedios que hacen a la creación de OAs, también representa un requerimiento de interés. Asimismo, la metodología ha sido utilizada en el marco de la *Facultad de Informática* de la *Universidad Nacional de La Plata*, por lo que se trata de una metodología ya probada.

A continuación se describe la manera en que el *framework* acompaña cada una de las etapas de la *Metodología CROA*:

- 1) ETAPA DE ANÁLISIS: El *framework* presenta una serie de preguntas, definidas por *CROA*, que permiten plantear la necesidad educativa que da origen al OA.
- 2) ETAPA DE DISEÑO: El *framework* determina tres secciones: (1) diseño instruccional, (2) estructura, (3) diseño multimedial. La primera sección presenta una serie de preguntas, definidas por *CROA*, que hacen al diseño instruccional del OA. La segunda sección permite diseñar y planificar el recorrido entre los diversos nodos del OA. Y la tercera, permite determinar los tipos de escenario que se utilizarán en el OA y el mapeo entre nodos y plantillas multimedia.
- 3) ETAPA DE DESARROLLO: El *framework* determina dos secciones: (1) implementación, (2) metadatos. La primera sección permite, a partir de la estructura del OA definida, armar e integrar los componentes del OA. Y la segunda, permite, de acuerdo a las consideraciones de *CROA*, cargar los metadatos del estándar *IEEE-LOM*⁵ que identifican y describen al OA y facilitan su reutilización.
- 4) ETAPA DE PUBLICACIÓN: El *framework* muestra las sugerencias de *CROA*, y los pasos a seguir, para realizar la publicación del OA en un repositorio o un *Entorno Virtual de Enseñanza y Aprendizaje* (EVEA), es decir, en diferentes entornos tecnológicos que respeten el estándar de empaquetamiento utilizado.
- 5) ETAPA DE EVALUACIÓN: El *framework* muestra las sugerencias de *CROA*, y los pasos a seguir, para evaluar el OA en contextos de uso específicos. Se propone considerar ciertos ejes para analizar la calidad del OA, según *CROA*.

El *framework* también acompaña el análisis de coherencia interna del OA que propone *CROA*, presentando una serie de preguntas que componen la guía, definida por la metodología, para asegurar la coherencia interna entre los componentes de diseño.

⁵ *IEEE-LOM*: <https://standards.ieee.org/findstds/standard/1484.12.1-2002.html>.

3.4 Diseño inicial

El diseño inicial realizado consta de 25 potenciales pantallas del *framework*. Se hizo una representación de cada pantalla con el objetivo de reflejar su funcionalidad, y no el estilo visual. A continuación se muestra y describe una de dichas representaciones, elegida por ser la más significativa del *framework*.

En la Fig. 1 se muestra una representación de la pantalla principal de creación de OAs del *framework*. Esta pantalla propone, al usuario del *framework*, hacer un recorrido específico por: las etapas que define *CROA* para la creación de un OA y su exportación; los pasos a seguir se muestran numerados (de 1 a 7), a modo de guía para el docente. Al comienzo, sólo la primera etapa estará habilitada y, a medida que el docente avance, se irán habilitando las etapas restantes; se intenta orientar al docente según lo establece *CROA* (por ejemplo: no sería recomendable que un docente aborde el desarrollo de un OA si aún no pensó en su diseño). Para cada etapa se indica (mediante un símbolo representativo) el porcentaje de avance que el usuario lleva realizado en esa etapa. Además, en esta pantalla, se tendrá al alcance la definición de OA según *CROA*, por si es necesario revisarla.

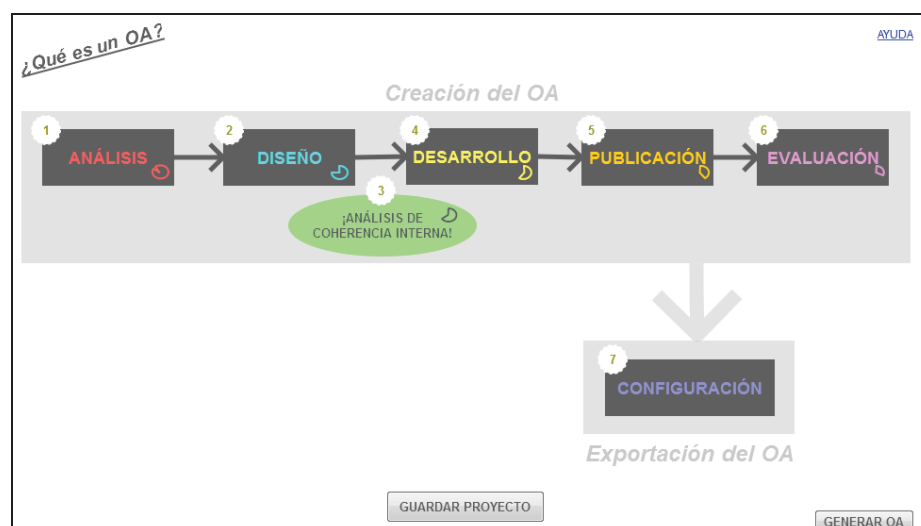


Fig. 1. Representación de la pantalla principal de creación de OAs del *framework*.

4 Primera evaluación del diseño inicial de MarCOA

En esta sección se presenta la primera evaluación realizada sobre el diseño inicial del *framework* propuesto. Resulta relevante para la construcción de *MarCOA*, ya que, se pretende alcanzar la versión final de manera evolutiva, realizando validaciones parciales a medida que se avanza y considerando los resultados. Se sigue así un diseño participativo. Aquí, en primer lugar, se explica cómo ha sido la técnica

utilizada para esta primera evaluación. Luego, se exponen y analizan los resultados obtenidos.

4.1 Juicio de expertos

Como primera instancia de evaluación, se realizó una muestra del diseño inicial del *framework* a expertos en la temática. Para ello, se consideraron dos perfiles de expertos y se confeccionaron dos tipos de encuestas respectivamente.

En cuanto a los expertos, los perfiles considerados fueron: (1) teóricos de OAs y (2) creadores de OAs. Por un lado, se encuentran aquellos expertos que se especializan en las bases teóricas de los OAs y conocen detalladamente la *Metodología CROA* y, por otro lado, están los docentes que han creado OAs en el marco de sus propuestas educativas y posiblemente sean futuros usuarios del *framework*.

En cuanto a las encuestas, los tipos confeccionados fueron: (1) para teóricos de OAs y (2) para creadores de OAs. La encuesta para teóricos apuntó a evaluar principalmente el diseño inicial del *framework* en relación a la *Metodología CROA*, detectando: en qué medida el *framework* refleja la metodología, con qué nivel de detalle y organización respeta las etapas de *CROA*, qué podría agregarse o sacarse del diseño, entre otros aspectos evaluados. En cambio, la encuesta para creadores pretendió evaluar principalmente el diseño inicial del *framework* desde la experiencia previa del experto creando OAs, detectando: cuán intuitivo y claro resulta este *framework*, cuán acompañado está el usuario en el proceso completo de creación de un OA, cuán comprensible es la manera en que se organiza, si resultaría favorable contar con este *framework* para crear OAs, entre otros aspectos evaluados. Además, para ambos tipos de encuesta, se consideraron preguntas que invitan al experto a dejar observaciones, comentarios adicionales, sobre determinadas cuestiones. Para los dos tipos de encuesta, se utilizaron respuestas: (1) en escala tipo Likert de 1 a 5 (aumentando gradualmente el nivel de acuerdo/coincidencia con lo planteado) y (2) abiertas (para desarrollar); para todas las respuestas en escala se dio la posibilidad de ampliar con comentarios.

Los expertos encuestados fueron 5 en total: 2 teóricos de OAs y 3 docentes creadores de OAs. La dinámica de las sesiones de trabajo individual con cada experto tuvo: una primera fase de presentación del diseño, en la cual se mostraron las potenciales pantallas del *framework* y se describieron sus características generales, y una segunda fase de resolución de la encuesta. Al mismo tiempo, se registraron los comentarios de cada experto a medida que hacía el recorrido por las pantallas diseñadas. Se pretende, en el corto plazo, continuar mostrando el diseño inicial del *framework* a otros expertos y contar con más devoluciones acerca del mismo.

4.2 Resultados preliminares

Los resultados preliminares, obtenidos a partir del juicio de expertos realizado, permiten reflexionar sobre diversas cuestiones referidas al diseño inicial de *MarCOA*.

Con respecto a la encuesta realizada a 2 teóricos de OAs, la gran mayoría de las respuestas obtenidas fue de nivel 5. Esto revela que existe concordancia, por parte de los expertos teóricos encuestados, al momento de observar una estrecha relación entre el *framework* y la *Metodología CROA*. Partiendo de la idea de *CROA* como base del *framework*, los resultados indican que se está por buen camino.

Con respecto a la encuesta realizada a 3 docentes con experiencia previa creando OAs, la gran mayoría de las respuestas obtenidas fue de nivel 5. Esto refleja que el diseño inicial del *framework*, desde la mirada de los expertos creadores encuestados, resulta ser: intuitivo, claro y útil. Si la premisa es desarrollar un *framework* que pueda ser utilizado por cualquier docente, tenga éste conocimientos sobre herramientas de *software* o no, los resultados son optimistas en este sentido. Además, pretendiendo proporcionar una herramienta que guíe el proceso completo de creación de OAs, los resultados indican que los docentes creen favorable contar con un *framework* así.

Para finalizar el análisis, se destaca que los 5 expertos encuestados dieron una devolución satisfactoria con respecto al *framework*, todas sus apreciaciones fueron positivas. Algunos de ellos sugirieron cuestiones concretas que serán consideradas y contribuirán a la evolución del *framework*. Por ejemplo: dos de los expertos encuestados sugirieron ampliar la etapa de evaluación, haciendo que el *framework* proporcione al docente posibles encuestas a realizar a sus alumnos para evaluar el uso del OA. Otro aporte: uno de los expertos encuestados sugirió ampliar la etapa de publicación, haciendo que el *framework* permita llevar adelante la publicación del OA, en un repositorio de OAs y/o en un EVEA, de forma automática.

5 Conclusiones y Trabajo futuro

Luego de la investigación realizada y de analizar los resultados obtenidos a partir de la primera evaluación sobre el diseño inicial de *MarCOA*, se presentan a continuación conclusiones a las cuales se ha arribado.

Habiendo realizado una revisión sobre *Metodologías de diseño de OAs*, se distinguieron tres clases: (1) con enfoque tecnológico, (2) con enfoque pedagógico y (3) con enfoque híbrido; correspondiendo a esta última clase la *Metodología CROA*, elegida como base del *framework* propuesto. Se detectó la falta de *Herramientas para la creación de OAs* basadas en metodologías de diseño, lo que motivó el origen del *framework*.

Se propuso un *framework* que pretende acompañar al docente durante el proceso de diseño que define *CROA* para la creación de un OA, mediante guías y ayuda. Se realizó una primera evaluación de su diseño inicial a través de un juicio de expertos. Esto arrojó resultados optimistas/positivos que muestran que, tanto para teóricos especializados en OAs como para docentes que poseen experiencias previas creando OAs, el diseño inicial del *framework* es claro e intuitivo y comprende completamente las etapas de *CROA*. Según los resultados, para los expertos, sería favorable que los docentes dispongan de este *framework* al momento de crear OAs.

Como trabajo futuro se pretende, en principio, continuar con el proceso de evaluación del diseño inicial del *framework* propuesto. Se cuenta con una primera validación del mismo, obtenida a través del juicio de expertos, pero se espera

completar el proceso de evaluación y atender a las sugerencias que se fueron realizando. De esta manera, se espera que *MarCOA* sea un *framework* que resulte de evoluciones sucesivas a partir de un diseño participativo.

Referencias

1. Adell Segura, J., Castañeda Quintero, L.: Los Entornos Personales de Aprendizaje (PLEs): una nueva manera de entender el aprendizaje. En Roig Vila, R. & Fiorucci, M. (Eds.) Claves para la investigación en innovación y calidad educativas. La integración de las Tecnologías de la Información y la Comunicación y la Interculturalidad en las aulas (2010)
2. Callejas Cuervo, M., Hernández Niño, E. J., Pinzón Villamil, J. N.: Objetos de aprendizaje, un estado del arte. *Entramado*, vol. 7, no. 1 (2011)
3. Herrera, M.: Las fuentes del aprendizaje en ambientes virtuales. Universidad Autónoma Metropolitana (2001)
4. Hodgins, H. W.: The future of learning objects. *The Instructional Use of Learning Objects: Online Version* (2000)
5. Astudillo, G. J.: Análisis del estado del arte de los objetos de aprendizaje. Revisión de su definición y sus posibilidades. Trabajo Final de Especialización, Facultad de Informática, Universidad Nacional de La Plata (2011)
6. Metodología CROA de la Universidad Nacional de La Plata, <http://croa.info.unlp.edu.ar/>
7. Sanz, C., Barranquero, F., Moralejo, L.: CROA: A learning object design and creation methodology to bridge the gap between educators and reusable educational material creation. *EDULEARN16 Proceedings*, pp. 4583--4592 (2016)
8. Redeker, G. H.: An educational taxonomy for learning objects. In *Advanced Learning Technologies, 2003. Proceedings. The 3rd IEEE International Conference on*, pp. 250--251 (2003)
9. Vidal Castro, C., Ferreira Satler, M.: Modelo Ontológico para la Secuenciación de Objetos de Aprendizaje. *IEEE-RITA*, vol. 5, no 2, pp. 63--67 (2010)
10. Motz, R., Badell, C., Barrosa, M., Sum, R.: La Extracción de Objetos de Aprendizaje con Metadatos de Diseño Pedagógico. *IEEE-RITA*, vol. 5, no. 2, pp. 49--55 (2010)
11. Silva Sprock, J., Ponce Gallegos, C., Hernández Bieliukas, Y.: Estado del Arte de las Metodologías para el Desarrollo de Objetos de Aprendizaje. *Conf. LACLO*, vol. 4, no. 1. (2013)
12. Parra Castrillón, E.: Propuesta de metodología de desarrollo de software para objetos virtuales de aprendizaje-MESOVA. *Revista Virtual Universidad Católica del Norte*, vol. 1, no. 34, pp. 113--137 (2011)
13. Sanz, C., Barranquero, F., Moralejo, L.: Diseño y Producción de Objetos de Aprendizaje. Curso de postgrado, Facultad de Informática, Universidad Nacional de La Plata (2016)
14. Baruque, L. B., Melo, R. N.: Learning theory and instructional design using learning objects. *Journal of Educational Multimedia and Hypermedia*, vol. 13, no. 4, pp. 343 (2004)
15. Violini, L., Sanz, C.: Herramientas de Autor para la creación de Objetos de Aprendizaje. *ESTADO DEL ARTE. XXII Congreso Argentino de Ciencias de la Computación* (2016)

ReadMe. Complemento de Aprendizaje Móvil

Mg. José Luis Filippi¹, I.S. Héctor Daniel Perez², I.S. Sofía Aguirre³, Mg. Rodolfo Bertone⁴

GIAU⁴ - Facultad de Ingeniería - UNLPam.
{filippij¹, daniel.perez², saguirre³}@ing.unpam.edu.ar
pbertone@ada.info.unlp.edu.ar⁴

⁴Grupo de Investigación de Ambientes Ubicuos

Resumen. El contexto socio-cultural, caracterizado por la presencia ubicua y el uso intensivo de las Tecnologías de la Información y la Comunicación, coloca a las instituciones educativas frente a la demanda de desarrollar en sus alumnos la alfabetización digital necesaria para la utilización competente de las herramientas tecnológicas [1]. Con el objetivo de explorar nuevas oportunidades que ofrecen los dispositivos móviles en el ámbito educativo universitario, se decide desarrollar una aplicación que permite al dispositivo leer la documentación que las cátedras distribuyen de formato digital; y facilitar al estudiante la incorporación de la información a través del audio.

Keywords: Aprendizaje móvil. Aplicación móvil. Texto a Voz. Tecnología Móvil.

1 Introducción

La presencia de la tecnología es un fenómeno ubicuo e irreversible en el mundo contemporáneo, y las instituciones educativas no pueden desconocer esta realidad si quieren formar a personas que sean capaces de integrarse en esta nueva sociedad de modo pleno. En efecto:

“(...) los sujetos que no sepan desenvolverse con la tecnología digital de un modo inteligente (conectarse y navegar por redes, buscar información útil, analizarla y reconstruirla, comunicarla a otros usuarios) no podrán acceder a la cultura y el mercado de la sociedad de la información. Es decir, aquellos ciudadanos que no estén calificados para el uso de las TIC tendrán mayores probabilidades de ser marginados culturales en la sociedad del siglo XXI. Este analfabetismo digital provocará, seguramente, mayores dificultades en el acceso y la promoción en el mercado laboral, indefensión y vulnerabilidad ante la manipulación informativa, (e) incapacidad para la utilización de los recursos de comunicación digitales.” [2].

Las redes virtuales móviles facilitan la movilidad del conocimiento, accesible en cualquier momento y en cualquier lugar. La portabilidad, interactividad e individualidad [3], sumado a la inmediatez, conectividad, ubicuidad y adaptabilidad de estos dispositivos aumentan las potencialidades educomunicativas de la tecnología móvil y, con ellas, las oportunidades para un cambio de paradigma educativo en el

contexto de la sociedad digital. A partir de la convergencia de los dispositivos móviles en las instituciones educativas, organismos de gobierno, entidades de acción social, entre otras, se impone en la sociedad actual un nuevo modelo de formación, el aprendizaje móvil.

Con el sentido permanente de alentar nuevas prácticas educativas acordes a las demandas que la sociedad actual realiza, se incorporan los dispositivos móviles en la formación universitaria como instrumento complementario a la formación presencial. El aprendizaje tiene lugar no solo en el aula, su accionar se despliega a cualquier lugar y en cualquier momento, se está gestando un nuevo método de formación, un aprendizaje permanente e ininterrumpido. [4]

Se trata de entender una educación que se caracterice por ser:

- Ubicua, presente en todas partes y al mismo tiempo.
- Inclusiva, que minimice las barreras para que toda la sociedad participe sin importar sus características físicas, mentales, sociales, culturales, entre otras. [5]
- Multiplataforma, con variados entornos educativos, como espacios de exploración e intercambio de información a partir de renovados contextos y tecnologías.
- Multimedia, con recursos tecnológicos de diversa naturaleza, herramientas libres y servicios en línea de fácil acceso.

Sin embargo, no todas las instituciones se adaptan al ritmo esperado por la sociedad vigente. En la Facultad de Ingeniería de la UNLPam., se observa una leve incorporación de diferentes aplicaciones móviles como recursos adicionales a los dispuestos por el docente en la formación tradicional. Se destacan aquellas aplicaciones móviles que permiten el intercambio de información y la mensajería de forma continua, a modo de ejemplo se puede señalar el uso de grupos de WhatsApp para mantener un canal de comunicación fluido con los estudiantes. Indagando la presencia de otros recursos que se caractericen por la movilidad, no se observa la incorporación de aplicaciones como instrumentos de apoyo a la formación presencial, que facilite al docente presentar materiales innovadores a partir de la mediación de las nuevas posibilidades que ofrecen los dispositivos móviles de última generación, y al estudiante ampliar sus habilidades intelectuales a partir de su agregación a la actividad cotidiana.

Se presenta un nuevo escenario educativo donde convergen diferentes tecnologías, que exige incorporar a las prácticas educativas tradicionales nuevas metodologías, que se caractericen por ser inmersivas y ubicuas. Gunther Kress en su conferencia sobre Exploraciones Metodológicas Multimodales considera que se debe enseñar a los estudiantes a incorporar diferentes medios de comunicación a través del cual se informa y al mismo tiempo, se fomentan las habilidades retóricas que necesitan para interactuar haciendo uso de diferentes plataformas. [6]

El entorno formativo cambia, se presenta desafiante, al mismo tiempo invita a reflexionar sobre cómo configurar nuevos ambientes educativos acordes a una sociedad irrupida por instrumentos tecnológicos inteligentes, reconocidos por toda la sociedad. En este contexto, incorporar y/o desarrollar aplicaciones móviles innovadoras, implica construir espacios que enriquezcan la práctica educativa, con propósitos claros, que estimulen tanto a docentes como estudiantes, trasladen más allá del espacio físico material, y permitan dar un valor agregado a la propia actividad.

2 Metodología

El grupo de investigación de ambientes ubicuos (GIAU) viene desarrollando actividades de investigación aplicada en situaciones concretas, destinada a la Facultad de Ingeniería de la UNLPam y restantes instituciones del medio local y regional en el que se encuentra inmerso.

Sin el propósito de efectuar un detalle minucioso de las diferentes acciones implementadas en la última década en el ámbito universitario, cabe destacar que se ha incorporado a la actividad educativa tradicional, formal y presencial, nuevas metodologías educativas mediadas por el uso de las nuevas tecnologías de la información y la comunicación, se pueden mencionar:

Incorporación de recursos informáticos que permiten el diseño y presentación de contenidos multimediales.

- Conformación de servidores para el almacenamiento permanente de datos.
- Disposición de redes que fomentan el trabajo colaborativo.
- Diseño de entornos educativos que posibilitan la formación en línea.
- Instauración de tutorías en comunidades de aprendizaje que posibilitan el aprendizaje permanente y continuo.
- Redefinición del diseño visual, acorde a dispositivos tecnológicos actuales.
- Agregación de una nueva narrativa configurada a partir de la inmediatez de los sucesos y particularidades de las aplicaciones móviles.
- Manejo de recursos de diversa índole.

En general, la formación influenciada por la aparición de los nuevos dispositivos tecnológicos, ofrecen métodos que permiten crear contenidos de forma dinámica y atractiva. Sin embargo, el uso de instrumentos tecnológicos sin la mediación pedagógica adecuada, conlleva una experiencia poco motivadora para los estudiantes.

Durante este tiempo con la incorporación de las TICs en la formación, es posible reafirmar que la tecnología cumple la función de facilitador del proceso educativo. Permite acceder rápidamente a cualquier tipo de información con solo un enlace, comunicarse con colegas que se encuentran en cualquier parte del mundo en videoconferencia y/o desarrollar trabajo colaborativo, compartir documentos electrónicos, entre muchas otras posibilidades.

Una problemática que se presenta a los estudiantes y/o docentes en permanente capacitación, yace en la gran cantidad de información que deben analizar a través de los diferentes documentos de texto. A la tradicional información que se presenta en soporte de papel se le añade información en formato digital; a través de documentos electrónicos, revistas científicas en línea, mensajería instantánea, presentaciones electrónicas, entre otros recursos que aparecen en forma diaria.

Es usual que la excesiva cantidad de información conlleve diferentes acciones por parte de los estudiantes y/o docentes, que al momento de destinar tiempo y lugar para leer y analizar de forma consciente la documentación, deciden efectuar diferentes estrategias acordes a las posibilidades de cada uno, por ejemplo:

- lectura rápida y superficial de la documentación que se les presenta.
- distribución entre varios de la información, con el objeto de reducir el tiempo y esfuerzo requerido para su lectura.
- utilizar el esfuerzo ajeno para leer comentarios y resumen de terceros.

- optar por no leer la documentación o leer en forma parcial.

Para atender esta problemática, se propone una opción complementaria a la lectura tradicional que incorpora el uso del dispositivo móvil. A sabiendas que el tamaño de la pantalla hace difícil la lectura visual, se solicita al dispositivo que reproduzca el texto en forma oral.

De igual forma que se escucha música, es posible que el teléfono móvil lea documentos de texto en voz alta, y los convierta en audio en el mismo instante en que se le solicita.

Algunas situaciones de la cotidianeidad que permiten hacer uso de esta propuesta, por ejemplo:

- En un medio de transporte público, el estudiante y/o docente se conecta a la red y accede a la documentación en formato digital, situada en algún entorno de aprendizaje.
- En aquellas prácticas deportivas que se desarrolla en forma individual no competitiva (caminata, ciclismo, etc.).
- En condiciones de espera de turnos en instituciones públicas o privadas (bancos, consultorios, etc.).

A sabiendas que tales escenarios se reiteran en forma periódica y haciendo uso de la ubicuidad que provee la tecnología actual (dispositivos inteligentes, redes wifi o redes de datos telefónicas), se propone implementar ReadMe una aplicación en proceso de desarrollo que permite a los usuarios acceder a documentos desde cualquier parte y en cualquier momento, y haciendo uso del sintetizador de voz convierte el texto en audio. Maggio incorpora propuestas de ésta índole al concepto de enseñanza enriquecida, donde los docentes componen escenarios educativos de alta disposición tecnológica. [7]

ReadMe es una aplicación móvil que se encuentra en fase de desarrollo. Diseñada para dispositivos móviles basados en sistema operativo Android versión 5 en adelante, efectuada en la Facultad de Ingeniería de la UNLPam por el grupo de investigación de ambientes ubicuos.

Es una aplicación que forma parte de un grupo de aplicaciones que se han desarrollado en el contexto del grupo de investigación, y que se ha de distribuir a los estudiantes y/o docentes que conforman la unidad académica.

3 Fases para el desarrollo de ReadMe

3.1 Estudio y selección de una plataforma de desarrollo

La plataforma que se utiliza para el desarrollo de la aplicación es Android Studio. Un entorno de programación multiplataforma de código abierto. La elección de esta plataforma se fundamenta en el alto índice de presencia del sistema operativo en los celulares, tabletas y otros dispositivos. Según encuesta efectuada en el segundo trimestre del año 2016 a nivel mundial, el 82,6% de dispositivos móviles utilizan sistema operativo Android. [8]

3.2 Análisis y diseño del sistema

Inicialmente se procedió a realizar un análisis de los que requerimientos del sistema que se pretende desarrollar. A partir de éstos, se examinaron aplicaciones de igual índole, encontrando aplicaciones gratuitas con publicidad, o de pago sin publicidad. Estas razones y la independencia de editar el código motivaron el desarrollo de ReadMe, una aplicación con particularidades propias.

3.3 Programación

Actualmente se encuentra en la etapa de programación en la cual se pueden destacar varias funcionalidades ya desarrolladas que corresponden a la versión 1.0.

Explorador de archivos, correspondiente a las diferentes unidades de almacenamiento que posee el dispositivo móvil, memoria externa e interna, y carpeta de descargas. (ver figura 1).

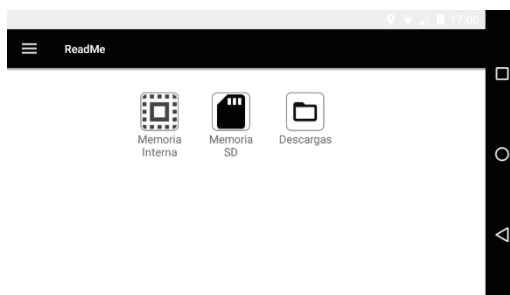


Fig. 1. Explorador de archivos

Multiplicidad de idiomas, permite conectar la aplicación al proveedor de idioma, descargar de internet el correspondiente empleado en el documento que se desea leer.

Visor de documentos de texto en formato txt, rtf y pdf. (ver figura 2).

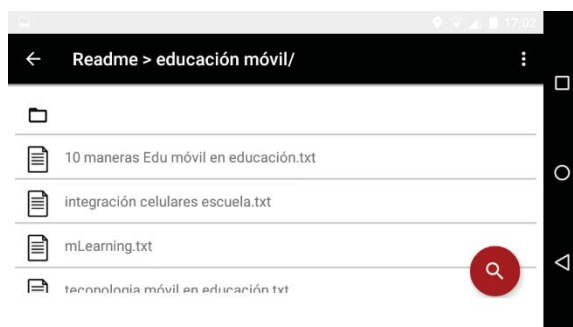


Fig. 2. Visor de documentos

Menú flotante en el margen inferior derecho, con variadas opciones que aparecen y se ocultan acorde a las necesidades del usuario como se puede observar en la figura 3. Las opciones disponibles son:

- Reproducir el documento completo.

- Reproducir por párrafo.
- Reproducir el texto seleccionado.
- Lupa, que permite agrandar el tamaño del texto para ampliar la vista.
- Guardar el documento.
- Compartir.
- Eliminar.
- Convertir el archivo de texto a archivo de audio.
- Reproducir archivos de audio.
- Capturar texto desde otras aplicaciones y almacenarlo en el dispositivo para ser tratado con posterioridad.
- Resumir.

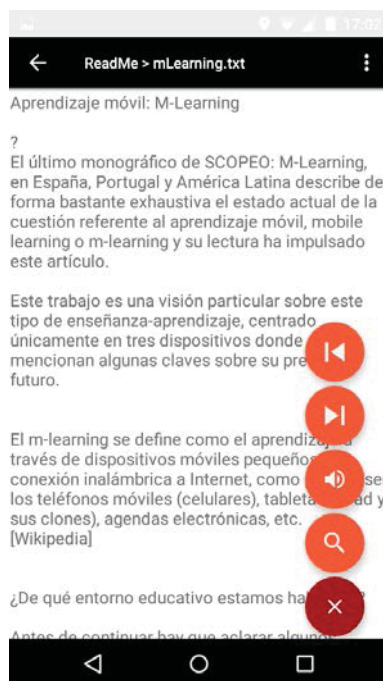


Fig. 3. Menú flotante

Otras funcionalidades que se pretenden incorporar a la próxima versión y que se encuentran actualmente en proceso de análisis y diseño son:

- Reproducir archivos de audio de terceros.
- Leer archivos de imágenes y gráficas en forma textual.
- Capturar texto desde documentos en soporte de papel y efectuar la digitalización automática para su posterior tratamiento.
- Pronunciar un texto en voz alta, capturar el audio e incorporar en formato de texto gramatical y/o audio.

4 Conclusiones

ReadMe es una aplicación móvil que facilita el acceso a la información disponible en formato digital textual y/o audio, permitiendo a estudiantes y docentes utilizar su funcionalidad en cualquier momento y lugar, gracias a la movilidad que ofrecen los dispositivos tecnológicos actuales.

Actualmente se está finalizando con la fase de desarrollo de la versión 1.0., con la expectativa que genera la fase de evaluación próxima, a la espera que los resultados sean satisfactorios para su difusión e implantación. La distribución se efectuará a partir del repositorio propio de la Facultad de Ingeniería de la UNLPam.

Al mismo tiempo que se iniciará el desarrollo de los módulos para la versión 2.0, según lo especificado en el apartado anterior.

5 Trabajos futuros

Se espera poder incorporar la aplicación como complemento innovador a las clases teóricas, explicaciones prácticas y textos recomendados por las cátedras. Asimismo, esto permitirá retroalimentar la aplicación con sugerencias tanto por parte de los docentes como de los mismos estudiantes.

Al mismo tiempo que se iniciará el desarrollo de los módulos para la versión 2.0, según lo especificado en el apartado anterior.

6 Referencias

1. Salinas, M. E.: Entornos virtuales de aprendizaje en la escuela: tipos, modelo didáctico y rol del docente. (2011).
2. Moreira, M. A., Salvat, B. G., & García-Quismondo, M. Á. M.: Alfabetizaciones y tecnologías de la información y la comunicación. Síntesis. (2008)
3. Klopfer, E., & Squire, K.: Environmental Detectives—the development of an augmented reality platform for environmental simulations. *Educational Technology Research and Development*, 56(2), 203-228. (2008)
4. Lam, J., Yau, J., & Cheung, S.: A review of mobile learning in the mobile age. *Hybrid learning*, 306-315 (2010)
5. Torres, C., & Sofia, K.: Diseño e implementación de una propuesta pedagógica para el fortalecimiento de las prácticas inclusivas a partir de los estilos de aprendizaje (Master's thesis, Universidad de la Sabana) (2016).
6. Multimodality: Methodological Explorations, <https://goo.gl/xF8KXc>.
7. Maggio, M.: Enriquecer la enseñanza: los ambientes con alta disposición tecnológica como oportunidad, Grupo Planeta Spain. (2012)
8. C. e. Español, “iOS vs. Android: ¿Quién va ganando en 2016?”, <https://goo.gl/v2Nieq>

Recta Numérica: Una Actividad de Códimo

Luciano Graziani, Gabriela Anahí Cayú, Matías Emanuel Sanhueza y Enrique Molinari

Universidad Nacional de Río Negro, Sede Atlántica, Viedma, Río Negro
<http://www.unrn.edu.ar/>
{lgraziani, gcayu, msanhueza, emolinari}@unrn.edu.ar

Abstract. El Pensamiento Computacional resulta de utilidad en situaciones de la vida cotidiana [9]. Introducirlo en la educación primaria y secundaria permitirá despertar el interés de los estudiantes en continuar sus estudios en ciencias de la computación, a la vez que posibilitará que desarrollen estas habilidades desde temprano para que puedan aplicarla en su vida cotidiana. En este trabajo presentamos una actividad diseñada para Códimo [5] y denominada Recta Numérica [1] que fue pensada para alumnos entre 6 y 8 años, y con la cual se busca incentivar de manera transparente uno de los aspectos del Pensamiento Computacional: el diseño de algoritmos. Los alumnos deben jugar a ubicar un número en la recta numérica, haciéndolo avanzar por un laberinto, mediante el uso de bloques (o comandos) simples. El alumno refuerza un conocimiento adquirido en clases (los números naturales, enteros y su ubicación en la recta), utilizando el pensamiento computacional.

Keywords: códimo, educación, pensamiento-computacional, objetos-de-aprendizaje, aprendiendo-a-programar, javascript.

1 Introducción

Desde el año 2014, la Universidad Nacional de Río Negro realiza distintas actividades para llevar la programación a las escuelas primarias y secundarias de la región. Durante todas ellas, el objetivo de las mismas fue contribuir a despertar vocaciones tempranas en carreras afines a las Ciencias de la Computación. A su vez, el Pensamiento Computacional es una herramienta de gran utilidad para todas las personas, no solo para aquellos decididos a continuar su vida profesional dentro de las Ciencias de la Computación [9].

Luego de dos años de experiencia llevando la programación a las escuelas utilizando herramientas como Alice, entendiendo las dificultades de docentes y alumnos, decidimos construir una herramienta que permitiera reforzar los conceptos dictados en clase utilizando habilidades del Pensamiento Computacional. A esta herramienta la nombramos Códimo.

En las actividades planteadas en Códimo, los alumnos refuerzan temas aprendidos en clases, ya sean de ciencias exactas, naturales, sociales u otra, aplicando conceptos de Ciencias de la Computación.

Este trabajo continúa en la sección 2 con la descripción de la propuesta, donde describimos en detalle cómo se juega y se aprende utilizando la actividad llamada “Recta Numérica”. En la sección 3 se describen detalles técnicos del desarrollo de Códimo; y en la sección 4 planteamos diferentes caminos posibles para continuar este trabajo.

2 “Recta Numérica”: Programar para Aprender

Es una actividad pensada para alumnos de primer ciclo de primaria (6 a 8 años). Se encuadra dentro del área de matemáticas, más concretamente luego de aprender números naturales, enteros, relaciones de orden y la recta numérica.

El objetivo siempre es ubicar el número dado en la recta numérica. La actividad presenta diferentes niveles de complejidad. En el nivel inicial se utilizan únicamente números naturales de una cifra; en el siguiente nivel números naturales de dos cifras; y en el último nivel, el más avanzado, se utilizan números enteros.

La consigna es desplazar cada número a través de un laberinto y ubicarlo dentro de la recta numérica, con la restricción de que solo es posible mover el número utilizando un conjunto limitado bloques. Cada bloque tiene una función específica, como mover el número en una dirección, o hacer que salte. Los bloques pueden encastrarse formando un conjunto de instrucciones que se aplicarán sobre el número. De esta forma, utilizando uno de los aspectos del Pensamiento Computacional, el diseño de algoritmos, los alumnos realizan la actividad.

El uso de bloques para programar tiene un fundamento y ventaja pedagógica por sobre la utilización de sentencias escritas de un pseudo lenguaje de programación. Los bloques siempre generan programas sintácticamente correctos, lo cual permite a los alumnos enfocarse únicamente en comprender la semántica o sea el efecto que producen los bloques sobre el número, en este caso, el movimiento del número a través del laberinto.

La Fig. 1 muestra uno de los ejercicios de nivel avanzado de la actividad “Recta Numérica”.

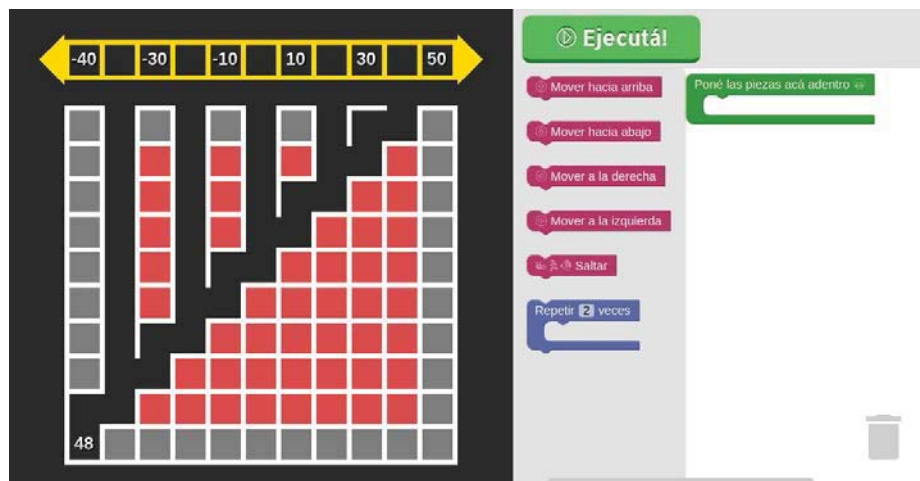


Fig. 1. Actividad Nivel Avanzado

Sobre la izquierda de la Fig. 1 se puede observar el laberinto y la recta numérica encima de él. La actividad inicia siempre con varios números sobre la recta y un número dentro del laberinto que hay que mover y llevar hacia la recta. En el ejemplo de la Fig. 1, el número que hay que llevar hacia la recta es el 48. La recta numérica posee huecos en los cuales el número puede ser ubicado. A partir del último ejercicio del nivel inicial en adelante, la recta numérica posee huecos donde el número se ubicará correctamente y donde no.

El objetivo de la actividad es llevar el número a través del laberinto hasta una posición de salida, aquellas donde termina el laberinto, y allí saltar hacia la recta.

Sobre la derecha de la Fig. 1 se encuentran los bloques que se utilizarán para mover el número a través del laberinto. Cada bloque posee una función específica que produce un efecto diferente sobre el movimiento que realizará el número. Dividimos las funciones de los bloques en tres categorías:

- Simples:
 - **Mover hacia arriba:** mueve el número hacia arriba una posición.
 - **Mover a la derecha:** mueve el número una posición hacia la derecha.
 - **Mover a la izquierda:** mueve el número una posición hacia la izquierda.
 - **Mover hacia abajo:** mueve el número hacia abajo una posición.
- Compuestos:
 - **Repetir N veces:** ejecuta N veces los bloques simples que estén dentro de él.
- Especiales de la Actividad:
 - **Saltar:** una vez que se llega a una posición de salida del laberinto, con este bloque se ubica el número en la recta numérica.

Los bloques, cualquiera sea su categoría, se deben ubicar dentro del contenedor de bloques, el cual se identifica por su color verde y su descripción "Poné las piezas acá adentro".

Al oprimir el botón “Ejecutar”, Códimo interpreta los bloques uno a uno, realizando el movimiento del número a través del laberinto. Terminada la ejecución, si el número finalizó dentro del hueco correcto de la recta numérica, se sugiere continuar con la siguiente actividad, en caso contrario se sugiere intentarlo nuevamente. Al reiniciar el ejercicio, el número volverá a su posición inicial y (probablemente) cambiará su valor dentro del rango de números posibles correspondientes al hueco que actualmente es el correcto. Esto así planteado permite que no sea un ejercicio sistemático, aun cuando la solución sea constante.

Los autores entienden que es importante destacar el hecho de que “Recta Numérica” es una actividad de matemática y no de programación. Al utilizar los bloques simulando ser sentencias en un lenguaje de programación, se está ejercitando uno de los aspectos del pensamiento computacional: el diseño de algoritmos; pero esto no forma parte del objetivo principal de la actividad que es afianzar el conocimiento sobre los números naturales, enteros y la recta numérica.

Todas las actividades que se pretenden generar con Códimo tendrán esa misma particularidad, lo que hace que puedan ser utilizadas por los docentes sin tener que desviarse de su currícula para introducir el Pensamiento Computacional en sus clases.

Al utilizar esta actividad, el alumno tiene que programar con los bloques para ejercitarse y afianzar este contenido específico de su currícula escolar. Por tal motivo decimos que no se está aprendiendo a programar, sino que se programa para aprender.

3 “Recta Numérica”: descripción técnica

Códimo es una aplicación web escrita únicamente en Javascript. Funciona en las últimas versiones de Firefox, Chrome, Edge y Safari. Además, como es de código abierto, no solo puede verse cómo está desarrollado, sino también es posible colaborar al informar sobre errores, proponer ideas, o incluso realizar modificaciones. El repositorio se encuentra en: <https://github.com/lgraziani2712/codimo>.

3.1 Tecnologías utilizadas

Para el desarrollo de Códimo fueron necesarias un conjunto variado de herramientas, tanto para facilitar la experiencia del desarrollo, como para asegurar un buen diseño de los componentes. A continuación se listan las más importantes y qué función cumplió cada una en la implementación.

Parte de la aplicación

- **Blockly**: “es una librería para diseñar editores visuales de programación” [4]. Permite implementar una interfaz de programación con bloques.
- **PixiJS**: “es un renderizador 2D para Javascript” [6]. Cumple la función de motor gráfico 2D, es decir, todas las animaciones del juego de la “Recta Numérica” están construidos sobre esta herramienta.

- **React**: “es una librería Javascript para construir interfaces gráficas” [3]. Se utilizó para poder crear la página web, donde fue posible desarrollar rutas dinámicas sin tener que construir un servidor. Permitió integrar PixiJS con Blockly.

En la Fig. 2 se puede apreciar de manera general cómo está diseñada la arquitectura de Códimo:

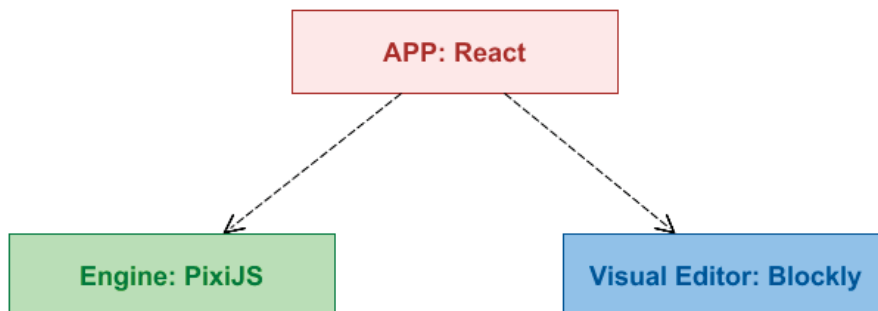


Fig. 2. Arquitectura general de Códimo

Como se evidencia en la Fig. 2, la arquitectura está integrada por tres componentes principales:

- **APP**: se encarga de manipular las URL de la aplicación, como la carga y eliminación de cada dependencia relativa a la actividad que se seleccionó.
- **Engine**: se encarga de renderizar la actividad y de animar cada objeto según la acción que recibe. Posee una interfaz que es utilizada por la APP para poder comunicarse.
- **Visual Editor**: se encarga de listar todas las posibles acciones, como también ofrecer un espacio de trabajo donde poder ir creando el programa. Ofrece una interfaz con la APP para poder comunicarse.

Cada vez que la APP intenta cargar una actividad, llama al *Engine* y le pide los elementos que necesita. Al mismo tiempo, llama al *Editor Visual*, y le pide que cree un editor con los elementos necesarios para la actividad en particular.

El usuario siempre interactúa con la APP y es ésta la que delega las tareas que recibe en los otros dos componentes.

Parte de la experiencia en el desarrollo

- **Storybook**: “es un entorno para desarrollo de Interfaces de Usuario” [8]. Con esta herramienta fue posible diseñar y desarrollar cada componente visual por separado. Para Códimo, un componente visual no sólo hace referencia al texto o los botones que acompañan a la actividad, sino el motor gráfico en sí. Así, fue posible diseñar, por ejemplo, cada animación del número de manera aislada.

- **Babel:** permite compilar código Javascript de las últimas versiones, a código que puede correr en las versiones más recientes de todos los exploradores. Esta tecnología nos permite confirmar que Códimo funciona, al menos, para las dos últimas versiones de Firefox, Chrome, Edge y Safari.
- **ESLint:** esta herramienta evalúa el código Javascript de manera estática en búsqueda de errores de sintaxis, de estructura y de estilo.
- **FlowType:** “es un analizador estático de tipos de datos para Javascript” [2].
- **Jest:** es un framework de testing para Javascript innovador en ciertos elementos.
Analiza qué archivos fueron modificados desde el último commit al repositorio (generalmente de tipo GIT), y ejecuta únicamente los tests relacionados a los mismos.
Durante la ejecución de los tests es posible filtrar por archivo, por nombre de test o por una expresión regular.
- **Snapshots:** este nuevo tipo de testing toma una variable y serializa el contenido en un archivo. Cada vez que se ejecuta el test, comprueba el contenido de la variable con lo serializado en el archivo, y si hay una diferencia, falla el test e informa la diferencia de la misma manera en que lo haría una herramienta de versionado de archivos. Esto es realmente útil cuando se quieren hacer tests unitarios de componentes HTML sin tener que hacer chequeos de cada elemento de la estructura.
- **Webpack:** es un module bundler, una herramienta que toma todos los archivos de un proyecto y los compila en único archivo. En particular Webpack puede considerar cualquier tipo de archivo (desde imágenes, json, de estilo, como css, entre otros) como un módulo de Javascript; siempre y cuando esté configurado para que sepa cómo manipular cada uno de ellos. Además, ofrece un servidor de desarrollo realmente sencillo de configurar donde es posible ir modificando el programa y ver los cambios en vivo sin tener que refrescar ninguna página.
La versión de producción de Códimo, con archivos que se descargan de manera asincrónica a medida que se requieren (por ejemplo, cada uno de los ejercicios), está generada a través de Webpack.

Parte del proceso de integración continua

- **CircleCI:** herramienta de integración continua donde es posible configurar eventos a partir de ejecutar distintas acciones cada vez que se hace un commit, un *pull request* o un merge al proyecto.
- **Netlify:** integra Códimo un proceso de deployment continuo.

4 Trabajo futuro

Los autores desean dar continuidad a este proceso de introducir el Pensamiento Computacional en las escuelas a través de las siguientes actividades:

- Completar todo el proceso de enseñanza, proveyendo un tutorial introductorio y un proceso de evaluación final. Aplicando FLOM [7] como arquitectura de diseño de actividades.
- Mejorar Códimo con la intención de brindar más y mejores funcionalidades junto a un conjunto de abstracciones que permitan a cualquier desarrollador implementar diferentes actividades reduciendo los tiempos de desarrollo.
- El Pensamiento Computacional abarca otros aspectos, no solo el diseño de algoritmos —el cual es ejercitado por medio de bloques—, como en la actividad descrita en este trabajo. El pensamiento abstracto, el reconocimiento de patrones, la descomposición de un problema complejo en múltiples más simples, son otros de sus aspectos relevantes. Es importante poder seguir generando actividades bajo el criterio de “programar para aprender”, que permitan evidenciar la utilidad de los diferentes aspectos del Pensamiento Computacional.

5 Conclusión

En este trabajo se presentó la actividad “Recta Numérica”, implementada con Códimo con el objetivo de poder encontrar evidencia sobre cómo la implementación del paradigma del Pensamiento Computacional a situaciones de aprendizaje específicas permite reforzar el conocimiento en sí a la vez que se desarrolla una manera de abordar problemas que será de utilidad para diversas situaciones. Introducir el Pensamiento Computacional en la educación primaria, brinda herramientas tanto al alumno como al docente. A los primeros porque permitirá despertar el interés de los estudiantes en el estudio de ciencias de la computación, a la vez que posibilitará que desarrollen estas habilidades desde temprano, para que puedan aplicarla en su vida cotidiana. Para los docentes es una forma entretenida de incorporar tecnología digital en el dictado de la materia, manteniendo el interés de los alumnos y brindando una forma novedosa y atractiva para afianzar las temáticas de cada asignatura.

6 Referencias

1. Códimo. (2017) ¡Donde el código, la imaginación y la motivación se encuentran! Recuperado de <https://codimo.netlify.com/>. Vigente junio 2017.
2. Facebook. (2017). Flow is a static type checker for javascript. Recuperado de <https://flow.org/>. Vigente junio 2017.
3. Facebook. (2017). React: A Javascript Library for Building User Interfaces. Recuperado de <https://facebook.github.io/react/>. Vigente junio 2017.
4. Google. (2017). A library for building visual programming editors. Recuperado de <https://developers.google.com/blockly/>. Vigente junio 2017.
5. Graziani, L. y otros (2016). Códimo: Desarrollo del pensamiento computacional en las escuelas. II Jornadas Argentinas De Tecnología, Innovación y Creatividad 2016 - Workshop sobre Innovación en Centros Educativos y de Investigación

- (WICEI). Recuperado de <http://jatic2016.ucaecemdp.edu.ar/trabajos/WICEI394TE-GrazianiJATIC2016.pdf>. Vigente junio 2017.
6. PixiJS. (2017). The HTML5 Creation Engine. Recuperado de <http://www.pixijs.com/>. Vigente junio 2017.
 7. Sanchez, A. J., Perez-Lezama, C. y Starostenko O. (2015). A Formal Specification for the Collaborative Development of Learning Objects. *Procedia - Social and Behavioral Sciences*. 182, 726–731. Recuperado de <http://www.sciencedirect.com/science/article/pii/S1877042815030955>. Vigente junio 2017.
 8. Storybook. (2017). Storybook: The UI Development Environment You'll ♥ to use. Recuperado de <https://storybook.js.org/>. Vigente junio 2017.
 9. Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, 49:33–35. [versión electrónica]. Recuperado de: <https://goo.gl/CHS4Yp>. Vigente junio 2017.

Revisión de herramientas para la creación de modelos 3D orientados a la enseñanza de la matemática con realidad aumentada

Tatiana Gibelli^{1,2}, Alfredo Graziani², Cecilia Sanz^{3,4},

¹Centro Interdisciplinario de Derechos Inclusión y Sociedad

²Sede Atlántica, Universidad Nacional de Río Negro

³III-LIDI Facultad de Informática Universidad Nacional de La Plata

⁴Investigador Asociado de la Comisión de Investigaciones Científicas de la Pcia. de Bs. As.
tgibelli@unrn.edu.ar, alfredograziani@msn.com, csanz@lidi.unlp.edu.ar

Resumen. Frente a las dificultades de los alumnos para la comprensión del espacio tridimensional, se propone el diseño de actividades con realidad aumentada, considerando que esta tecnología puede facilitar el desarrollo del pensamiento espacial. Uno de los requisitos en el diseño de actividades de este tipo es contar con los modelos 3D que se utilizarán. Si bien se pueden utilizar algunos de los modelos que se encuentran disponibles en bibliotecas digitales, en la realización de actividades específicas, surge la necesidad de diseñar los propios. Por ello se realizó una revisión de las herramientas de software, con especial énfasis en aquellas que permiten el trabajo con funciones matemáticas en 3D. En este artículo se presenta un estudio teórico y práctico de dichas herramientas, profundizando en dos de ellas: *Blender* y *SketchUp*. Se considera que este análisis comparativo puede resultar de interés para aquellos docentes que estén trabajando en temas de realidad aumentada.

Palabras clave: Modelos tridimensionales, Herramientas de modelado 3D, Realidad Aumentada.

1 Introducción

El aprendizaje de conceptos de geometría tridimensional suele presentar dificultades por parte de los estudiantes universitarios. En particular se observa esta situación en la Sede Atlántica de la Universidad Nacional de Río Negro (UNRN), en durante el cursado de las materias Matemática I, correspondiente al primer año de la carrera de Licenciatura en Sistemas e Ingeniería Agronómica y en Álgebra correspondiente al tercer año de la carrera de Contador Público. Se considera que esta problemática suele estar asociada a las dificultades para la visualización del espacio tridimensional, que permitiría una interpretación geométrica de dichos conceptos, facilitando la comprensión y el trabajo técnico-algebraico asociado [1]. De acuerdo a la teoría de Representaciones Semióticas, formulada por Raymond Duval, los objetos matemáticos no son accesibles físicamente, a través de evidencias sensoriales directas

o mediante el uso de instrumentos, la única forma de acceder y trabajar con ellos es a través de signos y representaciones semióticas [2]. Duval enfatiza en la importancia de que el alumno pueda representar y reconocer a un mismo objeto matemático con diferentes registros de representación semiótica. Para el aprendizaje de conceptos de geometría tridimensional, la integración de estos dos registros semióticos (algebraico y geométrico), permitiría una comprensión más acabada de los conceptos. Para la interpretación geométrica los estudiantes necesitan imaginar objetos tridimensionales en diferentes orientaciones, trasladar mentalmente la representación en el plano a tres dimensiones. Esta habilidad de pensamiento espacial es esencial para las carreras de Ingeniería y otras del ámbito artístico y científico, pues se utiliza para representar, manipular información en el aprendizaje y en la resolución de problemas [3].

Esta problemática fue abordada desde la investigación en el marco de un proyecto acreditado por la UNRN. El objetivo principal del mismo es investigar metodologías y estrategias innovadoras que favorezcan procesos educativos mediados por TIC, en particular aquellas consideradas emergentes como la Realidad Aumentada (RA). En este contexto se consideró el diseño de propuestas pedagógicas que incorporen el uso de la RA para facilitar el desarrollo de las habilidades de pensamiento espacial.

Para el diseño de actividades educativas con RA, uno de los aspectos a considerar, además de cuestiones pedagógico-didácticas, son las herramientas de software con las cuales se implementará la actividad. Para esta propuesta particular, se observa la necesidad de poder generar modelos 3D propios, acordes a la actividad planificada. Por ello, en este trabajo se presenta una revisión de las herramientas de software disponible para tal fin, con especial énfasis en aquellas que permiten el trabajo con funciones matemáticas en tres dimensiones, poniendo el foco en dos de ellas: *Blender* y *SketchUp*. Por un lado, se analizan las características de cada software, en relación al tipo de licencia, la plataforma de soporte, el nivel de usabilidad y los requerimientos de recursos, esto a partir de la documentación disponible. Por otra parte, se realiza un estudio práctico mediante una experiencia de modelado de funciones matemáticas tridimensionales con las diferentes herramientas de diseño 3D.

A partir de aquí, el artículo se organiza de la siguiente manera: la sección 2 presenta antecedentes teóricos sobre el uso de RA en educación y el software y hardware necesarios para su implementación. En la tercera sección se hace una revisión de las principales herramientas para el modelados 3D, profundizando el análisis teórico-práctico de dos de ellas: *Blender* y *SketchUp*. Finalmente, en sección 4 se presenta una discusión sobre cuál sería el software más apropiado según las necesidades específicas de la actividad de RA que se desea planificar.

2 Antecedentes teóricos

2.1 Realidad aumentada y usos en educación

Una de las tecnologías emergentes con posibilidad de inclusión en Educación es la Realidad Aumentada. Según Azuma et al. [4] la RA se caracteriza por: (a) una

combinación de objetos virtuales y reales en un escenario real, (b) usuarios interactuando en tiempo real y (c) una alineación entre los objetos reales y virtuales. Respecto a su potencialidad en el plano educativo, la RA permite incorporar multimedia a los procesos de enseñar y aprender, innovar en la práctica docente y promover el diseño de materiales educativos atendiendo a los requerimientos didácticos [5]. La RA puede proporcionar grandes oportunidades en distintas áreas como la ciencia o la ingeniería, puesto que estas disciplinas conllevan en su currículum un enfoque práctico en el aula [6]. Sin embargo, como señala Prendes Espinosa [7], aún siendo la tecnología lo suficientemente madura para ser útil en los entornos educativos, los nuevos desarrollos de RA llevarán a una nueva generación de aplicaciones que desplegarán mucho más su potencial.

Específicamente para la enseñanza de la matemática, Bujak et al. [8], analizan el uso de la RA desde tres perspectivas: física, cognitiva y contextual, presentando para cada una de ellas ejemplos de aplicaciones existentes de RA, así como pautas para el diseño de futuras experiencias, considerando cuestiones prácticas y tecnológicas. En la dimensión física destacan que las interacciones naturales con objetos físicos fomentan la comprensión de los conceptos en contexto. En la dimensión cognitiva mencionan que la alineación espacio-temporal de la información puede servir como andamiaje en la comprensión de los conceptos abstractos. Finalmente, en la dimensión contextual plantean las posibilidades que brinda la RA para el aprendizaje colaborativo en entornos no tradicionales, generando experiencias de aprendizaje significativas.

En cuanto al uso de RA para la comprensión del espacio tridimensional existen algunas experiencias. Se destaca la que desarrollan De la Torre et al. [3] donde proponen el uso de la RA y tabletas digitales multitáctiles para ofrecer la posibilidad de manipular un modelo digital en tres dimensiones de forma similar a como se hace con un modelo físico. Estos autores presentan resultados de una prueba piloto con 62 estudiantes de tres ámbitos educativos y muestran evidencias de que la realidad aumentada en tabletas multitáctiles es una alternativa válida para los objetivos educativos propuestos.

Sin embargo, además de lo planteado anteriormente, cuando se considera la planificación de una experiencia didáctica con integración de la tecnología de RA, es necesario plantearse la selección del software a utilizar, pues esa elección tendrá consecuencias metodológicas y pedagógicas.

2.2 Software y hardware para realidad aumentada

Las implementaciones de RA requieren de una combinación de software y hardware capaces de presentar una superposición de elementos virtuales en el mundo real. Todo sistema de RA ejecuta de manera secuencial las siguientes tareas [9]: 1) captura de la escena real observada por el usuario e identificación de la misma (recolección de información), 2) fusión de la realidad con objetos sintéticos y 3) visualización de escena aumentada (imagen compuesta ubicando objetos virtuales sobre la escena real). Para concretar cada una de estas etapas son indispensables elementos de

hardware, que pueden incluirse en un mismo dispositivo o en una combinación de varios: a) un dispositivo que capture las características del entorno físico y las transmita al software de realidad aumentada (cámaras web conectadas o integradas, y más específicos, sensores de geoposicionamiento como GPS, acelerómetros, giroscopios); b) un equipo capaz de sustentar el software para procesar dicha información y para enviar datos, que completen la realidad de lo que estamos viendo; y c) un dispositivo que plasme el contenido con la mezcla de la información real y virtual (una pantalla, una pizarra digital, unas gafas de realidad aumentada, un casco HMD, etc.).

En el caso específico del diseño y planificación de actividades educativas con RA, para el análisis y selección de software y hardware a utilizar, se deben tener en cuenta aspectos como el hardware disponible para el profesorado y el alumnado, las etapas del proceso de creación de RA que se quieren abarcar, el nivel de RA que se quiere utilizar, el coste del software necesario y el tipo de licencia (libre o propietario).

En el proceso de creación de RA, se pueden identificar las siguientes etapas, diferenciadas por el software y el hardware intervinientes en cada una [10]: 1) Creación del contenido virtual, 2) Programación de la experiencia de RA y 3) Reproducción del contenido de RA por el usuario final. Para este trabajo se pondrá el foco sobre la primera fase de creación del contenido virtual, principalmente en la indagación de herramientas de software disponibles para este propósito. En particular, el objetivo final es el diseño de una actividad con realidad aumentada para promover en los estudiantes la comprensión de conceptos de geometría tridimensional, por ello, esta primer fase se concentra en la creación de los modelos 3D que se utilizarán.

3 Software para modelado 3D

Existen muchos softwares que ofrecen la posibilidad de crear modelos 3D y animaciones. Mencionaremos a continuación algunos de los principales:

- *Adobe Photoshop CC*: es un potente software que permite modelar y animar en 3D. Es un software privativo, ya que requiere licencia para su uso. Está orientado a modelar, pues las herramientas de animación son escasas .
- *CAD*: toda la familia de software *CAD*, *TinkerCAD*, *FreeCAD*, *AutoCAD*, *OpenSCAD*, entre otros, constituyen herramientas de modelado que están orientados a la Ingeniería, Arquitectura o Mecánica. Permite crear modelos 3D y animaciones, con un enfoque relacionado a la construcción, por ejemplo, planos 3D de una casa o edificio, animación de un motor de un vehículo, etc. Con excepción de algunos como *FreeCAD*, los softwares de esta familia requieren licencia para su uso.
- *3ds Max*: *3DS Studio Max* es un potente software de modelado que cuenta con todo tipo de herramientas para hacer modelos muy profesionales. Requiere experiencia en la materia de modelos y animación, ya que no es sencillo su uso. Empresas como *Blizzard Entertainment* hacen uso de este

software para crear y animar modelos para juegos y películas animadas. Requiere una licencia y una cuota mensual para su uso.

- *Wings3D*: es un software libre de modelado 3D. Tiene una falta de herramientas para animación, lo que no lo hace la mejor opción para animar modelos. Tiene la ventaja que es gratuito y multiplataforma.

Frente a la necesidad específica de modelar funciones matemáticas tridimensionales, muchos de estos softwares, no cuentan con la posibilidad de crear los modelos a partir de una ecuación matemática. Si es posible el modelado a mano, lo que ya requiere conocimiento más profesional de modelado 3D. Por ello el análisis que aquí se presenta, se concentra en dos softwares que sí permiten esta posibilidad: *Blender* y *SketchUp*.

3.1 Análisis teórico y práctico de *Blender*

Blender es un software especializado en el modelado, creación y animación de gráficos tridimensionales. Actualmente, es un software gratuito de código abierto que puede ejecutarse en múltiples plataformas como *Windows*, *Linux*, *MacOS*, entre otros.

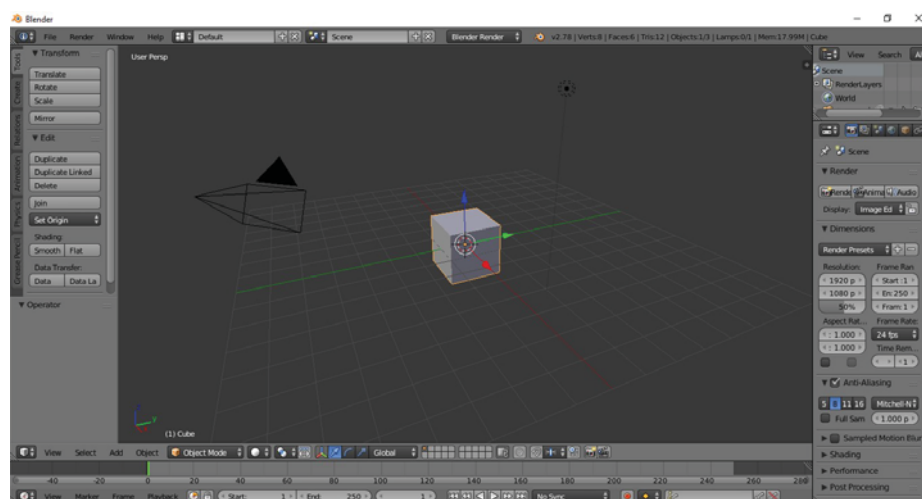


Fig. 1. Interfaz de *Blender* como se muestra apenas se instala

La interfaz de *Blender* (ver Figura 1) es poco intuitiva y resulta algo complicada de entender si se está iniciando con este software. Sin embargo, investigando un poco los menús del software se pueden descubrir algunas funcionalidades básicas. La interfaz de *Blender* está en varios idiomas, incluido el español. Para realizar el cambio de idioma, se debe ir a *File*, luego *User preferences*, pestaña “*System*”. En la parte inferior, menú “*International Fonts*”, se puede elegir el idioma deseado, indicando a quién afecta este cambio: “*Interface*”, “*Tooltips*”, “*New Data*”.

Este software cuenta con tutoriales *online*, oficialmente presentados por la empresa desarrolladora del software, que están en idioma inglés. También ofrecen soporte en

línea (foro) donde se brindan soluciones. Los usuarios tienen acceso a libros sobre el uso de *Blender* (muchos están a la venta, pero hay otros de libre acceso), que lo pueden ayudar a introducirse al programa, o bien, una vez familiarizado, llegar a un nivel más profesional.

Blender ofrece la posibilidad de animar los modelos creados, utilizando una línea de tiempo compuesta por fotogramas, donde en cada uno es posible decidir en qué estado y posición se desea ubicar el objeto. En el menú inferior se puede ver la línea de fotogramas, donde se puede modificar la cantidad de fotogramas para la animación. Para grabar la posición y estado del modelo en un fotograma dado, se necesita seleccionar el fotograma y acceder al menú *Animación* (tecla rápida *I*). En el mismo, se debe insertar una *Clave*, y elegir la opción deseada, como rotar, mover, etc.. Para hacer que rote, por ejemplo, si se ubica una posición inicial en el fotograma 1, y una posición final en un fotograma alejado, el 250 por ejemplo, la animación se encarga de ir rotando el objeto desde la posición inicial a la final automáticamente, no siendo necesario grabar la posición en cada fotograma. Este ejemplo se presenta para dar cuenta de ciertas facilidades que ofrece esta herramienta.

Una vez que se tiene el modelo deseado, *Blender* ofrece la posibilidad de exportarlo en distintos formatos. El predefinido es *Collada (.dae)*. Otros formatos son: *Alembic (.abc)*, *3D Studio (.3ds)*, *FBX (.fbx)*, *Stanford (.ply)*, *Wavefront (.obj)*, *X3D Extensible 3D (.x3d)*, *Stl (.stl)*.

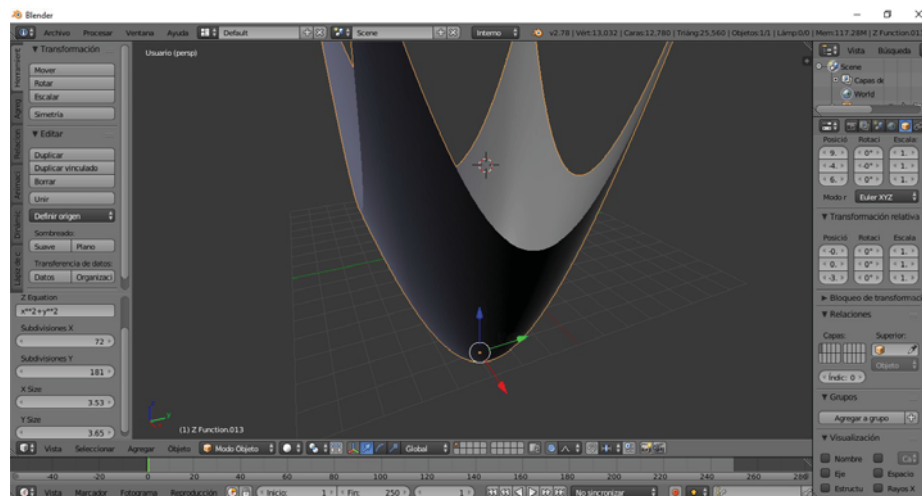


Fig. 2. Modelo gráfico de la función $z=x^2+y^2$ usando software *Blender*

Blender cuenta con complementos (o *plugins*), los cuales son aplicaciones desarrolladas por usuarios, que le agregan al software una funcionalidad específica. En el caso de esta investigación, se usa el complemento *3D Function Surface*¹ que permite incorporar una función dada por ecuación matemática. Una vez instalado el *plugin* (que en las últimas versiones viene por defecto), para modelar los gráficos de

1 https://wiki.blender.org/index.php/Extensions:2.6/Py/Scripts/Add_Mesh/Add_3d_Function_Surface

funciones en 3D, se debe ir al menú inferior de la interfaz, a la opción “Agregar”, y allí se debe seleccionar el submenú “Mallas”, luego “Math Function” y, por último, “Z Math Surface”. Realizada esta secuencia, se muestra un gráfico predefinido cuya ecuación puede verse en la sección inferior izquierda de la pantalla, junto a otras opciones como las del rango del gráfico. Todas estas opciones se pueden editar para obtener el modelo deseado (ver ejemplo en figura 2).

3.2 Análisis teórico y práctico de SketchUp

SketchUp es un software de modelado 3D privativo, el cual ofrece la posibilidad de crear y animar modelos 3D. Este software no es gratuito y su código no está disponible, por lo que se requiere una licencia para utilizarlo. Se ofrece un período de prueba de 30 días, en el cual se puede acceder a todo el contenido del software sin necesidad de pagar. Una vez terminado este período, el software se bloqueará hasta que se adquiera una licencia que otorga acceso a todo el contenido del software, a nuevas actualizaciones y soporte técnico. Las plataformas soportadas por *SketchUp* son *Windows* y *MacOS*. Actualmente no existe una versión para *Linux*.

Cuenta con una amplia gama de tutoriales, libros y ayuda *online*. La licencia del programa ofrece soporte técnico especializado, y hay foros donde consultar a la comunidad por soluciones. Para los principiantes, *SketchUp* desarrolló unos videos introductorios al programa para familiarizarse con el mismo. Estos videos están disponibles al público en general, y están en inglés.

La interfaz de este programa (ver Figura 3) es muy amigable, puede descargarse en idioma español. Cuenta con un menú superior con las herramientas para utilizar, la pantalla de modelado, y un menú lateral derecho que se divide en secciones, una de las cuales es un instructor, que explica y muestra gráficamente el uso de la herramienta que está actualmente seleccionada, haciendo fácil el aprendizaje de ellas.

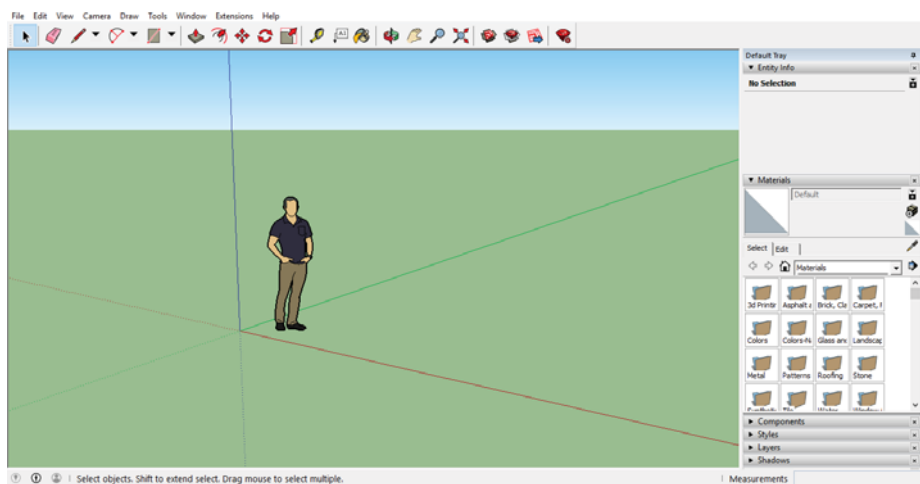


Fig. 3. Interfaz de SketchUp

Este software permite la posibilidad de animación básica donde es posible recorrer el modelo moviendo las cámaras. En el menú *View*, submenú *Animation*, se puede crear una escena con *Add Scene*. La vista del modelo actual es la que se va a guardar en la escena, por lo que si hay 2 escenas vistas de distintos ángulos, al reproducir la animación, el pase de la escena 1 a la 2 va a mover la cámara de un punto a otro recorriendo así el modelo.

SketchUp permite exportar los modelos creados en distintos formatos. El formato por defecto es *COLLADA (.dae)*. Los otros formatos de exportación son: *3DS (.3ds)*, *AutoCAD DWG (.dwg)*, *AutoCAD DXF (.dxf)*, *FBX (.fbx)*, *IFC (.ifc)*, *Google Earth (.kmz)*, *OBJ (.obj)*, *VRML (.wrl)*, *XSI (.xsi)*.

Al igual que *Blender*, *SketchUp* también cuenta con *plugins*, muchos de los cuales están desarrollados por usuarios que no tienen relación con la empresa propietaria del software. Para el modelado de funciones, se encontró un *plugin* específico que permite dada una ecuación del tipo $z=f(x,y)$ crear el modelo 3D del gráfico de la función, llamado *k_tools*². Su funcionamiento es sencillo, una vez descargado, el archivo se debe ubicar en la carpeta *plugins* de *SketchUp* y al reiniciar el programa aparece un nuevo menú en la barra superior llamado *Extensions*. Ahí aparecerá el *plugin* recién instalado y se puede acceder a sus funciones. Para crear un gráfico 3D, se debe acceder a la opción *Graphs*, seguido de *3D-Graph z=f(x,y)*. Seleccionada esta opción, se debe ingresar en una ventana el rango de las variables x e y , y la ecuación de la función matemática. Por último se debe seleccionar un estilo; y así finalmente puede verse el modelo generado (ver ejemplo en figura 4).

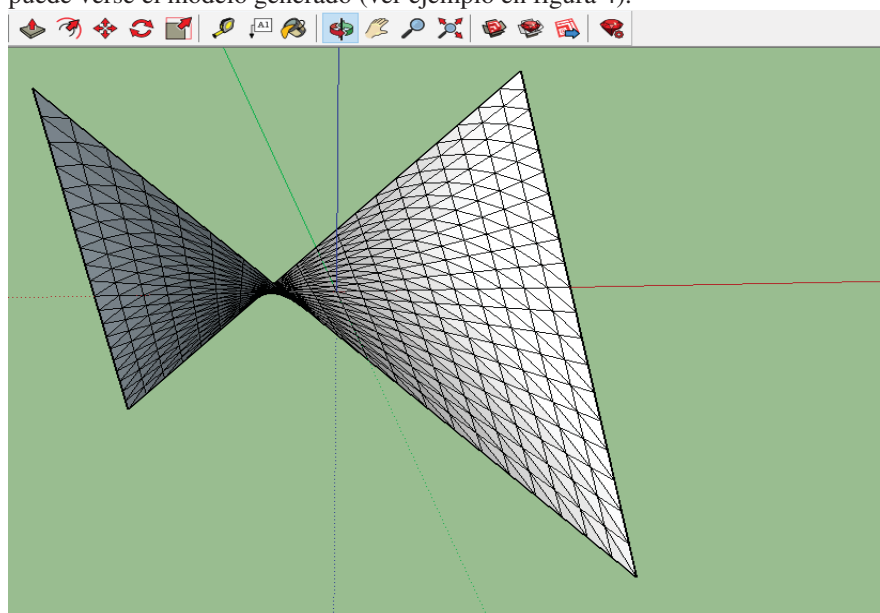


Fig. 4. Modelo gráfico de la función $z=x*y$ usando software *SketchUp*

² http://www.crai.archi.fr/rld/plugin_details.php?id=210

4 Discusión

Se presenta a continuación una comparación entre los dos softwares explorados para la necesidad específica del modelado de funciones matemáticas tridimensionales: *Blender* y *SketchUp*.

Luego del análisis teórico-práctico presentado (en tabla 1 se muestra un resumen de las características exploradas) es posible concluir que *Blender* es una de las mejores opciones para realizar la tarea de modelado de gráficos de funciones 3D. La principal razón es la posibilidad de hacer uso del software de forma libre y que es apto para varias plataformas de sistemas operativos. Si bien inicialmente puede mostrarse un poco difícil su utilización, acostumbrarse a la interfaz es más sencillo de lo que puede parecer en un principio. Cuenta con mucha ayuda *online* por parte de la comunidad y desarrolladores, lo que ayuda a un mejor y más eficiente aprendizaje. Por otra parte, *Blender* cuenta actualmente con el módulo para modelar gráficos de funciones 3D, no hay que buscar *plugin* de terceros para hacerlo. La animación de *Blender* es más potente que otros softwares, es posible ver cómo se está animando el modelo, teniendo una línea de tiempo y configurando qué pasa en cada fotografía.

SketchUp también puede llegar a ser una buena opción, el principal inconveniente es la licencia, la cual puede resultar costosa. Pero una vez adquirido el software, junto con el *plugin* de modelado de gráficos 3D, es un poco más sencillo que en *Blender* llegar al modelo 3D deseado y cuenta con mucho soporte especializado por parte de los desarrolladores, tales como videos online, guías, tutoriales y ejemplos que ayudan a un aprendizaje rápido para un uso básico del software. Sin embargo, resulta mucho más complejo que en *Blender* realizar animaciones en sus modelos, y requiere de más experiencia en el uso del software y conocimiento general de animación.

Tabla 1. Resumen de principales características de los softwares *Blender* y *SketchUp*.

Característica	Análisis de software Blender	Análisis de software SketchUp
Licencia	Gratuita. Código abierto	Privativa. Requiere licencia
Formatos de exportación	3D: Alembic, .3DS, .DAE, Filmbox, Autodesk, Wavefront, DirectX, Lightwave, Motion Capture, SVG, STL, VRML	3D: .3ds, .dwg, .dxf, .dae, .fbx, .ifc, .kmz, .obj, .wrl, .xsi
Plataforma	Windows (7, Vista, 8, 10), Linux, Ubuntu, Fedora, SunOS, Arquitectura 32 y 64 bits	Windows (10, 8+, 7+), MacOS 10.10+
Nivel de usabilidad	Intermedio. La interfaz no es muy intuitiva, y no está en español.	Sencillo. Interfaz básica, cuenta con idioma español.
Requerimiento de recursos	Mínimo: Intel Core i3, Tarjeta gráfica con 1024MB de memoria dedicada.	Mínimo: Procesador 2+ GHz, 8+ GB RAM, Tarjeta gráfica con 1GB de memoria y soporte aceleración de hardware.
Modelado de funciones 3D	Es difícil dar con la herramienta adecuada, una vez encontrada, crear los gráficos es sencillo.	El plugin necesario para graficar es difícil de encontrar. Una vez instalado, graficar es muy básico.

5 Conclusiones

En este trabajo se ha presentado un análisis de diferentes herramienta para la creación de modelos 3D. En particular, se han revisado en profundidad dos herramientas: *Blender* y *SketchUp*. Uno de los ejes de análisis se ha puesto en las posibilidades para generar modelos a partir de las ecuaciones matemáticas, requisito referido, acorde a la motivación de este trabajo, vinculada a la creación de actividades educativas basadas en RA para la enseñanza de la matemática, y más específicamente la geometría tridimensional. La experiencia práctica realizada con ambas herramientas, así como el análisis teórico, constituye un aporte para docentes que se encuentren en este camino. Esta investigación continuará a partir de la implementación de las actividades de RA utilizando *Blender* para el modelado 3D. Estas actividades están diseñadas y se planifica implementarlas en la segunda parte del año.

Agradecimientos. Este trabajo fue financiado por el proyecto de investigación “La mediación de las tecnologías de la información y la comunicación en procesos educativos. Innovaciones para mejorar los procesos de enseñanza y aprendizaje” (PI-UNRN-40-C-486), acreditado por la Universidad Nacional de Río Negro. También se hace un agradecimiento al proyecto REFORTICCA financiado por la Comisión de Investigaciones Científicas de la Prov. de Buenos Aires, a partir del cual se han tomado referencias de estudios sobre Realidad Aumentada.

Referencias

1. Del Río, L. S., Búcarí, N. D. y Sanz, C. V. (2016). Uso de recursos hipermediales para la enseñanza y el aprendizaje de la matemática. In *II Congreso Internacional de Enseñanza de las Ciencias y la Matemática. Tandil, 6 al 9 de septiembre de 2016*.
2. Duval, R. (2006). Un tema crucial en la educación matemática: La habilidad para cambiar el registro de representación semiótica. *La gaceta de la RSME*, 9(1), 143-168.
3. De la Torre Cantero, J., Martín-Dorta, N., Pérez, J. L. S., Carrera, C. C., & González, M. C. (2015). Entorno de aprendizaje ubicuo con realidad aumentada y tabletas para estimular la comprensión del espacio tridimensional. *Revista de Educación a Distancia*, (37).
4. Azuma, R., Bailiot, Y., Behringer, R., Feiner, S., Julier, S., & MacIntyre, B. (2001). Recent advances in augmented reality. *Computer Graphics and Applications, IEEE*, 21(6), 34-47.
5. Avendaño, V. y Domínguez, L. A. (2012). Realidad aumentada: Una exploración al escenario de la virtualidad educativa. Madrid: Editorial Académica Española.
6. Andújar, J. M.; Mejías, A.; Márquez, M. A. (2011). Augmented Reality for the Improvement of Remote Laboratories: An Augmented Remote Laboratory. *IEEE Transactions on Education*, 54 (3), (492-500)
7. Prendes Espinosa, C. (2015). Realidad aumentada y educación: análisis de experiencias prácticas. *Pixel-Bit. Revista de Medios y Educación*, (46), 187-203.
8. Bujak, K. R., Radu, I., Catrambone, R., Macintyre, B., Zheng, R., & Golubski, G. (2013). A psychological perspective on augmented reality in the mathematics classroom. *Computers & Education*, 68, 536-544.
9. Mazen Abdulmuslih, A. (2012). “Análisis de sistemas de realidad aumentada y metodología para el desarrollo de aplicaciones educativas.”. Universidad Rey Juan Carlos.
10. Pajares Ortega, E. P. (2015). Diseño de actividades didácticas con Realidad Aumentada. Trabajo Final. Master de Educación y Comunicación en la Red. UNED.

Selector de actividades de AulasWeb: la utilización de personajes virtuales para acompañar las decisiones de los docentes

Alejandro Héctor González¹, Cristina Madoz¹, Leandro Matías Romanut²

¹Instituto de Investigación en Informática III – LIDI – Facultad de Informática UNLP

²Dirección de Educación a Distancia y Tecnologías UNLP

{agonzalez,cmadoz}@lidi.info.unlp.edu.ar

leandro.romanut@presi.unlp.edu.ar

Abstract. Este trabajo tiene sus orígenes en una propuesta de investigación de una tesina de grado de la Facultad de Informática de la UNLP. El producto se está utilizando actualmente en el entorno virtual de enseñanza y aprendizaje AulasWeb, basado en Moodle. Se utiliza un personaje virtual en el selector de actividades y recursos como complemento para orientar a los docentes en la inclusión de actividades a las aulas virtuales. Así también se muestran los resultados logrados a través de la realización de las experiencias en los cursos y las nuevas modificaciones agregadas al personaje.

Keywords: entornos virtuales de enseñanza y aprendizaje, trabajo colaborativo, personajes virtuales, b-learning

1 Introducción

En la Dirección de Educación a Distancia y Tecnologías (EADyT) dependiente de la Secretaría de Asuntos académicos de la Universidad Nacional de la Plata (UNLP) se realizan, desde el año 2004, diferentes capacitaciones dirigidas a docentes en general y en particular a profesores universitarios que quieran incorporar los conceptos de educación a distancia y uso de tecnologías digitales a sus prácticas de enseñanza. En diferentes ediciones se observa que los docentes poseen poca experiencia en el uso de entornos virtuales, y al momento de la creación y diseño de las aulas, se encuentran con variados recursos y actividades para utilizar sin tener en claro el objetivo pedagógico con el cual incorporarlos.

Se analizaron diferentes entornos virtuales de enseñanza y aprendizaje: Moodle (AulasWeb¹ y Qoodle²), Sakai³, IDEAS⁴, utilizados en distintas universidades, revisando cómo están organizados a nivel sistema y en particular qué herramientas proveen para realizar la creación y diseño de las aulas. Se profundizó en el estudio de las herramientas colaborativas provistas por estos entornos y se analizó que

¹<https://aulasweb.ead.unlp.edu.ar/aulasweb/>

²<https://acceso.uvq.edu.ar>

³<https://sakaiproject.org/>

⁴<https://ideas.info.unlp.edu.ar/login>

herramientas dispone cada uno, si son de trabajo individual o grupal y que nivel de ayuda presentan cada una de ellas. Revisando la asistencia brindada en cada caso para las herramientas y, tomando las mismas como base, se propuso generar para el entornos AulasWeb, utilizado en la dirección de EADyT, un conjunto de instructivos que incluyan posibles situaciones pedagógicas para utilizar las actividades de trabajo colaborativo. Estos instructivos se crearon basados en encuestas realizadas a profesionales en el uso y coordinación de aulas virtuales y en un análisis de uso de herramientas de trabajo colaborativo en las aulas virtuales del entornos AulasWeb. Los instructivos están acompañados de un personaje virtual que colabora como guía en la creación y diseño de las aulas virtuales.

2 Marco teórico

Los procesos educativos han incorporado un variado conjunto de tecnologías digitales que permiten pensar y repensar las propuestas de enseñanza. Entre ellas podemos mencionar los denominados “entornos virtuales de enseñanza y aprendizaje”, como Moodle, Sakai, y algunos con software propietario como WebCT o Blackboard; entre otros.

Las funcionalidades de los entornos virtuales han evolucionado a lo largo de los años y desde la llegada de la Web 2.0 se han incrementado las posibilidades de realizar propuestas educativas colaborativas en línea.

La incorporación de las tecnologías de la información y comunicación (TIC) en la educación, produce una serie de cambios y transformaciones en las formas en que se representan y llevan a cabo los procesos educativos en general.

Desde una visión constructivista, explica Barriga, “el aula”, lejos de referirse a un conjunto de recursos físicos, consiste en un sistema interactivo en el cual ocurren una serie de transacciones comunicativas. Este sistema genera un ambiente particular de trabajo para la construcción del aprendizaje, determinado por una serie de reglas de organización y participación, este ambiente se denomina “ambiente de aprendizaje”. [1]

2.1 Modalidades educativas

Las modalidades educativas están relacionadas con el ambiente en el que se desarrollan los procesos de enseñanza y aprendizaje. Definen la utilización de los medios, de los recursos educativos y establece las acciones de los agentes del proceso -estudiantes, tutores, coordinadores-. [2]

Bates explica que la modalidad educativa puede cambiar de acuerdo a la cantidad de e-learning que interviene en el proceso. En la modalidad presencial, los estudiantes comparten un mismo espacio físico y tiempo, la formación es sincrónica y en el mismo lugar y es la más evolucionada dada la historia misma de la educación.

“El e-learning es un término conveniente para cubrir una amplia gama de usos de la tecnología en el aprendizaje y la enseñanza. Todas las actividades de la computadora y basadas en internet que apoyan en la enseñanza y el aprendizaje tanto “on-campus” como en la distancia. Esto incluye usos administrativos, así como

académicos de las tecnologías de información y comunicación que apoyan el aprendizaje.”[3]

A medida que incrementa el uso de e-learning en las modalidades de enseñanza aparecen dos términos: el b-learning y el aprendizaje distribuido.

González define al b-learning (Aprendizaje híbrido o combinado) como la modalidad que combina la enseñanza presencial con la no presencial de tal manera que ambas experiencias de aprendizaje se vuelven imprescindibles para completar con éxito los objetivos de aprendizaje. [4][5]

Al involucrar de manera total el e-learning, se hace referencia a la educación a distancia, dado que involucra diferentes formas de comunicación.

Algunos autores definen a la educación a distancia como:

“La enseñanza a distancia es un sistema tecnológico de comunicación bidireccional (multidireccional), que puede ser masivo, basado en la acción sistemática y conjunta de recursos didácticos y el apoyo de una organización y tutoría, que, separados físicamente de los estudiantes, propician en éstos un aprendizaje independiente (cooperativo).” [6]

“La educación a distancia es un aprendizaje planificado que ocurre normalmente en un lugar diferente al de la enseñanza, por lo tanto requiere de técnicas especiales de diseño de cursos, de instrucción, de comunicación, ya sea por medios electrónicos u otro tipo de tecnología, así como de una organización especial.” [7]

Por su lado, Mena, Diez y Rodríguez señalan a la educación a distancia como una modalidad educativa que, mediatizando la mayor parte del tiempo la relación pedagógica entre quienes enseñan y quienes aprenden a través de distintos medios y estrategias, permite establecer una particular forma de presencia institucional más allá de su tradicional cobertura geográfica y poblacional, ayudando a superar problemas de tiempo y espacio.[8]

De las definiciones anteriores puede pensarse que la educación a distancia en la actualidad es un proceso educativo que mantiene una relación pedagógica entre los docentes y los estudiantes a través de distintos recursos y estrategias, siendo este proceso mediado por tecnologías. Los entornos de aprendizaje y enseñanza apoyan la comunicación y el intercambio.

2.2 Aprendizaje Colaborativo

El aprendizaje colaborativo es un proceso que se caracteriza porque los alumnos son quienes diseñan su estructura de interacciones y mantienen el control sobre las diferentes decisiones que repercuten en su aprendizaje.

“El aprendizaje colaborativo es la adquisición de destrezas y actitudes que ocurren como resultado de la interacción en grupo.”[9]

Gros (2000) agrega que en un proceso colaborativo, las partes se comprometen a aprender algo juntos. Lo que debe ser aprendido sólo puede conseguirse si el trabajo del grupo es realizado en colaboración. Es el grupo el que decide cómo realizar la tarea, qué procedimientos adoptar, cómo dividir el trabajo, las tareas a realizar. La comunicación y la negociación son claves en este proceso. [10]

2.3 Escritura Colaborativa

Storch explica que las posibilidades del trabajo en grupo o entre pares para escribir se limitaban o al proceso de generación de ideas (brainstorming) o al proceso final de revisión (peer review). Destaca que algunas investigaciones mostraron que el proceso de revisión entre pares, generaba beneficios para los participantes relativos a su capacidad de considerar las posiciones de los otros; y en el desarrollo de habilidades tanto para la escritura como para la lectura analítica y crítica. Sin embargo, la autora plantea también que la tendencia a potenciar y estudiar la escritura colaborativa sólo como proceso de revisión entre pares supone una deficiencia en la aproximación dado que se centra en el producto y no considera el proceso de escritura. [11]

En el proceso para la escritura colaborativa, los estudiantes deberían participar a lo largo de todo el desarrollo de la escritura donde los participantes deberían compartir la responsabilidad en la producción de un texto tanto a nivel de estructura, de contenido y de lenguaje. Las ventajas de una posición de coautoría, más que de revisión entre pares, se centran en aspectos como los siguientes:

- Se favorece el pensamiento reflexivo (sobre todo si los participantes ponen en marcha mecanismos para defender o explicar mejor sus ideas);
- Los participantes pueden rebasar los niveles centrados en la ortografía o la gramática para abordar cuestiones relativas al discurso; por último,
- Puede favorecer el conocimiento que los participantes desarrollan sobre el lenguaje.

3 Entornos Virtuales de Enseñanza y Aprendizaje

Un LMS (Learning Management System - Sistema de Gestión de Aprendizaje) es un software online que permite administrar, distribuir, monitorear, evaluar y apoyar las diferentes actividades previamente diseñadas y programadas dentro de un proceso de formación completamente virtual (e-learning) o de formación semi-presencial (b-learning), explica Cañellas Mayor. [12]

3.1 Análisis de los Entornos Virtuales de Enseñanza y Aprendizaje

Algunos de los entornos de enseñanza y aprendizaje que se utilizan para abordar el estudio de las actividades de trabajo colaborativo son desarrollos de comunidades de software libre, como Moodle y Sakai, y otros son espacios que están a cargo de alguna entidad educativa: Qoodle (Universidad Nacional de Quilmes), AulasWeb (Presidencia - Universidad Nacional de La Plata) e IDEAS (Facultad de Informática - Universidad Nacional de La Plata).

Para cada uno de los entornos elegidos se analizó: cómo surgieron, cómo están organizados a nivel sistema, qué roles de usuario poseen, qué herramientas proveen para realizar la creación y diseño de las aulas, si poseen herramientas externas además de las que vienen de manera nativa.

Luego se realizó un análisis de las herramientas de trabajo colaborativo que provee cada entorno.

Algunos ejemplos: la actividad Base de Datos es propia de Moodle, pero no está presente en Sakai, ni en IDEAS. El foro es una actividad provista por todos los entornos pero con distintas características. En Moodle existen distintos tipos: Uso general, Preguntas y respuestas, Cada persona plantea un tema, Formato de blog y Debate sencillo. En Sakai no hay tipos, los aportes pueden ser moderados antes de que sean visibles, se necesita un tema para inaugurar el foro, de otra manera, no será visualizado. La actividad Wiki está disponible en Moodle y en Sakai, en el primero permite crear Wikis de manera colaborativa o individual y soporta varios formatos (HTML, Creole, NWiki) y en el segundo no posee ninguna característica destacable, por su parte, no está como posibilidad en Ideas.

3.2 Trabajo Colaborativo y Entornos Virtuales de Enseñanza y Aprendizaje

Cuando el concepto de aprendizaje colaborativo se extiende en el entorno virtual, el concepto permanece, pero las condiciones, y por ello las posibilidades, cambian sustancialmente. El entorno virtual brinda herramientas para que los docentes puedan proponer y diseñar actividades innovadoras para la colaboración, la comunicación y la producción de conocimientos y aumenta las posibilidades para poder aprender y trabajar en equipo a las cuales se veía limitada hasta ahora en un entorno de trabajo presencial. [13]

Barriga y Morales realizan una selección de herramientas digitales que permiten el aprendizaje colaborativo en entornos virtuales: [14]

- Chat
- Blog
- Wiki
- Foro
- Salas de trabajo (breakoutrooms)
- Pizarrón de mensajes
- Conferencias en línea
- Pizarra compartida
- Correo electrónico

4 Justificación y contexto de desarrollo

La tesina se enmarca dentro el proyecto de investigación: “*Tecnologías para sistemas de software distribuidos. Calidad en sistemas y procesos. Escenarios educativos mediados por TIC*”, del Instituto de Investigación en Informática III-LIDI de la Universidad de La Plata. Se tiene en cuenta para el desarrollo la experiencia en capacitación de profesores llevadas a cabo en la Dirección de Educación a Distancia y Tecnologías de la UNLP. Se observa desde hace tiempo la falta de inclusión de actividades orientadas al trabajo colaborativo en las propuestas de los cursos a través

de las herramientas provistas por el entorno AulasWeb (perteneciente a la dirección y basado en Moodle).

Para poder profundizar sobre las herramientas de trabajo colaborativo en los entornos virtuales de enseñanza y aprendizaje, se llevó a cabo una encuesta de carácter anónimo compuesta por preguntas abiertas para poder brindar mayor libertad de expresión. Estas encuestas se realizaron a los docentes y coordinadores de aulas virtuales de la Universidad Nacional de La Plata, con el fin de dar a conocer su trabajo con entornos desde una perspectiva no solo de usabilidad, sino de su experiencia como docentes (interactuando con alumnos y materiales) y acompañando diferentes propuestas educativas de distintos niveles (pregrado, grado, postgrado, extensión). De las encuestas realizadas (54 en total) se observa que no todos utilizan herramientas de trabajo colaborativo y, generalmente, las más utilizadas son los foros y las tareas.

Las encuestas realizadas a los docentes colaboraron para el diseño y desarrollo de breves instructivos que identifiquen algunas situaciones pedagógicas y didácticas donde se puedan utilizar las actividades de trabajo colaborativo.

Para poder revisar las respuestas de los encuestados se realizó un análisis de la presencia de herramientas colaborativas en los cursos del entorno AulasWeb. De las encuestas realizadas a docentes y coordinadores de aulas y el análisis del uso de las actividades colaborativas en los cursos, se puede suponer que a los profesores se le presentan problemas al momento de utilizar las herramientas de trabajo colaborativo:

- Algunas de las actividades de trabajo colaborativo poseen cierta complejidad en su configuración lo que provoca que los docentes se abstengan de usarlas.
- Los docentes habitualmente no intentan dinamizar las actividades del curso incluyendo las actividades colaborativas. Generalmente trabajan con texto (recursos de tipo archivos en AulasWeb) convirtiendo al aula en un sitio “estático”, donde los alumnos se descargan el documento y trabajan con la actividad. Esto pasa sin que haya un real motivo, tanto en la modalidad de extensión del aula como, en las propuestas a distancia.
- Los docentes están dispuestos a trabajar con las actividades colaborativas, pero no rediseñan ni reformulan las prácticas educativas, no adecuándolas para usar tales actividades.

5 Herramienta desarrollada

Con el fin de tener una herramienta flexible y que pueda ser incorporada fácilmente en cualquier entorno Moodle, se optó por realizar una modificación al código que permita incorporar un personaje virtual que colabore en la selección de actividades y recursos de Moodle.

Se tomó como base el selector de actividades de Moodle. En la interfaz de este selector, el docente puede ver el listado de todas las actividades y recursos que posee Moodle para agregar a las aulas. Cuando selecciona una de las actividades o recursos, se muestra sobre la misma interfaz una descripción de la herramienta. A partir de este selector, se extienden tres funcionalidades:

- Organizar el selector de actividades en tres secciones:
 - **Actividades de Trabajo Colaborativo** (Base de Datos, Chat, Foro, Glosario, Taller, Tarea, Wiki, y otras herramientas externas como VPL y BigBlueButton).
 - **Otras Actividades** (Consulta, Cuestionario, Herramienta Externa, Lección, Paquete SCORM).
 - **Recursos** (Archivo, Carpeta, Etiqueta, Libro, Página, URL).
- Ofrecer para cada actividad de trabajo colaborativo una ayuda pedagógica sobre en qué situaciones educativas podrían emplear dicho recurso.
- Ofrecer un asistente que permite reproducir la ayuda a través de un audio y ver un ejemplo de uso de la actividad seleccionada.

Con el fin de seguir trabajando en la herramienta, se extendieron los instructivos pedagógicos a cada actividad (Consulta, Cuestionario, Herramienta Externa, Lección, Paquete SCORM) y recurso (Archivo, Carpeta, Etiqueta, Libro, Página, URL) del entorno AulasWeb, basado en Moodle. Así también, se incluyó el asistente virtual con sus funcionalidades ya desarrolladas: reproducir el instructivo y mostrar un ejemplo de la actividad/recurso seleccionada.

5.1 Personaje virtual

Se creó un personaje virtual a través del sitio “Pocoyó” mediante su aplicación Pocoyize⁵, la cual permite crear avatares de caricaturas y descargarlos de manera gratuita.

La incorporación de un personaje intenta generar una estrategia que oriente a los docentes en la comprensión de las actividades de trabajo colaborativo.

En esta oportunidad se decide trabajar con un personaje que cumple el rol de tutor tecnológico, que es aquel que realiza el acompañamiento durante el aprendizaje de los alumnos y colabora en el entendimiento y uso de las tecnologías involucradas en el aula virtual.

La creación final del personaje se define a partir de un conjunto de “rasgos personales”, así como también de un escenario, una historia y uno o más objetivos. Según Rib Davis, los ingredientes para el personaje son los que resultan de uno mismo y que hacen que cada uno sea diferente de los demás. Si bien en la obra de Rib se describe el armado de los personajes para teatro, cine y literatura, los elementos de creación resultan adaptables a la creación de personajes virtuales. [15][16]

Para lograr un personaje se deben tener en cuenta tres aspectos:

1. ¿Cómo es el personaje cuando nace (por su genética y su entorno)?.
2. ¿Cómo es el personaje por lo que va aprendiendo y llega a ser a través de la experiencia?
3. ¿Cómo es el personaje ahora?

⁵<http://www.pocoyo.com/pocoyizador>



Fig. 1. Manu, el tutor tecnológico

6 Primera experiencia

Se instaló y utilizó la modificación desarrollada al selector de actividades; en el servidor donde reside el entorno AulasWeb (basado en Moodle) de la Dirección de Educación a Distancia y Tecnologías de la UNLP.

Se realizó una prueba piloto del asistente en el curso “Introducción al uso de los entornos virtuales de la UNLP soportados en Moodle (Aulas Web – Cavila – Cursos externos)”. El objetivo del curso busca que los profesores se apropien de conceptos básicos para poder implementar sus cursos y se acerquen a la edición básica de las herramientas de las aulas virtuales del entorno Moodle. El curso se dictó bajo la modalidad b-learning con la duración de 5 semanas con 2 encuentros presenciales, en los cuales se trabaja en forma de taller utilizando las herramientas de Moodle, en un lapso de dos horas y media cada uno. El curso se dictó durante el 2015. Asistieron 18 docentes de las distintas unidades académicas de la UNLP.

El curso propone realizar la creación y el diseño de un aula virtual tomando como referencia la propuesta educativa de alguna cátedra a la que pertenezcan los participantes. Los estudiantes tuvieron que leer la documentación y las guías de ayuda desarrolladas para las actividades de trabajo colaborativo, así como también utilizar el asistente para ver algunos ejemplos de uso de las herramientas.

Se realizó una encuesta, a través de un formulario online creado con la herramienta Google Drive, con el objetivo final de indagar el nivel de aceptación del prototipo en el entorno AulasWeb y extraer, algunas sugerencias y mejoras que puedan realizarse sobre el desarrollo realizado para este trabajo. Se pueden mencionar dos de las preguntas realizadas:

La primera pregunta *¿Utilizó los instructivos disponibles como ayuda para diseñar las prácticas educativas que involucren las actividades de Trabajo Colaborativo?*, Todos los encuestados respondieron que sí. Esto indica que recurrieron a la ayuda que ofrece la herramienta para planificar y comprender el uso desde una perspectiva más pedagógica y no sólo tecnológica.

La segunda pregunta *¿Cuál es su opinión acerca de los instructivos?*, basada en la respuesta afirmativa de la pregunta anterior, arrojó varias opiniones, las cuales fueron procesadas y agrupadas en porcentajes, con los comentarios textuales de los docentes:

- El 50% dijo: *“Son útiles y claros”*.
- El 25% dijo: *“Son de gran ayuda, sobretudo la primera vez que uno comienza a trabajar con la plataforma.”*
- El 12% dijo: *“La información de los instructivos es muy buena, da una idea general de cada herramienta y ejemplos dónde utilizarla.”*

- El 13% dijo: *“Son una buena herramienta de apoyo para los que todavía no estamos del todo insertos en el sistema.”*

7 Conclusiones

Se realizó el estudio de diferentes entornos virtuales de enseñanza y aprendizaje. Se estudiaron los distintos LMS usados tanto en la UNLP como en otras universidades, como es el caso de Sakai y Qoodle. El estudio comparativo de las herramientas en general y las de trabajo colaborativo en particular permitieron armar los instructivos adecuados para el personaje, tomando las fortalezas y debilidades de las indicaciones de uso de cada una.

Se desarrolló una herramienta que colabora con la creación, diseño, uso e inclusión de las actividades colaborativas en las propuestas educativas que los docentes llevan adelante con sus cursos.

La primera experiencia con la herramienta permite ver buenos resultados, tanto en la comprensión de las actividades de trabajo colaborativo desde una perspectiva pedagógica, así como también desde un punto de vista tecnológico. Los docentes se manifestaron a favor de tener en contexto la ayuda, y que se encuentre con indicaciones pedagógicas.

Los ejemplos que provee el asistente son analizados como útiles y claros, esto aproxima a la suposición que se tenía sobre el tipo de información que ofrece el manual de ayuda de Moodle. Permiten comprender e interpretar el texto de ayuda y utilizar los ejemplos para el diseño de las propuestas educativas.

La inclusión de un asistente virtual involucrando la creación de un guión para el mismo, logra recorrer un proceso creativo que permite incorporar diversos elementos ligados a la multimedia que van desde la presentación del personaje, su historia y su forma de ser hasta los procedimientos para la asistencia en el uso de las actividades colaborativas. La utilización del canal de audio fue bien recibido por los docentes y es un primer intento de mejorar la accesibilidad a personas con capacidades diferentes.

Se realizó un aporte tecnológico y pedagógico a la Dirección de Educación a Distancia y Tecnologías de la UNLP, incorporando el asistente en los LMS que se utilizan actualmente: AulasWeb, CAVILA - UNLP y Cursos Externos, todos basados en Moodle.

8 Trabajo futuro

Se proponen las siguientes mejoras a futuro para aplicar sobre la herramienta desarrollada:

- Realizar nuevas encuestas para ver el impacto en la comunidad docente y obtener retroalimentación para mejorar las ayudas y sugerencias didácticas para cada herramienta.
- Agregar nuevas funciones para personalizar el asistente en aspectos de imagen de acuerdo a como lo prefiera el docente del curso: cambiar la ropa, el color de piel, el color y corte de pelo, los ojos, la boca, el cuerpo, la voz.

- Sumar nuevas funcionalidades al asistente Manu:
 - Proveer un espacio de preguntas frecuentes con respuestas automatizadas.
 - Sugerir un listado de herramientas externas que se puedan incluir (embeber) dentro del entorno AulasWeb. Este listado constaría de una breve descripción, un tutorial de uso y el enlace a la página web de la aplicación. Por ejemplo: construir presentaciones online (Prezi), diseñar videos interactivos (Moovly), confeccionar collages (Fotor - BeFunky), organizar líneas de tiempo (Capzles), elaborar infografías (RAW), armar nubes de etiquetas (Tagxedo - Word it Out), desarrollar murales colaborativos (Padlet), construir imágenes interactivas (Thinglink), etc.

9 Bibliografía

1. Barriga, F. D. Educación y nuevas tecnologías de la información y la comunicación: ¿Hacia un paradigma educativo innovador? (2008)
2. Jiménez, D. Nuevos paradigmas educativos y modalidades educativas. Recuperado de: http://es.slideshare.net/auri_desi/modalidades-educativas-5551315 (2010)
3. Bates, A. T. Technology, e-learning and distance education. Routledge. (2005)
4. González, A. H. Educación a Distancia y Tecnologías Digitales en la Enseñanza Universitaria. Recuperado de https://prezi.com/v-vkyghu8zsc/2015-especializacion-en-docencia-universitaria-unlp/?utm_campaign=share&utm_medium=copy (2015)
5. González, C. Cuadro comparativo: Blended learning, Distributed Learning, Online Learning. Recuperado de <http://es.calameo.com/read/001790350da562655eacd> (2012)
6. García Aretio, L. La Educación a Distancia. De la teoría a la práctica. Ariel editorial S. A. (2001)
7. Moore, M. & Kearsley, G. Distance education: A systems view. Boston, MA: Wadsworth Publishing. (1996)
8. Mena, M., Diez, L. M., y Rodríguez, L. El Diseño de proyectos de Educación a distancia. Colección Itinerarios. Ediciones La Crujía y Editorial Stella. Buenos Aires. (2005)
9. Salinas, J. El aprendizaje colaborativo con los nuevos canales de comunicación, 199 – 227. En Cabero, J. (Editor). Nuevas tecnologías aplicadas a la educación. Madrid: Síntesis. (2000)
10. Gros, B. El Aprendizaje Colaborativo a través de la Red: Límites y Posibilidades. En Aula de innovación educativa, ISSN 1131-995X, N° 162, 2007, págs. 44-50. (2000)
11. Storch, N. Collaborative writing: Products, process, and students reflections. Journal of Second Language Writing, 14, (153-173). (2005)
12. Cañellas Mayor, A. LMS y LCMS: funcionalidades y beneficios. Recuperado de <http://www.centrocp.com/lms-y-lcms-funcionalidades-y-beneficios/> (2014)
13. Harasim, L., Hiltz, S. R., Turoff, M. y Teles, L. Redes de aprendizaje. Guía para la enseñanza y el aprendizaje en red. Barcelona: Gedisa. (2000)
14. Barriga, F. D. y Morales R. Aprendizaje colaborativo en entornos virtuales: un modelo de diseño instruccional para la formación profesional continua. En Revista de Tecnología y Comunicación Educativas. México, ILCE, Año 22-23, No. 47-48, pp. 4-25. (2009)
15. Rib, D. Escribir guiones: desarrollo de personajes. Editorial Paidós, manuales de escritura. Barcelona. España. (2004)
16. González, A. H. TICs en el proceso de articulación entre la Escuela Media y la Universidad. Tesis de Maestría, Facultad de Informática, UNLP. (2008)

SocialSchool: una plataforma web para instituciones educativas

Franco Brusatti¹, Sonia Permigiani¹, María Marta Novaira¹, Rodrigo Ariño²,
Eduardo Depetris², and Saul Coria²

¹ Dpto. de Computación, FCEFQyN, Universidad Nac. de Río Cuarto, Argentina

² PlumLabs (www.plum.com.ar)

{fbrusatti, spermigiani, mnovaira}@dc.exa.unrc.edu.ar

{saul, eduardo, rodrigo}@plum.com.ar

Resumen En la actualidad la tecnología está presente en muchas de las tareas cotidianas, la actividad escolar de nuestros hijos no es ajena a esta exposición pero hemos identificado algunos puntos donde es posible profundizar el uso de las tecnologías para que la experiencia escolar sea aún más eficiente. Con este propósito hemos creado SocialSchool[2] una aplicación que intenta acercar todos los actores involucrados en una institución educativa tutores, directivos, docentes, preceptores y alumnos. En este paper vamos a introducir las principales funcionalidades de la aplicación, hablaremos de las experiencias aprendidas en su implementación e instanciación en algunos colegios de la zona y concluimos con un análisis sobre lo observado en las implantaciones.

Keywords: Programación, Herramientas Tecnológicas, Aplicación Web

1. Introducción

La educación es una columna importante de la sociedad y necesita adaptarse a los cambios de generación de la misma.

Argentina siendo un país de tercer mundo logró acercar la tecnología a los alumnos, mediante asignaturas orientadas y entrega de materiales necesarios (plan del gobierno “netbook para todos”). De esta manera el alumno tiene la posibilidad de interactuar con nuevas fuentes de estudios, investigar, experimentar y realizar tareas a partir de las TICs disponibles hasta el momento.

Sin embargo, la mayoría de los establecimientos educativos no acompañaron el ritmo de este avance y siguen utilizando las mismas metodologías de administración (sistema de notas, planillas, control de asistencia, método de avisos a los padres, ficha médica, etc.) desde hace 30 años. De esta manera se está perdiendo información importante, que si bien está físicamente dentro del establecimiento, sigue siendo de poca utilidad estando aislada, de difícil manipulación y acceso.

Otro eje de esta problemática, la constituyen los padres. Según estadísticas de la OIT (Organización Internacional del Trabajo)[1], la cantidad de horas laborales promedio viene en aumento desde el año 2000. De acuerdo a las cifras de este relevamiento, el promedio de algunos rubros, como el comercio, ya supera

las 46 horas semanales, mientras que uno de cada tres hombres trabaja 48 horas o más, lo que es considerado exceso de trabajo. Por estas razones, entre otras, es común escuchar que los padres se enteran tardíamente del rendimiento de sus hijos, las inasistencias y mucho más tarde de las sanciones disciplinarias. Nunca se enteran a tiempo de lo que pasa en el colegio de sus hijos: actos, reuniones, actividades, etc. La falta de control debido a las restricciones de tiempo disponible más la poca conexión entre establecimiento y tutores es un problema común en el sistema educativo.

La NO digitalización de la información en las escuelas, es un factor que no se está priorizando. Por ejemplo, contar con información actualizada en tiempo real podría alertar a los involucrados antes de que sea demasiado tarde. Una herramienta que solucione este atraso podría mejorar la calidad educativa.

SocialSchool[2] es una plataforma web que pretende resolver los problemas mencionados anteriormente. Provee una forma para comunicar a los tutores, estudiantes, profesionales y personal administrativo de una manera innovadora utilizando tecnologías accesibles y agiliza la gestión administrativa de los establecimientos.

Cada institución en donde se implantó el sistema tenía diferentes necesidades, la implementación de las soluciones para una institución también tomaron valor para otras y esto provocó que la plataforma crezca en robustez.

A través de la experiencia en el uso de la plataforma pudimos observar una significativa mejora en el tiempo insumido en las actividades de gestión de las instituciones, otro punto donde la plataforma ha mostrado interesantes desempeños es en la notificación a los tutores, para algunas instituciones esto se ha convertido en un alivio.

A lo largo de este documento vamos a describir las dificultades y aprendizajes obtenidos tanto en la implementación como en la implantación de la plataforma.

En la sección dos se describe el contexto y las problemáticas de la educación y se plantea a la herramienta como una solución a las mismas, en la sección siguiente, se detalla la implementación de la plataforma: las herramientas utilizadas y algunas de las funcionalidades que posee. En la cuarta sección se muestran las conclusiones. Finalmente, en la quinta sección, se describe los trabajos futuros y lecciones aprendidas.

2. La Institución educativa y su contexto

Al ser la educación una necesidad fundamental de implementación de todo estado por ser un derecho humano fundamental, reconocido por la ONU como tal, las necesidades asociadas a la misma son siempre crecientes ya que están relacionadas al crecimiento de la población, que se encuentra siempre en aumento. Por otro lado, el sector evidencia un fuerte atraso en la implementación de TICs que potencien y optimicen sus actividades, a diferencia de otros segmentos que ya las han adoptado, por lo que presenta una fuerte necesidad de actualización.

Esta realidad es más marcada y evidente en países con menor grado de desarrollo socio-económico, mientras que en países más desarrollados la implementa-

ción de las TICs en educación tiene desde hace años mayor grado de penetración. Es así que se plantea como solución a esta problemática este proyecto, que intenta resolver cómo aplicar TICs en procesos de gestión y organización en el sector educativo, con foco en entidades educativas del segmento primario, secundario y terciario.

SocialSchool tiene tres objetivos principales sobre las instituciones:

1. Ayudar a la institución a digitalizar la información y reducir los tiempos de ejecución de las tareas diarias y toma de decisiones.
2. Reducir la distancia de los tutores con las actividades de sus hijos: La comunicación entre los tutores, alumnos y la institución se realiza en tiempo real por medio del software.
3. Administrar cuotas, pagos y reducir la morosidad por medio de diferentes mecanismos de avisos a los tutores.

3. SocialScool: herramienta tecnológica para instituciones educativas

SocialSchool es una plataforma web de gestión educativa, su nombre tiene origen en la idea que la escuela tiene que estar estrechamente vinculada con las tecnologías de comunicación social que hoy día gobiernan la atención de nuestra sociedad.

Los usuarios del sistema son el personal de la institución, los alumnos y los tutores de los mismos. Cada uno se conecta con la aplicación por medio de un usuario y una contraseña. Cada uno tiene restringido sus acciones dentro de la plataforma de acuerdo al perfil que posee. Por ejemplo un profesor puede cargar notas y asistencias, un alumno puede consultar sus notas y asistencias. Un tutor puede ver las notas y asistencia de su hijo/a únicamente; y el estado de las cuotas.

La seguridad es un punto en donde se hizo mucho incapié en la implementación de la plataforma debido a la información sensible que se manipula, es por esto que se han seguido altos estándares de seguridad en su implementación, en particular se han seguido los estándares de seguridad sugeridos para certificar PCI-DSS[3]. También se firma un contrato de protección de datos con cada institución que se trabaja.

El sistema está concebido para cubrir en su versión final todos los aspectos de la gestión y administración de la entidad educativa.

3.1. Implementación

El equipo de desarrollo involucrado en la implementación del sistema es un equipo experimentado en desarrollo web, se utilizó una metodología de desarrollo llamada SCRUM[5][6] para el desarrollo de la plataforma, esta metodología permite tener rápidos resultados en corto tiempo, esto permitió una rápida retroalimentación con los colegios que estaban dispuestos a hacer una prueba temprana del sistema.

La plataforma fue desarrollada en un framework para desarrollo web llamado Ruby on Rails [7][8]. Y hace un uso intensivo de algunas librerías javascript para que la experiencia de usuario sea placentera y veloz.

La plataforma puede y ha sido desplegada en la nube usando los servicios de Amazon S3 y EC2[9]. Esto le permite a los colegios prescindir de un servidor en el establecimiento y acceder a la plataforma a través de internet.

3.2. Funcionalidades

En esta sección hablaremos de las principales funcionalidades que posee la plataforma. Cabe mencionar que los siguientes son sólo algunos de los módulos que la plataforma ofrece, por supuesto existen todos los módulos para gestionar alumnos, tutores, personal y demás entidades en el sistema.

- **Asistencias.** Los preceptores o profesores pueden tomar asistencia utilizando un celular o una tablet. Se puede observar que en este caso la digitalización de la información es inmediata, la plataforma contabiliza el total de faltas justificadas y no justificadas y puede enviar avisos a los tutores de manera automática e inmediata. Estas notificaciones pueden ser configuradas por cada institución. Al tener digitalizada las asistencias permite obtener rápidamente información sobre los alumnos en riesgo de abandono y actuar en consecuencia.
- **Notas.** Existen diferentes tipos de notas para diferentes cursos. Las notas pueden ser de tipo cuantitativa, cualitativa o narrativa. Los jardines de infantes, por ejemplo, hacen uso de las calificaciones narrativas para hacer el seguimiento de los niños. Las notas pueden ser consultadas en cualquier momento en la libreta digital por las personas autorizadas a hacerlo.
- **Libreta.** Cuenta con un sistema de representación digital de la libreta del alumno. Estas libretas son configurables, se elaboran automáticamente y se pueden imprimir. El sistema de calificaciones se personaliza para adaptarse a las necesidades de cada institución.
- **Cuotas.** Permite la administración de cuotas. Sus principales funcionalidades son:
 - Generación de cuotas.
 - Cálculo de intereses por mora.
 - Notificaciones a los tutores.
 - Administración de moroso.
 - Integración con entes de cobro (RapiPago, Pagos Mis Cuentas, etc.).
- **Notificaciones.** En un principio creíamos que las notificaciones era algo que toda institución desearía pero hemos encontrado establecimientos que se han encontrado ante la problemática de familias en proceso de divorcios y

las notificaciones a ambos padres no siempre es bienvenida en uno de ellos. Es por eso que todas las notificaciones del sistema son configurables y cada institución decide qué enviarle a los tutores y qué no. Este módulo es transversal en la aplicación y fue un desafío interesante su implementación. Se puede notificar inasistencias, mora de cuotas, comunicados de actividades en las instituciones, sanciones disciplinarias, etc. También la notificación es efectiva a través de sms, email o whatsapp.

- **Estadísticas.** Es uno de los módulos que mayor valor tiene para los directivos de los colegios, dado que les permite conocer la situación exacta en la que se encuentra la institución en cada momento. Se puede obtener la información de los alumnos en riesgo de abandono debido a la cantidad de inasistencias. Se puede conocer el rendimiento de cada alumno y/o del curso; como así también el rendimiento de una materia determinada de un curso. Se puede determinar las materias previas de un alumno, de un curso determinado o de la institución. También nos han sugerido reportes particulares que son requeridos por los ministerios de educación provinciales (San Luis y Córdoba). Estos reportes han sido relativamente fáciles de generar dado que al tener toda la información de la institución digitalizada y centralizada es simple poder representarla a través de reportes.

También la plataforma cuenta con un equipo de asistencia al usuario que está disponible en los horarios que los colegios trabajan por si les surge alguna duda o problema con la plataforma.

4. Conclusión

Como hemos mencionado cada institución ha representado un desafío diferente y de cada una se aprendió y se adaptó la plataforma para su uso.

La centralización y digitalización de la información en los establecimientos ha mostrado considerables beneficios a la hora de generar reportes y ha reducido los tiempos de la gestión administrativa de manera satisfactoria.

Por el lado de los padres, se han presentado deseos de disponer una herramienta de este tipo que los conecte más con la escuela y poder estar al tanto de lo que le pasa a sus hijos. Los padres impulsan este tipo de soluciones. La incorporación de tutores al proceso de comunicación y seguimiento de sus hijos ha favorecido el contacto de la familia con la institución, y en los lugares donde se utiliza parece tener buena aceptación.

La morosidad también se vio reducida en las instituciones donde se implementó la plataforma y esto se le atribuye al sistema de notificaciones temprana de avisos y a la amplia gama de opciones que se ofrecen a la hora de abonar la cuota.

Otros beneficios que la herramienta aporta son, entre otros, ahorro en papel contribuyendo a la reducción de las impresiones ya que la información puede

verse online; la implementación es inmediata, no depende de la instalación de un software; manejo simple e intuitivo de la aplicación (requiere poca capacitación para los usuarios).

Uno de los puntos a resaltar sobre la herramienta es el de la **seguridad** ya que cuenta con los niveles de seguridad más altos que existen actualmente en el mercado.

5. Trabajos a futuro

La problemática del bullying tiene a toda la comunidad muy sensibilizada, por lo que creemos que una solución en esta dirección podría ser muy beneficiosa para las instituciones que adopten la plataforma. Actualmente estamos investigando qué herramientas existen en este campo y cómo poder adaptarlas en la plataforma para poder atacar este problema.

Además, se pretende incorporar a la herramienta el legajo de cada integrante de la comunidad educativa: directivos, docentes, secretarios, preceptores, bibliotecarios, etc. De esta forma se mantiene actualizada la información, de fácil y ágil acceso. Otra información importante para incorporar a la herramienta son los proyectos institucionales e interdisciplinarios que la escuela lleva adelante. Esto permitirá que todos los integrantes de la institución tengan la posibilidad de acceder a los mismos.

5.1. Lecciones aprendidas

El proyecto inicialmente fue solventado con un Fondo Fiduciario de Promoción de la Industria del Software - FONSOFT[4], si bien el planteo de la aplicación estaba claro, aplicar TICs a la educación, nos comenzamos a dar cuenta después de algunas reuniones con directivos de colegios educativos de la zona que las necesidades que tenían eran más básicas de las que habíamos previsto inicialmente, tenían problemas de gestión en vez de calidad educativa.

En un principio creímos que la notificación a los padres era algo deseable para todas las instituciones, luego nos encontramos que no siempre es el caso.

Los módulos en la herramienta debieron ser adaptados según las exigencias y necesidades de cada institución en diferentes provincias.

Referencias

1. <http://www.ilo.org/global/lang--es/index.htm>
2. <http://socialschool.com.ar>
3. <https://www.pcisecuritystandards.org/>
4. <http://www.agencia.mincyt.gob.ar/frontend/agencia/fondo/fonsoft>
5. [https://en.wikipedia.org/wiki/Scrum_\(software_development\)](https://en.wikipedia.org/wiki/Scrum_(software_development))
6. "Essential Scrum: A Practical Guide to the Most Popular Agile Process", Kenneth S. Rubin, Addison-Wesley, 2008.
7. <http://rubyonrails.org/>

8. "Ruby on Rails Tutorial: Learn Web Development with Rails", Michael Hartl, Addison-Wesley, 2015.
9. <https://aws.amazon.com/>

Una Propuesta de Evaluación de Herramientas CASE para la Enseñanza

Nicolás Battaglia, Roxana Martínez, Carlos Neil, Marcelo De Vincenzi
Facultad de Tecnología Informática - Universidad Abierta Interamericana
{nicolas.battaglia, roxana.martinez, carlos.neil, medevincenzi}@uai.edu.ar

Abstract. Este trabajo presenta un marco teórico de las Herramientas CASE y UML en base al sustento de análisis realizado por varios autores y el propio del equipo. Por lo que, en base a dicho estudio, se propone un método de evaluación para conocer en qué grado una herramienta CASE se adapta con la funcionalidad requerida. Además, se sugieren criterios de evaluación a tener en cuenta con su importancia correspondiente para cada uno de los casos. Presentamos en base a esto, las características deseables de una herramienta CASE destinada a la enseñanza de la Ingeniería de Software.

Keywords: Herramientas CASE, UML, Evaluación, Ingeniería del Software.

1 Introducción

Actualmente en Argentina, el sector de software ofrece un potencial considerable en términos de crecimiento económico para las empresas de software. Este sector puede ser visto mediante dos perspectivas, por un lado, como avance a futuro sobre la modernización tecnológica en el país y, por otro lado, como un sector capaz de desarrollar procesos de innovación y aprendizaje. Partiendo de este último concepto, en el desarrollo de sistemas de información, es notorio el interés constante por mejorar la productividad y la calidad en los productos desarrollados. El avance en el software, impacta directamente en los costos de desarrollo y mantenimiento, por lo que es vital, establecer un marco de trabajo disciplinado y estructurado el cual permita identificar en forma clara y puntual, las tareas a realizar por un equipo de trabajo, siendo un punto fundamental para evitar resultados inesperados.

La inclusión de las herramientas CASE (*Computer Aided Software Engineering*), están tomando cada vez más relevancia en la planeación y ejecución de proyectos que involucran sistemas informáticos, debido a que permiten aumentar la productividad en el desarrollo de software. Estas herramientas, suelen inducir a sus usuarios a la correcta utilización de metodologías que implican una reducción tanto en el costo del proyecto, como en el tiempo de desarrollo del producto de software final. Para esto, es importante considerar que tipo de herramienta CASE es necesaria al momento de llevar adelante una actividad relacionada con la Ingeniería de Software (IS), por lo que es importante conocer y analizar las distintas taxonomías y la usabilidad relacionada en cada fase del ciclo de vida del desarrollo de software a construir.

Si bien, la importancia de la utilización de las herramientas CASE se relaciona con el mercado laboral, también debe ser considerada en el sector educativo, porque

impacta directamente en los alumnos que serán los que se encuentren a futuro, insertados en dicho mercado laboral. Por lo que, es necesario conocer y estudiar las diferentes limitaciones, ventajas y desventajas de utilizar herramientas CASE en el proceso de enseñanza en una carrera de ingeniería en sistemas, donde es fundamental hacer hincapié en el concepto de modelado, durante el análisis, diseño e implementación.

1.1 Herramientas CASE

A lo largo de los años, la introducción de nuevos componentes y circuitos integrados en la IS, permitió que varias aplicaciones comenzaran a ser propuestas válidas y a materializarse en productos mucho más amplios y complejos, dejando como experiencia que el enfoque informal en la construcción de software no era bueno, y que se generaban productos más caros de lo presupuestado en los proyectos, los cuales eran difíciles de mantener, con desempeños muy pobres y de mala calidad.

Los costos de software fueron aumentando considerablemente, mientras que los de hardware iban por el camino opuesto, por lo que surgió la necesidad de nuevas técnicas y métodos que permita controlar la creciente complejidad y así también, controlar los costos [1]. Fue notoria la gran demanda de calidad en el desarrollo de software en el mercado laboral, por lo que se buscaba una manera confiable y económica de administrar los recursos y así cumplir con la entrega del producto de software en tiempo y forma. Es por ello que, surge la necesidad de obtener productos más fáciles de mantener, con más capacidad de adaptación y con costos más accesibles, y, en consecuencia, comenzó a hacerse necesario automatizar las diferentes actividades que propone la IS. En base a esto, IEEE (*Institute of Electrical and Electronics Engineers*) desarrollo diferentes estándares y una definición completa, la cual apunta a la aplicación de un enfoque sistemático, disciplinado y cuantificable en el desarrollo y mantenimiento del software.

Pressman asegura que “las herramientas CASE ayudan a garantizar que la calidad se diseñe, antes de llegar a construir el producto” [2]. Y propone entender la IS como una tecnología con diferentes capas, apoyándose en el compromiso con la calidad. Luego el proceso, como base de control; los métodos, que brindan la experiencia necesaria para elaborar software por medio de un conjunto de tareas (como el análisis, diseño, implementación y pruebas). Por otro lado, las primitivas herramientas CASE, se dirigieron principalmente a la automatización de la documentación y la comunicación como una mejora clave de la productividad del software [3].

1.2 Características adicionales

La utilización de modelos para crear software utilizando lenguajes, técnicas y herramientas dentro de la IS es una práctica común durante las etapas de análisis y diseño. Esto conlleva al desarrollo de un producto con calidad de Ingeniería para la resolución de un problema [4].

Todas las herramientas CASE prestan soporte a un lenguaje de modelado para acompañar la metodología que, en su gran mayoría, soportan UML (*Unified Modeling*

Language) el cual permite realizar una notación orientada a objetos, mediante diferentes diagramas, de los cuales, se representan las distintas etapas del desarrollo de un proyecto. UML es de amplia aceptación en el ambiente de IS, ya que su valor conceptual y visual, proporciona facilidad para representar elementos particulares de aplicaciones a desarrollar. Si una herramienta CASE soporta UML, la funcionalidad de intercambio de modelos con otras herramientas CASE se convierte en una meta alcanzable, teniendo en cuenta la estandarización sintáctica y semántica que persigue dicho lenguaje. En esta línea de estandarización la OMG (*Object Management Group*) que es un consorcio dedicado al cuidado y el establecimiento de diversos estándares de tecnologías orientadas a objetos, como ser: UML, definió un estándar de intercambio llamado XMI (*Metadata Interchanger*), que está basado en el meta modelo de UML, el cual sirve para el intercambio de metadatos utilizando XML.

Existen distintos objetivos de las herramientas CASE, de acuerdo al trabajo de McClure [3] algunos de los objetivos principales de las herramientas CASE son: Aumentar la productividad en el desarrollo; Dar calidad a los productos desarrollados; Reducir el costo del software; Automatizar los chequeos de errores; Acelerar el desarrollo de las aplicaciones; entre otros.

Cabe destacar que la carencia de la enseñanza de la IS como característica es recurrente, por lo que, en un entorno académico, los alumnos deben utilizar herramientas que estén pensadas para utilizar en el ámbito profesional y no sólo disponer de aquellas funcionalidades requeridas en el ámbito académico [5] [6][7].

1.3 Modelado UML

Un modelo es una representación de algo en un medio determinado. Los modelos capturan los aspectos de lo que se desea modelar con un punto de vista en particular y provee un nivel de abstracción sobre los aspectos menos importantes [8].

Un modelo posee dos aspectos importantes, información semántica y presentación visual por un lado y el contexto por otro. Los aspectos semánticos capturan el significado de las construcciones que son descriptos de forma separada, pero están relacionadas como parte de un modelo único y coherente. La presentación visual muestra la información semántica para que pueda ser usada por humanos. El contexto incluye información interna para que los modelos no carezcan de sentido. Es decir, que un modelo tiene un sentido distinto según el contexto en el que se utilice.

Un modelo de software se utiliza para diversos propósitos, entre ellos destacamos los más importantes: a) Para que un equipo interdisciplinario pueda entenderse; b) Para capturar y precisar requerimientos funcionales; c) Para Diseñar un sistema; d) Para crear productos de trabajo; y e) Para organizar grandes sistemas.

UML surge de la necesidad de tratar de entender, diseñar, navegar, configurar y mantener toda la información relativa a los sistemas de software. No está pensado como un método completo de desarrollo ni un marco metodológico de trabajo, sino como una herramienta que, en concurrencia con una metodología, brinde soporte al proceso de desarrollo en cada una de sus etapas. Por esto es que, los autores del estándar UML describen en sus libros que UML se crea con el objetivo de unificar experiencias anteriores sobre técnicas de modelado y así, incorporar las mejores

prácticas en un único estándar. Hoy por hoy, UML es soportado por la mayoría de las herramientas comerciales que se utilizan para modelar software [9].

2 Propuesta de una Herramienta CASE para Modelado

Toda herramienta CASE de Modelado, debería soportar algún lenguaje de modelado estándar que permita simplificar tanto la comunicación como la documentación del sistema objeto de estudio [10]. Es por lo que las tendencias metodológicas de los últimos años están altamente influenciadas por el enfoque de Orientación a Objetos (OO), permitiendo identificar las abstracciones necesarias para proponer una solución consistente con los elementos del dominio del problema. Para lograr esto, es necesario disponer de una herramienta de modelado que brinde soporte a la comunicación entre participantes y representar estos conceptos a lo largo proceso propuesto por la IS, y que brinde una semántica común [2] [11].

La propuesta de este trabajo, pretende optar por una herramienta CASE relacionada con la enseñanza de modelado UML en el marco de la IS, y que los posibles usuarios son alumnos de nivel universitario con, posiblemente, poca o nula experiencia en modelado de software y el uso de herramientas CASE.

El primer elemento a considerar como característica es la posibilidad de brindar soporte completo a UML y sus características a nivel notación, según propone González Génova [12]. En el mismo trabajo, los autores proponen 2 tipos de herramientas CASE para modelado según su notación: a) Las herramientas sintácticas, que solo permiten dibujar diagramas correctos en base a las reglas notación de UML, beneficiando a los usuarios con asistentes que evitan errores de diseño y, b) Las herramientas Semánticas que proveen mecanismos de validación para garantizar construcciones con sentido y coherencia.

2.1 Método de Evaluación

Si bien es sencillo encontrar información cualitativa sobre las herramientas objeto de análisis, muchas veces resultan insuficientes debido a que están atadas a promesas de sus fabricantes y hasta campañas publicitarias que no caracterizan las herramientas como el cliente espera. Es por esto que, es necesario definir una metodología que permita valorar de forma cuantificable, con el fin de reducir la ambigüedad inherente a la interpretación humana.

Existen numerosos trabajos que proponen métodos para evaluar herramientas CASE acorde a un proyecto o empresa. Entre ellos, Rojas et. al, propone considerar aspectos técnicos y organizacionales (cómo la empresa se prepara para asimilar el nuevo producto). Para considerar aspectos técnicos se utilizan un conjunto de métricas definidas que refieren a la cobertura de la herramienta respecto a etapas del ciclo de vida, metodologías soportadas, trabajo concurrente, plataformas, manejo de bases de datos, generación de código, ingeniería de reverso, horario de atención del soporte técnico, material de apoyo, y temas relativo al costo [13]. Todos estos puntos que se mencionan, sugieren un ranking para comenzar con la selección de la herramienta. Se considera que este punto propone una división general sin tener en

cuenta aspectos puntuales como, por ejemplo, tipos de bases de datos que utiliza o si posee soporte para generar código a un lenguaje en particular. Creemos que la primera división es interesante para luego especificar según criterios a definir en base al uso en particular. En otro trabajo, los autores definen un conjunto de características deseables agrupados en cuatro enfoques: Enfoque procedimental, definiendo como las herramientas utilizan las metodologías para guiar al usuario a través de un proceso de IS, el segundo aspecto es sobre el soporte al modelado que proveen las herramientas, evaluando la capacidad de modelado UML para apoyar la definición de un sistema. El tercer enfoque lo denominan apoyo al repositorio, refiriéndose a la capacidad de la herramienta de mantener trabajo colaborativo permitiendo niveles de robustez y consistencia de los datos y, por último, un enfoque funcional en base a como las herramientas ayudan a los usuarios a desarrollar un sistema de información con mayor o menor facilidad, utilizando criterios de versionado, navegación de diagramas, trazabilidad, entre otros [11].

Para poder evaluar con objetividad una solución informática es importante centrarse en aquellas características deseables que describen el tipo de herramienta, por ejemplo, para una herramienta CASE de modelado es de suma importancia que posea soporte para UML no así para una herramienta CASE para análisis de riesgos o de gestión de proyectos. En este sentido, la selección de una herramienta CASE puede verse afectada por la cantidad de aspectos a considerar [13].

Otro trabajo relacionado a la selección de Herramientas CASE de modelado utiliza criterios para definir el soporte UML desde el punto de vista de notación, es decir, que la herramienta provea de todos los métodos gráficos, sintácticos y semánticos para poder dibujar diagramas, darles significado a los componentes y darles sentido a las relaciones existentes, utilizando como base el estándar UML [12]. En este trabajo, los autores proponen definir un conjunto de características y contabilizar los números de respuesta afirmativas y negativas de cada herramienta. Si bien, se plantean la dificultad de poder catalogar todas las posibles características, proponen un método dinámico para ir confeccionando las preguntas solo cuando alguna herramienta objeto de análisis no la cubre. Un último punto a considerar para definir los criterios de evaluación, se basa en la selección de un enfoque metodológico y que soporte propone UML para tal. Para ello, se tomó en cuenta el trabajo de Kruchten quien propone utilizar la metodología denominada RUP (Rational Unified Process) [14].

Es importante generar valor para el cliente porque es una meta obvia en todo desarrollo de software. Por este motivo, RUP propone como característica que el desarrollo esté guiado por casos de uso (CU). Un caso de uso describe requisitos funcionales, en termino de como los actores interactúan con el sistema. Este enfoque “*guiado por casos de uso*” permite describir como los usuarios y la interacción, van a existir para conseguir un objetivo funcional, de forma temporal y mostrando las características para una funcionalidad determinada. En RUP esta actividad es la más importante para formalizar la elicitación resultante en requisitos funcionales.

3 Propuesta de Evaluación

En base al estudio anterior proponemos el siguiente método de evaluación de herramientas CASE, el mismo está centrado en estas características deseables mencionadas en la sección anterior.

El objetivo principal es enmarcar y valorizar la herramienta ideal como punto de comparación definiendo como criterio de evaluación a un conjunto de características que a su vez dispondrán de ítems denominados *funcionalidades*, según si son aspectos funcionales a la IS o funcionales al usuario. Cada ítem identificado, podrá valorarse como *verdadero* o *falso* lo que dará un número finito que identifique en qué medida cumple las expectativas de las funcionalidades requeridas.

El total de puntos posibles estará dado por el total de funcionalidades posibles que tiene la herramienta ideal para el trabajo en cuestión, pudiendo utilizar un porcentual para identificar en qué grado cumple con las funcionalidades esperadas de una herramienta ideal, la cual es necesaria para una determinada actividad.

Por ejemplo, si la herramienta ideal posee 200 funcionalidades aceptadas y la herramienta arroja un total de 150, nos indicará que el acercamiento hacia el ideal es un 75% (aplicamos una regla de tres simples). La puntuación total posible, denominado *nivel de aceptación*, está dada por la cantidad de funcionalidades o especialidades requeridas para la evaluación.

3.1 Definición de los Criterios de Evaluación

La herramienta a evaluar será una herramienta CASE para modelado UML de tipo semántica, con utilidades para la enseñanza y aprendizaje de la IS. Para desarrollar el correcto análisis de las herramientas propuestas, se definieron criterios que consideramos acordes en base a las necesidades y dificultades detectadas con el objetivo de calificar aquellas herramientas que sirvan de forma referencial para obtener una puntuación objetiva. Escoger una herramienta CASE acorde a la aplicación que se requiere, supone mejoras durante el proceso de la IS. Para esto es necesario conocer ciertas características fundamentales. Para ello se proponen dos enfoques, uno funcional a la IS y otro funcional a la experiencia del usuario.

En base a los trabajos estudiados en el apartado anterior y desde el enfoque funcional a la IS proponemos las siguientes características a considerar para la definición de los criterios de evaluación: a) Enfoque procedimental; b) Apoyo metodológico; c) Soporte completo UML; d) Especificación de casos de uso; e) Facilidad de extensión del lenguaje; f) Modelado de Datos; g) Autogeneración de código; h) Ingeniería Inversa; i) Métricas; j) Apoyo a lenguajes formales; k) Soporte al modelado arquitectónico; l) Apoyo al modelado por capas; m) Apoyo a la enseñanza y aprendizaje de la IS. Desde el punto de vista de la experiencia del usuario, las características a considerar deberán ser las siguientes: a) Control de concurrencia; b) Versionado; c) Navegación; d) Manejo y compatibilización de diagramas; e) Visualización; f) Trabajo Online; g) Colaboración entre usuarios.

4 Criterios de Evaluación

En base a las características enumeradas en la sección anterior, se definen los siguientes conjuntos de *criterios de evaluación*. A cada criterio de evaluación se le definió un conjunto de *funcionalidades* esperadas en base al tema tratado. 1) Extensivo al proceso completo de desarrollo; 2) Adaptabilidad y flexibilidad; 3) Amigable para el usuario; 4) Producir documentación sólida; 5) Trabajo online y colaborativo; 6) Apoyo al proceso de enseñanza y aprendizaje; 7) Ingeniería Inversa y generación de código; 8) Apoyo metodológico e interpretación de invariantes OCL. A continuación, se explica brevemente cada una de éstas.

4.1 Extensivo al Proceso Completo de Desarrollo

Muchas de las herramientas actualmente vigentes en el mercado no contemplan el proceso de ingeniería de software de forma integral, sino que se especializan en una parte específica del proceso. “La herramienta de desarrollo deberá soportar todas las etapas del ciclo de vida del desarrollo de un Sistema de Información, a saber: Análisis, Diseño, Implementación y Prueba” [15].

RUP propone un conjunto de fases y disciplinas que deberán recibir soporte UML en consecuencia. También es importante que la herramienta pueda dar un soporte completo de UML, adaptador a la norma, es decir la capacidad de las herramientas de construir todos los diagramas que propone este modelo, o por lo menos los más relevantes: el diagrama de casos de uso, que representa la funcionalidad o alcance del sistema; el diagrama de clases, que escribe la estructura de objetos y sus relaciones; el diagrama de interacción (secuencia o colaboración), diagramas de actividad para describir un proceso de negocio y en algunos casos, el diagrama de estados, que visualiza el ciclo de vida de un objeto.

Las características a evaluar son las siguientes: 1) Diagramas de casos de uso: Según RUP, centrar el desarrollo de software en casos de uso permite acercarse a la validez funcional propuesta por los requisitos descriptos durante el análisis; 2) Especificación de casos de uso: describir la interacción entre el actor y el sistema de forma detallada para promover una corrección funcional; 3) Diagrama de dominio: Permite describir por medio de un diagrama de clases UML simplificado la estructura abstracta del sistema, conociendo solo entidades y relaciones más importantes. Es un modelo de análisis que será tratado independiente al diagrama de clases; 4) Diagrama de Actividades: permite modelar procesos de negocio; 5) Diagrama de clases: al refinar el diagrama de dominio obtenemos un diagrama de clases UML que permite especificar relaciones, estado y comportamiento de las entidades que se transformaran en objetos software. Es un modelo de diseño; 6) Diagrama de secuencia: Es un modelo de diseño que permite relacionar las clases software con la arquitectura y el comportamiento con el objetivo de darle una vista dinámica al sistema que se está diseñando; 7) Aspectos notacionales UML; 8) Validación de modelo UML: Para mantener la semántica de los modelos, es necesario que se puedan realizar validaciones.

4.2 Adaptabilidad y Flexibilidad

Se valora la posibilidad de que la información de distintos proyectos, guardados por una herramienta en un formato específico, pueda ser leída y compatibilizada por otras herramientas CASE. Por lo que las características a evaluar son: 1) Compatibilidad XMI: Es una especificación que permite compatibilizar entre herramientas CASE que interpreten modelos UML; 2) Exportación de imágenes: Permite portar los diagramas a otras plataformas como por ejemplo editores de texto.

4.3 Amigable para el Usuario

Según los estándares ISO dedicadas a la calidad del Software (ISO 9241 e ISO 14598) la usabilidad es “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”. Es por este motivo que resulta sumamente subjetiva la evaluación de usabilidad. Por lo que las características a evaluar son las siguientes: 1) Ayuda en Línea: El usuario puede acceder a una ayuda indexada, actualizada e idioma propio; 2) Manual de Usuario: La herramienta tiene un manual de uso disponible desde el momento de su instalación; 3) Soporte Técnico Flexible: Disponibilidad de acceso al soporte técnico de forma.

4.4 Producir Documentación

Se deberá considerar un repositorio que permita centralizar, administrar y gestionar las versiones o estados de un proyecto en el que se requieren revisiones frecuentes. La documentación deberá ser personalizable y estar basada en diagramas, modelos.

4.5 Trabajo Online y Colaborativo

Consideramos para nuestro análisis la herramienta de Google denominada *Trends* [16] que muestra un resumen de tendencias sobre un determinado criterio de búsqueda en base a un plazo de tiempo determinado. *Google Trends* es una herramienta que proporciona información sobre las búsquedas que los usuarios ingresan en Google durante un periodo de tiempo y en un área geográfica determinada, es decir el volumen total de consultas para el término de búsqueda. Los datos de consulta se analizan con un método de muestreo ya que considera solo aquellas consultas realizadas un número significativo de veces. Esta herramienta sirve para predecir el futuro, pero se puede utilizar como herramienta para “Predecir el Presente”. En estos términos, se puede lograr una predicción inmediata sobre temas de particular interés [17]. En base a tendencias analizadas, se deduce que durante los últimos años hubo una merma significativa en la cantidad de veces que se realizaron consultas relacionadas a estos conceptos de colaboración y trabajo on-line para con Herramientas CASE.

4.6 Apoyo al Proceso de Enseñanza y Aprendizaje

Para que una herramienta se adapte al proceso de enseñanza y aprendizaje debe contar con espacios específicos para el docente y para el alumno, con la posibilidad de realizar no solo trabajo de forma individual, sino también de forma colaborativa. Deberá tener también, el espacio necesario para realizar evaluación y seguimiento de trabajos [6][7].

4.7 Ingeniería Inversa y Generación de Código

Las herramientas que asisten al proceso de implementación, pueden generar código fuente. Si bien UML no es un lenguaje de programación, se puede usar para escribir programas, pero no posee la sintaxis y ni la semántica necesaria. Diferentes herramientas pueden generar código de tal manera que puedan importarse en un lenguaje de programación para luego utilizarse durante el proceso de implementación,

Generar e interpretar código existente hace que una herramienta CASE pueda participar activamente del proceso de Implementación. Para esto la herramienta debe lograr analizar diferentes lenguajes y debe ser escalable en cuanto a la diversidad de lenguajes existente. Las características a evaluar son las siguientes: 1) Generación de Código; 2) Ingeniería Inversa: Permite importar el código fuente para realizar el diagrama de clases correspondiente.

4.8 Apoyo Metodológico e Interpretación de Invariantes OCL.

Es importante que la herramienta provea soporte para alguna metodología tradicional de desarrollo de software para realizar trabajos de modelado en todas las fases principales: Análisis, Diseño e Implementación. De esta manera, el usuario deberá conocer en todo momento que rol está cumpliendo al momento de operar la herramienta. Las características a evaluar son las siguientes: 1) Soporte para identificación de Roles: deberá identificar en todo momento que rol está cumpliendo el usuario. Por ejemplo, si se está haciendo un diagrama de clases, podría indicarse que está trabajando en el diseño.

7 Conclusiones y Trabajos Futuros

En este trabajo se presentó un método de evaluación para las herramientas CASE teniendo en cuenta determinados criterios considerados en base a diversos análisis y estudios de investigación por parte de varios autores. Cada uno de los criterios que se propusieron, representan un conjunto preciso de funcionalidades o especialidades esperadas por un desarrollador de software. Se espera continuar en el estudio comparativo de diferentes herramientas CASE que se utilizó en el proceso de selección propuesto por Topper et al [18]. Este proceso consta de un conjunto de pasos bien definidos: 1) Realizar una revisión exhaustiva de las herramientas disponibles. 2) probar un pequeño grupo de una selección del paso anterior, probarlas

a través de un proyecto piloto o de una evaluación más detallada y 3) presentar una puntuación de las herramientas y seleccionar la de puntuación más alta. Mediante ello, se analizará el nivel de aceptación de las herramientas CASE de modelado más comunes en el mercado. Por último, se utilizará esta metodología para evaluar la integración propuesta en trabajos anteriores [6] [7].

References

1. Sommerville, I., & Galipienso, M. I. A. (2005). Ingeniería del software. Pearson Educación.
2. Pressman, R. S. (2010). Software engineering: a practitioner's approach. Palgrave Macmillan.
3. Mc Clure, C. O., & Ortega, J. M. (1993). Case: la automatización del software.
4. Battaglia, N., Neil C., Cardacci, D., De Vincenzi M., Martínez R. (2016, Septiembre). Evaluación y Seguimiento Durante el Proceso de Enseñanza y Aprendizaje del Modelado UML en Entornos Colaborativos. In V Workshop de Innovación en Educación en Informática (WIEI), Congreso Argentino de Ciencias de la Computación (CACIC).
5. Neil C., De Vincenzi M., Battaglia N., Martínez R. (2016). Herramientas Colaborativas Multiplataforma en la Enseñanza de la Ingeniería de Software. In XVIII Workshop de Investigadores en Ciencias de la Computación
6. Battaglia, N., Neil, C., De Vincenzi, M., & Martinez, R. (2016, Junio). UAICase: integración de un entorno académico con una herramienta CASE en una plataforma virtual colaborativa. In XI Congreso de Tecnología en Educación y Educación en Tecnología (TE&ET 2016).
7. Battaglia, N., Neil, C., De Vincenzi, M., Martínez, R., González, Dana. (2017, Junio). uCASE-CL: Aprendizaje Colaborativo de la Ingeniería de Software en Entornos Virtuales Ubicuos. In XII Congreso de Tecnología en Educación y Educación en Tecnología (TE&ET 2017).
8. Rumbaugh, J., Jacobson, I., & Booch, G. (2005). Unified Modeling Language Reference Manual, Second Edition, Addison-Wesley.
9. Watson, A. (2008). Visual Modelling: past, present and future. Online, Object Management Group, <http://www.uml.org>.
10. Larman, C. (1999). UML y Patrones. Introducción al análisis y diseño orientado a objetos. Ed. Pearson
11. Quintero, J. B., de Páez, R. A., Marín, J. C., & López, A. B. (2012). Un estudio comparativo de herramientas para el modelado con UML. revista universidad eafit, 41(137), 60-76.
12. Génova, G., Fuentes, J., & Valiente, M. (2006). Evaluación comparativa de herramientas CASE para UML desde el punto de vista notacional. Novática, 181, 59-64.
13. Rojas, T., Pérez, M., Grimán, A., Ortega, M., & Diaz, A. (2000). Modelo de decisión para soportar la selección de herramientas CASE. Revista de la Facultad de Ingeniería, UCV, 15(2), 117-144.
14. Kruchten, P. (2004). The rational unified process: an introduction. Addison-Wesley Professional.
15. Rojas, T., Pérez, M., Grimán, A., Ortega, M., & Diaz, A. (2000). Modelo de decisión para soportar la selección de herramientas CASE. Revista de la Facultad de Ingeniería, UCV, 15(2), 117-144.
16. Google Trends. (2017). Tendencias de Google. [online] Available at: <https://www.google.com.ar/trends> [Accessed 6 Jan. 2017].
17. Choi, H., & Varian, H. (2012). Predicting the present with Google Trends. Economic Record, 88(s1), 2-9.
18. Topper, A., Ouellette, D., & Jorgensen, P. (1994). Structure Methods Merger Models. Techniques an CASE Mc Graw Hill.

Uso de Controladores Difusos en el Proceso de Evaluación en Matemática

Graciela C. Lombardo , Marcelo J. Marinelli
Universidad Nacional de Misiones. Ruta 12 km 7,5, Miguel Lanus
{gracielalombardo, marcelomarinelli} @gmail.com

Resumen. Este trabajo es un avance del proyecto de investigación “Uso de la Lógica Difusa en el Proceso de Evaluación”. El objetivo fue utilizar técnicas de inteligencia artificial para obtener un sistema de calificaciones que se adapte al desempeño de los alumnos durante el cursado de una asignatura. Se optó por la lógica difusa debido a su interpretabilidad y adaptabilidad a los requerimientos del experto. Los objetos de estudio fueron las evaluaciones de una asignatura del Profesorado en Matemática de la Universidad Nacional de Misiones, correspondientes a la cohorte 2016. Para la implementación de este sistema se diseñaron controladores difusos tipo MAMDANI. Se contrastaron los resultados provenientes del sistema de controladores difusos con los procedentes de la evaluación tradicional correspondientes al mismo período. Los resultados preliminares mostraron que el sistema se adapta a las características particulares de la cátedra y la singularidad y desempeño del alumno.

Palabras clave: Evaluación, Matemática, Lógica difusa, Controladores difusos.

1 Introducción

Se considera la evaluación en un sentido amplio debido a que incluye el diagnóstico para la enseñanza, para la comprensión y finalmente para la acreditación. Interpretando que las actividades evaluativas no son todas mensurables numéricamente, es la lógica difusa una técnica apropiada debido a su amplia interpretabilidad.

Para desarrollar esta investigación se propone como objetivo desarrollar un sistema de evaluación basado en lógica difusa. Los objetivos específicos son: a) Definir, para cada instancia de evaluación, las variables lingüísticas de entrada y de salida, y determinar la totalidad de las reglas difusas capturadas a partir de la experiencia del docente. b) Seleccionar el o los métodos adecuados para la defusificación. c) Comparar los resultados obtenidos, mediante la técnica difusa, con los procedentes del proceso de evaluación histórico del mismo período.

2 Marco teórico

2.1 Evaluación

La evaluación, dentro del proceso de enseñanza y aprendizaje, es una variable compleja y controversial, que sirve para acreditar como para diagnosticar, retroalimentar, reflexionar, regular y optimizar los aprendizajes [1].

Tres son las fases o etapas que pueden distinguirse en la evaluación, claramente diferenciadas, aunque complementarias: diagnóstico inicial, evaluación diagnóstica continua y acreditación [2]. Las dos primeras tienen como objetivo determinar los conocimientos que poseen los alumnos en el área temática correspondiente a la disciplina y que deberían haberse adquirido en los cursos previos y verificar los que se han aprehendido en el período presente. De esta manera y con esta información se pueden delinear las estrategias de enseñanza y aprendizaje customizadas al grupo de alumnos. En tanto que la tercera corresponde a la certificación de conocimientos ante la institución y la sociedad [3].

Es esencial incorporar instrumentos de evaluación para obtener datos para el diseño de actividades en consecuencia con los objetivos planteados [4], entre los que se pueden citar: entrevistas clínicas, técnicas asociación libre y relaciones, gráficos o esquemas, cuestionarios, UVE heurística, mapas conceptuales, etc.

2.2 Lógica difusa

El precursor de la Lógica difusa es Lotfi A. Zadeh, quien desarrolla su actividad como profesor e investigador de la Universidad de California en Berkeley. En 1965 publicó un artículo denominado *fuzzy set*, en el cual postuló una lógica donde la pertenencia de un objeto o elemento a una clase o conjunto es gradual. Esto da origen a una lógica multivaluada a diferencia de la lógica clásica que es bivaluada (la pertenencia de un objetivo a una clase es verdadera o falsa sin permitir valores intermedios). De esta forma en la lógica difusa (*fuzzy logic*) los valores pueden tomar infinitos valores de verdad en el intervalo $[0,1]$. Además, permite utilizar términos lingüísticos de fácil interpretación y de manera parecida a como percibimos las variables, en la vida cotidiana, por medio de los cinco sentidos. Por ejemplo, puede describirse la humedad ambiente como "baja", "muy baja", "alta", "muy alta". A estos términos, los que describen a la variable, se los denomina "etiquetas lingüísticas". A diferencia de los conjuntos clásicos (*crisp set*), un elemento en el campo de la lógica difusa puede pertenecer a más de un conjunto difuso y con distintos grados de pertenencia. Se define la función de pertenencia de un conjunto difuso μA como se indica en (1):

$$\mu A: X \rightarrow [0,1] \quad (1)$$

El universo X es denominado "Universo de Discurso de la variable x ", también suele simbolizárselo con la notación U . Esto indica que la función de pertenencia está definida en el intervalo real $[0,1]$. Definiendo entonces al conjunto difuso A como se muestra en (2):

$$A1 = \{\mu_{A1}(x) / x : x \in X, \mu_{A1}(x) \in [0,1] \in \mathbb{R}\} [5]. \quad (2)$$

El razonamiento difuso se aborda mediante el tratamiento de proposiciones (reglas difusas) del tipo si-entonces (*if-then*); pero se diferencia de la lógica clásica en que los valores proposicionales pueden variar en el intervalo $[0,1]$ y la resolución de una o más reglas se realiza a través de leyes de inferencia o reglas de inferencia. Estas leyes pueden ser: modus ponens generalizado o modus tollens generalizado [6].

En 1974 el ingeniero inglés Ebrahim H. Mamdani aplicó la teoría de la lógica difusa en el proceso de control de un motor de vapor. A partir de entonces se han sucedido una gran cantidad de aplicaciones en la industria [6].

Los controladores difusos son dispositivos de inferencia que cuentan con un número determinado de variables difusas de entrada y una o más variable de salida o de control que permiten interpretar las reglas del tipo si-entonces, las cuales cuantifican en forma difusa, la descripción lingüística indicada por el experto, acerca de cómo realizar el control de una determinada tarea [7]. En la Figura 1 [5] puede observarse el diagrama en bloques de un controlador difuso (CD). Las variables de entrada y salida ingresan a las etapas de fusificación y defusificación respectivamente, ya que se supone que proviene de variables *crisp*. La base de conocimiento está constituida por la información de las particiones difusas de todas las variables y la base de reglas difusas que relacionan las variables de entrada con la o las variables de salida.

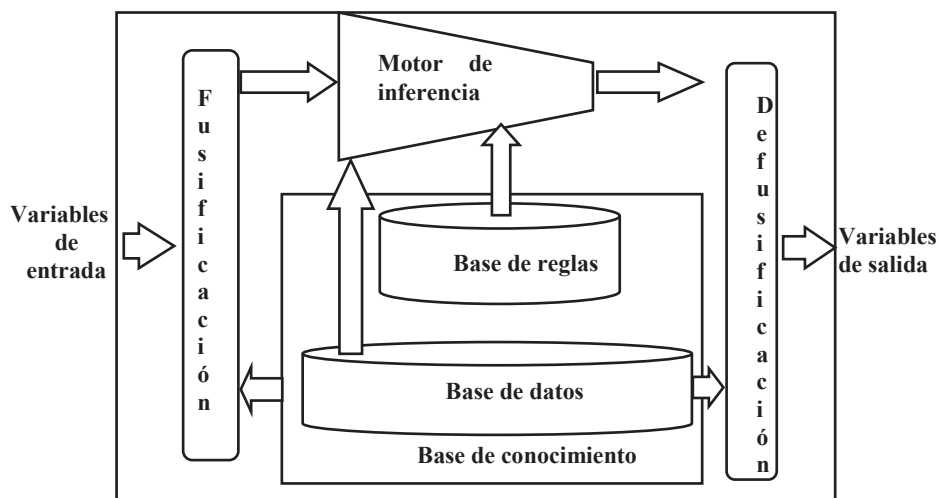


Figura. 1. Diagrama de un Controlador difuso

3 Materiales y Métodos

La investigación, del tipo exploratoria, se realizó en un curso correspondiente a la cohorte 2016, de la asignatura Geometría III (Proyectiva), del Profesorado en

Matemática de la Facultad de Ciencias Exactas, Químicas y Naturales (FCEQyN) de la Universidad Nacional de Misiones (UNaM), la cual se dicta en el tercer año de la carrera durante el segundo cuatrimestre del ciclo lectivo. La metodología se dividió en cuatro aspectos que se describen a continuación:

3.1 Delimitación de las instancias de evaluación y acreditación

Se han diseñado actividades de exposición teórica, análisis y resolución de situaciones problemáticas, trabajo de estudio grupal de temas ignotos para los estudiantes. Para esto último se conformaron grupos de trabajo con la posterior comunicación de resultados y defensa por parte de los integrantes (Entrevistas Clínicas Grupales – ECG).

El sistema de evaluación, previsto en el reglamento interno de la asignatura, fue aprobado en el año 2013, en tanto que los criterios y estándares utilizados en el proceso de evaluación son:

3.1.1 Logros y aprendizajes significativos

A efectos de evaluar el desempeño del estudiante, ya sea en exámenes parciales o en ECG, se diseñó la “Rúbrica 1”, donde se describen las evidencias a evaluar (Tabla 1) [8].

Tabla 1. Rúbrica 1

Nivel	MB	B	R	Insuficiente
Evidencia	Contesta con precisión la consigna, y muestra un gran dominio de los conceptos involucrados.	Contesta con mediana precisión la consigna, o muestra dominio de algunos conceptos involucrados.	Contesta la consigna con dificultades, o muestra un escaso dominio de los conceptos involucrados.	No responde la consigna o la responde de manera incorrecta

3.1.2 Desempeño individual y/o grupal

Para evaluar el desempeño del estudiante o del grupo al que pertenece se contemplaron los siguientes aspectos, que conforman la Actitud-proactiva, y permitieron diseñar la “Rúbrica 2”: comportamiento participativo en clase, cumplimiento de tareas o actividades asignadas, interacción con los restantes integrantes del grupo de pertenencia como con los restantes integrantes del curso, compartir sus aprendizajes con sus compañeros (Tabla 2) [8].

Tabla 2. Rúbrica 2

Nivel	Alto	Medio	Bajo
Evidencia	<p>Participa muy frecuentemente mostrando dominio de los temas abordados en clases anteriores.</p> <p>Asume un alto compromiso de compartir sus aprendizajes con sus compañeros.</p> <p>Cumple con todas las actividades asignadas.</p>	<p>Participa frecuentemente mostrando un mediano dominio de los temas abordados en clases anteriores.</p> <p>Asume un moderado compromiso de compartir sus aprendizajes con sus compañeros.</p> <p>Cumple con la mayoría de las actividades asignadas, haciendo necesario el reclamo de las faltantes.</p>	<p>Escasa participación en las clases mostrando un relativo dominio de los temas abordados en clases anteriores.</p> <p>No se involucra con el aprendizaje de sus compañeros.</p> <p>Es necesario el reclamo de un gran número de actividades faltantes, asignadas por las docentes.</p>

3.2 Selección de la herramienta de modelado difuso

Para modelar los conjuntos difusos y sistemas de inferencia se utilizó el software Matlab®. Para esto se seleccionó la herramienta “FIS” (*Fuzzy Inference System*), que es un editor para modelar controladores difusos del tipo Mamdani o Sugeno. Para la definición de las reglas de inferencia se optó por la arquitectura Mamdani.

3.3 Diseño y desarrollo del sistema de controladores difusos

Controlador difuso N° 1: Este controlador fue diseñado para la evaluación de los exámenes parciales, recuperatorios y ECG. Cuenta con cinco variables de entrada, (correspondientes a cada uno de los ejercicios de los exámenes parciales y/o núcleos temáticos de las ECG) y una de salida (una décima de la calificación). De entre las herramientas que brinda el editor FIS de Matlab® se utilizaron: el método *Prod* (producto) para el *And* (conjunción), el método *Max* (máximo) para el *Or* (disyunción), *Sum* (suma) para la agregación y el Centro de gravedad para la transformación de variables lingüísticas (*defusificación*). Las funciones de pertenencia, tanto de las variables de entrada como la de salida, se definieron como tipo gaussiana (*gaussmf*). De igual forma, se han considerado tres *set* difusos para las cinco variables de entrada, cuyas etiquetas lingüísticas son: mal, regular y bien. Tanto las variables de entrada como de salida se han normalizado en un intervalo de valores comprendido entre 0 y 1, el cual constituye el respectivo universo de discurso de la variable. En la Figura 2 se presentan las funciones de membresía con las respectivas etiquetas modeladas con FIS.

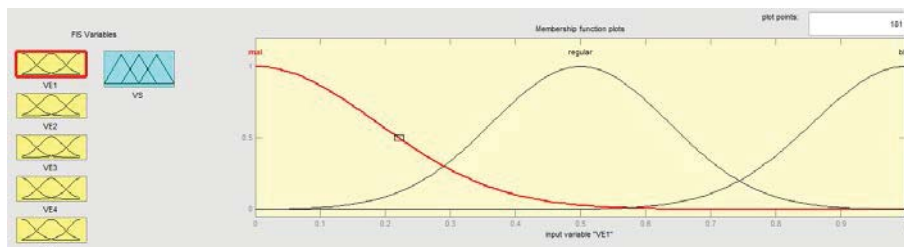


Figura 2. Set difuso de la primera variable de entrada del CD N° 1

Por otra parte, los set difusos correspondientes a la variable de salida son: desaprobado, regulariza y promociona (Figura 3). Con la combinación de las variables de entrada y salida hicieron que la arquitectura de este controlador conste de un total de 201 reglas de inferencia.

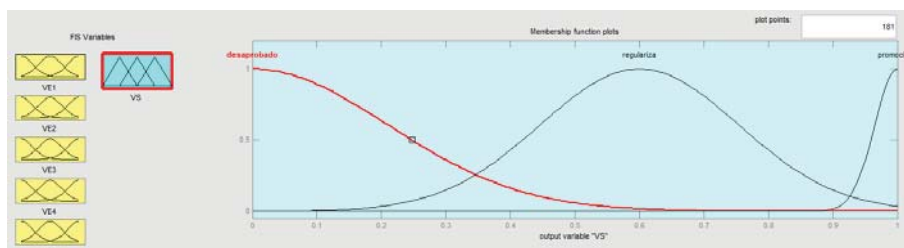


Figura 3. Set difuso de variable de salida del CD N° 1

Controlador difuso N° 2: Este controlador fue diseñado para la obtención de calificaciones, en la etapa de acreditación, de aquellos alumnos que se hallaban en condiciones de promocionar la asignatura. El mismo cuenta, por un lado con cinco variables de entrada: 1er-parcial, 2do-parcial, 1er-ECG, 2da-ECG, Actitud-proactiva; y por otro con una variable de salida: Acreditación. Los métodos que se utilizaron de la herramienta FIS son los mismos que los adoptados para el CD N°1 (Figura 4); en tanto que, para la quinta variable de entrada, Actitud proactiva, las etiquetas definidas son: baja, media, alta (Figura 5).

Respecto de la variable de salida las etiquetas son: muy bien, distinguido, sobresaliente (Figura 6). Tanto las variables de entrada como de salida se han normalizado en un intervalo de valores comprendido entre 0,6 y 1, el cual constituye el respectivo universo de discurso de la variable. La existencia de correspondencias causales entre las variables de entrada y de salida, derivó en la definición de un total de 192 reglas de inferencia.

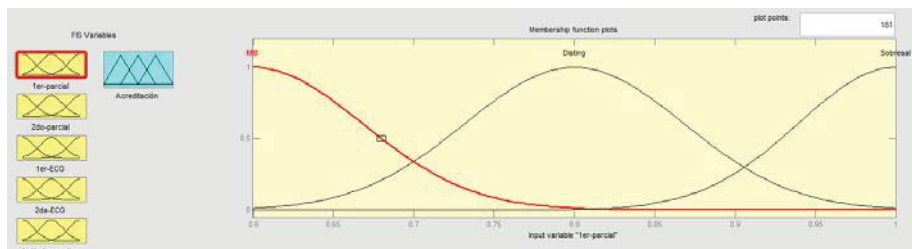


Figura 4. Set difuso de la primera variable de salida del CD N° 2

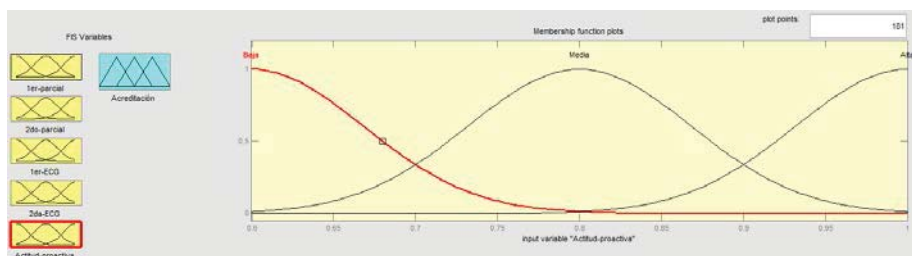


Figura 5. Set difuso de la quinta variable de entrada del CD N° 2

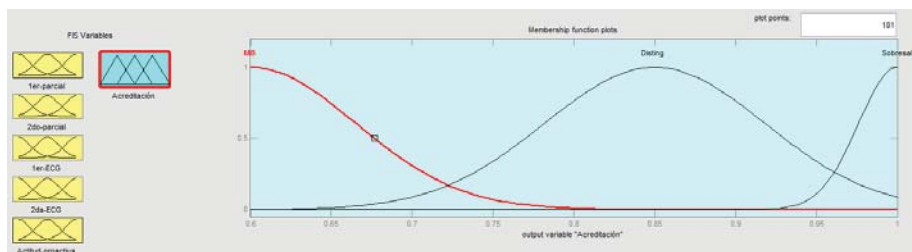


Figura 6. Set difuso de la variable de salida del CD N° 2

4 Resultados

La cantidad de alumnos que cursaron la asignatura, en el ciclo lectivo 2016, fue 34. A efectos de identificar a cada uno de los alumnos se asignó un código de manera de identificarlo por un alias.

A continuación, se presenta un ejemplo a efectos de mostrar el funcionamiento de ambos controladores difusos y la comparación con la calificación tradicional.

En la Tabla 1 se exponen los valores y notas de salida, obtenidos con el CD N° 1, de dos alumnos testigos, a partir de los valores crisp de entrada, correspondientes a las cuatro instancias evaluativas (1er parcial o recuperatorio, 2do parcial o recuperatorio, 1era ECG, 2da ECG), acompañados, en cada caso, por las notas asignadas por el cuerpo docente en la instancia de evaluación respectiva.

Tabla 1. Valores y notas de salida de instancias de evaluación obtenidos CD N° 1

	Alumno	Ej. 1	Ej. 2	Ej. 3	Ej. 4	Ej. 5	Valor salida	Nota salida	Nota Prof.
1er parcial	B1	0,8	0,25	0,5	0,8	0,5	0,605	6,05	6
	D6	0,75	0	0,5	0,75	0	0,405	4,05	4
Rec. 1er parcial	D6	0,5	0	0,8	0,8	0,5	0,606	6,06	6
2do parcial	B1	0	0,75	1	0	0	0,243	2,43	2
	D6	0	0,75	0,8	1	0	0,548	5,48	4
Rec. 2do parcial	B1	0	0,5	0	0,8	0,75	0,445	4,45	4
	D6	0,5	0	0,8	0,8	0,5	0,606	6,06	6
1era. ECG	B	0,9	0,9	0,9	0,9	0,5	0,966	9,66	9
	D	1	1	1	0,75	0,75	0,952	9,52	9
2da. ECG	B	1	0,8	0,8	0,8	1	0,95	9,5	9
	D	1	1	0,75	1	0,9	0,972	9,72	9

Si bien B1 alcanzó la nota de promoción (seis) en el primer parcial, en el segundo debió recuperarlo, no promoviendo esta evaluación, por lo que no estuvo en condiciones promover la asignatura, quedado finalmente como alumno “Regular”. Por otra parte, D6 tuvo que recuperar tanto el primero como el segundo parcial, logrando la promoción de ambas evaluaciones.

En la Tabla 2 se exponen las notas de salida, obtenidas a expensas de FIS, para las cuatro instancias evaluatorias a las que se le incorporó el valor crisp para la quinta variable “Actitud-proactiva”. Como B1 sólo pudo regularizar la asignatura, es que únicamente se cargaron, en el CD N° 2, los valores de salida del CD N° 1 pertenecientes a D6.

Tabla 2: Determinación de la nota de salida con CD N° 2 para acreditación de D6

VALORES DE SALIDA DEL CD N°1						Act. Pro-act	Valor de salida CD N° 2	Nota de salida	Nota prof.
1er Parc.	Rec	2do Parc	Rec	1era ECG	2da ECG				
0,405	0,606	0,548	0,606	0,952	0,972	0,8	0,936	9,36	8

5 Conclusiones

En el presente trabajo pudo diseñarse y construirse un sistema de dos controladores difusos para la obtención de calificaciones a efectos de la acreditación de la cohorte 2016 de Geometría III, del Profesorado en Matemática de la FCEQyN de la UNaM. Se han considerado todas las instancias y fases de evaluación. La modelización de los controladores creados permitió determinar la nota final de acreditación con un nivel mayor de certidumbre, ya que cada uno se encontraba en el umbral correspondiente (o en proximidades del umbral). Asimismo, se confrontaron los resultados vertidos por el sistema de controladores difusos con los provenientes del proceso de evaluación tradicional del mismo período. Pudo establecerse que la metodología propuesta resultó más equitativa que la tradicional debido a que, por un lado, se eliminaron posibles errores y, por otro, se modelaron las variables de entrada y salida de manera que los controladores proveen una función que las vincula dando como resultado un valor acorde a la singularidad y desempeño de cada alumno, cualidad asociada a través de variable "Actitud proactiva". Además, a diferencia del promedio se pueden dar peso a las reglas y operar con funciones de pertenencia que sean no lineales, lo cual no es posible con el promedio matemático. Debido a su interpretabilidad permite ser utilizado por usuarios no expertos en lógica difusa.

Con el diseño, construcción e implementación del sistema de controladores difusos para la obtención de calificaciones, se cumplieron tanto el objetivo general como los específicos, propuestos en el proyecto de investigación.

Referencias

1. Anijovich, R., González, C.: Evaluar para aprender. Conceptos e instrumentos. Aique Grupo Editor, Buenos Aires (2013)
2. Palou de Maté, M.: La enseñanza y la evaluación. Una propuesta para matemática y lengua. GEEMA-UNCo, Buenos Aires (2003)
3. Lombardo G., Chrobak, R.: Uso de la Entrevista Clínica en la Evaluación Diagnóstica Continua en Geometría Proyectiva. Revista de Ciencia y Tecnología, vol. 24, pp. 65-69. Recuperado el 25 de abril de 2017 de http://www.scielo.org.ar/scielo.php?script=sci_abstract&pid=S1851-75872015000200010&lng=es&nrm=iso&tlng=es. (2015)
4. Rodríguez, M.: Resolución de Problemas. En Pochulu, M., Rodríguez, M. (comp.) Educación Matemática. Aportes a la Formación Docente desde distintos enfoques teóricos, pp. 153-174. Editorial UNGS – EDUVIM, Buenos Aires (2012)
5. Marinelli, M.: Control automatizado de cultivos hidropónicos mediante Lógica Difusa. Tesis Doctoral no publicada. Facultad de Ciencias Exactas, Químicas y Naturales de la Universidad Nacional de Misiones, Posadas (2015)
6. Martín del Brío, B., Sanz Molina, A.: Redes neuronales y sistemas borrosos. Alfaomega, México (2007)
7. Gómez Salas, F.: Sistemas difusos jerárquicos para modelado y control. Tesis de Maestría. Centro de investigación y de estudios avanzados del Instituto Politécnico Nacional. Recuperada el 30 de enero de 2014 de <http://www.ctrl.cinvestav.mx/~yuw/pdf/MaTesGSFJ.pdf>

8. Lombardo, G.: Aplicación de la lógica difusa en el proceso de evaluación en Matemática. Tesis Doctoral no publicada. Facultad de Ingeniería de la Universidad Nacional del Comahue, Neuquén (2015)

Utilización de Técnicas de Análisis Automático para la Asistencia en Entornos de Introducción a la Programación

Maximiliano Chaves¹, María Marta Novaira¹,
Sonia Permigiani¹, and Nazareno Aguirre^{1,2}

¹ Departamento de Computación, FCEFQyN, Universidad Nacional de Río Cuarto,
Río Cuarto, Córdoba, Argentina

{mchaves, mnovaira, spermigiani, naguirre}@dc.exa.unrc.edu.ar

² Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina

Resumen Durante el proceso de aprendizaje de conceptos elementales de programación, los estudiantes cometen errores que derivan en programas incorrectos. Las características y tipos de errores cometidos son influenciados por diversos factores, desde el lenguaje de programación y paradigma utilizados, hasta el tipo de ejercicios de programación elegidos y la forma metodológica adoptada para la enseñanza de la programación, entre otros. En este trabajo realizamos una evaluación del tipo de errores cometidos por estudiantes de una asignatura introductoria a la programación, en el primer año de las carreras de Computación de la Universidad Nacional de Río Cuarto. Además de realizar un análisis del tipo de errores más frecuentes, y los conceptos que resultan más difíciles para los estudiantes, estudiamos cómo se podría aprovechar técnicas de análisis automático para desarrollar una herramienta de asistencia a la programación, para entornos introductorios. Analizamos el tipo de herramientas de análisis automático que pueden utilizarse para dar apoyo a los estudiantes, evaluamos su alcance y poder en un conjunto de ejercicios de programación reales, y damos algunos de los lineamientos que guiarán el desarrollo de un entorno de asistencia a la programación, que incorpore estas técnicas.

1. Introducción

La programación es una actividad intelectualmente desafiante, y su aprendizaje, como en otros campos con estas características, es complejo, puede ser frustrante y desalentador. Esto trae aparejado que la asistencia presencial, a través de docentes que acompañen el aprendizaje, asistan con las tareas en prácticas y laboratorios, puede influenciar significativamente la efectividad en el aprendizaje de los estudiantes [2]. Sin embargo, resolver problemas algorítmicos es en muchos casos una tarea que demanda tiempo, y que para muchos estudiantes es difícil de completar enteramente en el marco de actividades supervisadas, como clases prácticas y laboratorios. Más aún, en general los recursos humanos destinados a asignaturas introductorias de programación son limitados, y, especialmente en

cursos numerosos, la relación entre número de docentes y de alumnos dificulta aún más la posibilidad de asistir adecuadamente a muchos estudiantes [2]. Estas observaciones, acompañadas del creciente interés en formar estudiantes con las bases de la programación, y los avances tecnológicos y de comunicaciones actuales, han generado que se considere seriamente la posibilidad de construir asistentes automáticos para la programación. En efecto, si bien algunos de los antecedentes en esta dirección son bastante previos [3], las mencionadas innovaciones tecnológicas han reflatado el interés en este tipo de herramientas de asistencia [11].

Las características y tipos de errores cometidos por estudiantes en contextos introductorios a la programación son variados, y son influenciados por diversos factores, desde el lenguaje de programación y paradigma utilizados, hasta el tipo de ejercicios de programación elegidos y la forma metodológica adoptada para la enseñanza de la programación, entre otros [5,4]. En este trabajo realizamos una evaluación del tipo de errores cometidos por estudiantes de una asignatura introductoria a la programación, en el primer año de las carreras de Computación de la Universidad Nacional de Río Cuarto. La fuente utilizada para el análisis proviene de ejercicios y resoluciones correspondientes a exámenes de varias ediciones de la materia, todas de similares características: la materia se basa en la adopción del paradigma imperativo, utiliza el lenguaje de programación Pascal, pone énfasis en programación estructurada y el uso de diversas abstracciones procedimentales, e incluye un uso básico de especificaciones. En base al análisis del tipo de errores más frecuentes, y los conceptos que resultan más difíciles para los estudiantes, estudiamos cómo podríamos aprovechar los mismos para desarrollar una herramienta de asistencia a la programación, para entornos introductorios. En particular, evaluamos si algunas herramientas de análisis automático modernas tienen el potencial de asistir a los estudiantes en resolver los errores identificados. El objetivo final será incorporar estas técnicas de análisis en una herramienta de asistencia a la programación para entornos introductorios.

2. Análisis de Errores y sus Categorías

Los errores que cometen los estudiantes en cursos introductorios de programación son muy variados, pero pueden ser agrupados en categorías. En este trabajo tomamos como punto de partida dos categorías previamente propuestas, una más general sobre programación [4], y otra más específica para el lenguaje Java [5]. Debido a la motivación de este trabajo, descartamos algunas categorías de errores, que en general son identificados y/o capturados por un compilador estándar (y por lo tanto, ya cuentan con soporte automático para asistir al estudiante). Un ejemplo claro de este tipo de errores es la confusión entre el operador de igualdad y la asignación, que prácticamente cualquier compilador moderno ayuda a detectar. Otras categorías de errores que decidimos ignorar son aquellos que no se aplican al lenguaje Pascal, tales como errores de clases y objetos, de interfaces visuales o de formato en la entrada/salida. Finalmente, también

dejamos fuera del estudio errores correspondientes a temas más avanzados, y que exceden los vistos en cursos introductorios, como errores en implementaciones de tipos abstractos de datos, excepciones, etc. Las categorías resultantes se describen a continuación.

Resolución de Problemas. Esta categoría describe los errores que devienen de la dificultad de un estudiante de comprender un problema y/o cómo implementar una solución. Esto resulta en soluciones parciales de un algoritmo, o desarrollar algo completamente diferente a lo pedido.

Punteros. Esta categoría corresponde a errores con manejo de punteros, como por ejemplo en la implementación de estructuras enlazadas. Dentro de esta categoría se destacan el acceso incorrecto a punteros, la pérdida de referencias, y la liberación incorrecta de memoria.

Ciclos. Corresponde a errores en el uso de estructuras cíclicas del lenguaje. Incluye ciclos infinitos, y errores de tipo “off by one”.

Arreglos. Esta categoría incluye los errores específicos del uso de arreglos, como errores “off by one” y acceso fuera de límites.

Condicionales. Corresponde a los errores en el uso de estructuras condicionales, como lógica incorrecta en las condiciones, o casos Else faltantes o incorrectos en los bloques Case.

Parámetros. Esta categoría incluye errores de pasaje de parámetros, tales como pasajes por referencia innecesarios, parámetros pasados por valor que necesitan ser modificados, funciones con parámetros pasados por referencia, etc.

Sintácticos. Esta categoría agrupa errores en el uso de Pascal, como el uso incorrecto de punto y coma, posicionamiento inadecuado de Begin y End, definiciones de tipos incorrectas, errores en la escritura de estructuras cíclicas o condicionales, etc.

Retornos. Los errores de retornos incluyen: funciones que no retornan valores para todas las ramas, uso de salida estándar dentro de un módulo con variable de retorno o funciones, uso de variables globales para guardar información de retorno, etc.

2.1. Proceso de Captura

Luego de la identificación de las categorías relevantes de errores, realizamos una investigación que consistió en identificar cuáles son los errores de programación más comunes en los estudiantes de la asignatura introductoria a la programación de las carreras de Computación de la Universidad Nacional de Río

Cuarto. El análisis se basó en las categorías definidas anteriormente, y en la realización de una evaluación cuantitativa de cantidad de errores en cada categoría.

Para el análisis se tomaron exámenes parciales de la asignatura *Introducción a la Algorítmica y Programación* de las carreras de Computación de la Universidad Nacional de Río Cuarto. Se seleccionaron exámenes de los años 2011 y 2012, de temáticas introductorias a la programación tales como estructuras condicionales, estructuras cíclicas, abstracciones funcionales y procedimentales, tipo de pasaje de parámetros, implementación de estructuras de datos usando memoria dinámica (con punteros), uso de arreglos, etc. Cabe mencionar que la materia utiliza un pseudo-código para el diseño de soluciones algorítmicas, que luego son implementadas en Pascal.

2.2. Análisis de Datos

El análisis de los errores de los exámenes se presenta en la Tabla 1. Los errores de Parámetros son los más comunes, un 22% de los exámenes presentaban este tipo de errores. Los problemas más comunes en esta categoría son errores de pasajes de parámetro, esto es, modificar una variable o estructura que fue pasada por valor, o problemas de malas prácticas en pasar parámetros por referencia innecesariamente. En segundo lugar están los errores de Resolución de Problemas. Los problemas en esta categoría refieren a ejercicios en que los alumnos no entendieron la consigna del problema a resolver, o que entendido los requisitos la solución, no resuelven correctamente lo planteado. El tercer lugar es compartido por dos tipos de errores, Punteros y Retornos. En el primero se encuentran errores de tratamiento de secuencias simplemente encadenadas, como actualizaciones erróneas de referencias que derivan en la pérdida de propiedades de la secuencia. En el caso de errores de retorno, éstos significan una falta de comprensión por parte del estudiante del concepto de retorno de un valor en el caso de funciones, y el manejo de variables de resultado en el caso de una acción. Esto resulta en errores de retornos dentro de solo una rama de una condición, funciones con parámetros que guardan resultados, funciones que escriben por pantalla (salida estándar) resultados; y procedimientos que manejan variables globales para los resultados, o que también escriben los resultados por pantalla.

Daremos dos ejemplos específicos de las categorías de errores más comunes. Más ejemplos pueden encontrarse en el reporte técnico disponible en <http://dc.exa.unrc.edu.ar/staff/naguirre/education/>.

Ejercicio 1 (Segundo Parcial 2011) Dada una matriz cargada, de n filas y m columnas, desarrollar una función que cumpla con la siguiente especificación:

$$PRE : a = [a_{(1,1)}, a_{(1,2)}, \dots, a_{(n,m)}] \wedge 1 \leq n \wedge 1 \leq m \wedge num = num_0$$

$$DEF : esConstante(a, num) = \exists c(1 \leq c \leq m) : \forall f(1 \leq f \leq m) : a[f, c] = num$$

Errores	2011	Porcentaje	2012	Porcentaje	Total	Porcentaje Total
Parámetros	2	24 %	13	15 %	77	22 %
Resolución de problemas	56	21 %	10	12 %	66	18 %
Punteros	26	9 %	21	25 %	47	13 %
Retornos	2	15 %	5	6 %	46	13 %
Arreglos	2	9 %	5	6 %	30	8 %
Condicionales	2	7 %	9	10 %	29	8 %
Sintácticos	2	9 %	5	6 %	31	8 %
Ciclos	2	3 %	15	18 %	23	6 %

Cuadro 1. Tabla de errores más comunes

La siguiente resolución no resuelve correctamente el problema. El ciclo no termina cuando se encuentra el elemento; no hay retorno en el caso de no haber elemento que cumpla con el criterio de búsqueda.

```
Function esConstante (a : TArray; num:Integer): boolean;
Var
  i,j : Integer;
Begin
  p := True;
  i := 1;
  For j := 1 To Column Do
    While p = True And i<> Row Do
      If a[i,j] = num Then
        p:= True
      Else
        p:= False
        i := i + 1;
    End;
  If p = True Then
    ReturnesConstante:= True
  Else
    p := True
  End;
End;
End;
```

Ejercicio 2 (Segundo parcial 2012) Dado un arreglo de 120 valores lógicos, determinar si la cantidad de valores Verdaderos es mayor a la cantidad de Falsos, en la primera mitad del arreglo. Resolver mediante una función.

La siguiente resolución tiene un error de retorno; falta indicar qué se retorna en una rama de la condición.


```
...
Begin
  i:= 1;
  While (i<= 60) Do
    If arre[i] = True
      then
        v := v+1
      else
        f := f+1
      i := i+1;
    end;
  if v > f
  Then
    CountVF := v
  end;
```

3. Herramientas de Análisis

En base al análisis de los errores más comunes de programación presentados en la sección anterior, se derivan dos tipos de herramientas de asistencia para su solución. Por un lado, herramientas de análisis de código que puedan detectar errores que exceden aquellos que un proceso de compilación estándar normalmente identifica. En este tipo entran todas las categorías de errores a excepción de los errores de Resolución de Problemas. Asistencias relativas a este tipo de errores ayudarían, por ejemplo, a corregir y sugerir soluciones a formar de pasajes de parámetros erróneas, sugerencias sobre el uso de entrada y salida en módulos, detección de código duplicado, asistencia en la construcción de expresiones lógicas, etc.

En el otro tipo de herramienta de análisis debería apuntar a asistir en los errores que surgen de resolución de problemas. Este tipo de errores es sustancialmente más difícil de detectar, dado que esencialmente corresponden a detectar *diferencias semánticas*, es decir, discrepancias entre el comportamiento esperado de un programa y el comportamiento real del mismo. Es sabido que detectar tales problemas requiere de una especificación formal del comportamiento esperado del programa, y el problema es indecidible [6], pues corresponde esencialmente a comprobar que un programa satisface su especificación. Existen sin embargo, como mostraremos más abajo, formas de comprobar, al menos parcialmente, que un programa cumple con el comportamiento esperado del mismo.

3.1. Errores de Parámetros: Pascal Analyzer

Para la resolución de errores de tipos de pasajes de parámetros y similares, una técnica utilizada con frecuencia es el análisis estático extendido [7]. Como en el caso particular de nuestro entorno de evaluación el lenguaje utilizado es

Pascal, necesitamos una herramienta de este tipo para este lenguaje. Lamentablemente, estas herramientas no abundan para este lenguaje. El único ejemplo que conseguimos encontrar es una herramienta comercial denominada Pascal Analyzer (PAL). Esta herramienta realiza un análisis estático extendido y genera un reporte en base a una implementación en Delphi (Pascal orientado a objetos). La herramienta recibe un conjunto de archivos fuente, y presenta en distintas categorías, informes, correcciones, sugerencias y advertencias sobre el código, generados en base a un análisis estático extendido.

Esta herramienta produce salidas sumamente detalladas para una cantidad de errores posibles, muchos de los cuales sólo se aplican a programas Delphi. En nuestro caso, sólo estamos interesados, en principio, en los errores que tienen que ver con pasaje de parámetros. La herramienta es razonablemente potente en este sentido, permitiendo detectar errores tales como variables pasadas por referencia y no seteadas, variables pasadas por valor y modificadas en el cuerpo de la rutina, entre muchos otros.

3.2. Errores de Resolución de Problemas: Pex

Como mencionamos anteriormente, los errores en resolución de problemas son sustancialmente más difíciles, en relación a la provisión de asistencia automática al estudiante. En este aspecto, nuestro candidato inicial, como tecnología de análisis automático a utilizar, es la ejecución concreta/simbólica [10]. Específicamente, proponemos utilizar Pex [9], una herramienta de generación automática de tests, en un sentido similar al propuesto en [11]. Pex genera tests automáticamente intentando cubrir ramas en el código. La idea esencial consiste en utilizar una implementación secreta, considerada correcta, y provista por el docente. Luego, se utilizará Pex para intentar cubrir ramas en el siguiente programa:

```
program (x) {  
    if (implementacionEstudiante(x) != implementacionSecreta(x))  
        throw new Exception();  
}
```

Pex intentará activamente, mediante constraint solving, cubrir ramas de este programa, en particular la rama en la cual se lanza una excepción. Notemos que si tal rama es cubierta, lo que Pex proveerá es valores para los parámetros del programa para los cuales los resultados obtenidos por la solución del estudiante y la implementación secreta del docente, difieren.

Una alternativa al uso de Pex sería simplemente contar con una cantidad de tests provistos por el docente, que capturen, para una cantidad limitada de escenarios particulares, el comportamiento esperado del programa a desarrollar. La ventaja de utilizar una herramienta como Pex es que el análisis es sustancialmente más exhaustivo que el que, en principio, se puede realizar sólo mediante el uso de una cantidad acotada, ad hoc, de tests.

En el marco de este trabajo, el uso de Pex como herramienta de asistencia no es directo. La herramienta funciona para proyectos en cualquier lenguaje de programación sobre .NET, por ejemplo, C#. No soporta Pascal, lamentablemente.

Esto nos obliga entonces al uso de traductores de código Pascal a C# (u otro lenguaje .NET). Esta traducción no es simple, y dificulta la aplicación del análisis en muchos casos. Por ejemplo, la traducción de un programa Pascal a C# hace que el programa C# resultante, al ser compilado, detecte muchos problemas (no de tipo semántico) que el compilador Pascal pasa por alto. Concretamente, el compilador Pascal compila normalmente si un algoritmo tiene una función que no retorna en todas las ramas, mientras que el compilador de C# sí. Por otra parte, el funcionamiento de los dos lenguajes involucrados en la traducción tienen semánticas, en particular en lo referido a pasaje de parámetros, muy diferentes. En particular, módulos con parámetros pasados por referencia en Pascal no son fáciles de traducir a C#, donde el único tipo de pasaje de parámetros es por valor.

3.3. Evaluación de las Técnicas de Análisis

Nuestra evaluación de las técnicas de análisis elegidas se realizó sobre un amplio número de casos de estudio, tomados de resoluciones reales de exámenes parciales. En esta sección, y por razones de espacio, sólo mostramos los resultados obtenidos por parte de las herramientas antes descritas, para los dos ejemplos mencionados anteriormente en el artículo. Más datos pueden encontrarse en el reporte técnico extendido correspondiente a este artículo.

Ejercicio 1. Dado el tipo de error en la resolución ejemplo de este ejercicio, la herramienta a utilizar aquí es Pex. Al proveer a Pex de una versión correcta del programa (secreta para el estudiante) y contrastarla como hemos indicado anteriormente, con la resolución del estudiante, obtenemos el siguiente contra-ejemplo:

```
a = {{0,1},{0,0}}; num = 1
```

En este caso el programa del estudiante retorna true (debido a que el valor de la variable *i* no es reseteado luego de la ejecución del while), mientras que el programa secreto (correcto) retorna false. Este es sólo un ejemplo de un error detectado por la diferencia semántica realizada por Pex. Otros inputs reportados por la herramienta detectan problemas de terminación en el programa del estudiante.

Ejercicio 2. Dado el tipo de error en la resolución ejemplo de este ejercicio, la herramienta a utilizar aquí es Pascal Analyzer. Pascal Analyzer produce, entre el enorme reporte de problemas para este ejercicio, lo siguiente:

```
Local variables that are referenced before they are set:  
f : Integer                               Var, Local
```

Efectivamente, el analizador estático extendido detecta en este caso que existen caminos de ejecución del programa en los cuales se referencia (hace uso) de `f`, antes de ser seteada (en un momento en el cual contiene basura).

Es interesante mencionar además que otros errores de tipo de manejo de parámetros y entrada/salida, como en este ejemplo el de no tener un valor de retorno en cada rama, es detectado al intentar utilizar Pex, durante la traducción y compilación en C#.

4. Conclusiones y Trabajos Futuros

En este artículo hemos analizado cuantitativamente los errores que más afectan a estudiantes de las carreras de Computación de la Universidad Nacional de Río Cuarto, en la asignatura introductoria a la programación. La finalidad de este estudio es, como explicamos en el artículo, evaluar si algunas técnicas de análisis automático modernas podrían aplicarse para dar asistencia automática a los estudiantes, y así contribuir a lidiar con algunos problemas comunes de aprendizaje en los estudiantes. Identificamos los errores más comunes, y analizamos para dos de ellos la utilización de análisis estático extendido, y de ejecución concreta/simbólica, como tecnologías de análisis que podrían aprovecharse para la asistencia. Elegimos herramientas concretas de análisis y las aplicamos a un número de ejemplos concretos, tomados de resoluciones reales de exámenes parciales. A pesar de las dificultades que tienen que ver con el lenguaje de programación usado en la asignatura (Pascal) y la casi total ausencia de herramientas de análisis para este lenguaje, descubrimos que muchos de los errores cometidos por los estudiantes pueden ser identificados a través del uso de las técnicas mencionadas, satisfactoriamente. Sin embargo, la forma en que estos errores son reportados es en general bastante críptica y técnica, y por lo tanto se debería post-procesar la salida de estas herramientas para dar feedback a los estudiantes de manera más comprensible y dirigida.

En cuanto a trabajos futuros, como hemos mencionado en diversas ocasiones en el artículo, nuestra finalidad es desarrollar una herramienta de asistencia a los estudiantes. En base al estudio realizado y a la evaluación preliminar de las técnicas de análisis presentada, comenzaremos a implementar un primer prototipo de asistente. Planeamos poder evaluar la adopción del asistente, como así también en valor del feedback provisto por los estudiantes, en un estudio futuro. También extenderemos el análisis presentado en este artículo a otras categorías de errores, que seguramente demandarán incorporar otras tecnologías de análisis, además de las ya analizadas.

Finalmente, es parte de nuestra intención desarrollar a futuro esquemas de gamification [8] que faciliten la adopción de la plataforma de asistencia automática a la programación, siguiendo el relativo éxito que ha tenido esta política en contextos similares al nuestro [1].

Referencias

1. J. Bishop, R. Horspool, T. Xie, N. Tillmann y J. de Halleux, *Code hunt: experience with coding contests at scale*, Proceedings of the 37th International Conference on Software Engineering ICSE 2015, IEEE Press, 2015.
2. B. Bloom, *The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring*, Educational Researcher 13(6), 1984.
3. R. Farrell, J. Anderson y B. Reiser, *An Interactive Computer-Based Tutor for LISP*, Proceedings of the National Conference on Artificial Intelligence, AAAI Press, 1984.
4. A. Hansen, *An Empirical Study of Students Programing Bugs*, Utah State University, Logan, UT, 2011.
5. M. Hristova, A. Misra, M. Rutter y R. Mercuri, *Identifying and Correcting Java Programming Errors for Introductory Computer Science Students*, en Proceedings de 34th SIGCSE Technical Symposium on Computer Science Education SIGCSE 2003, ACM, 2003.
6. C. Nelson, *Techniques for program verification*, Doctoral Dissertation, Stanford University, 1980.
7. F. Nielson, H. Nielson y C. Hankin, *Principles of Program Analysis*, Springer, 2010.
8. K. Seaborn y D. Fels, *Gamification in theory and action*, International Journal of Human-Computer Studies 74(C), Academic Press, 2015.
9. N. Tillmann y J. de Halleux, *Pex: white box test generation for .NET*, Proceedings of the 2nd international conference on Tests and proofs TAP 2008, LNCS, Springer, 2008.
10. N. Williams, B. Marre, P. Mouy, *On-the-Fly Generation of K-Path Tests for C Functions*, Proceedings of the 19th IEEE International Conference on Automated Software Engineering (ASE 2004), IEEE Computer Society, 2004.
11. T. Xie, N. Tillmann y J. de Halleux, *Educational software engineering: where software engineering, education, and gaming meet*, Proceedings of the 3rd International Workshop on Games and Software Engineering GAS 2013, IEEE Press, 2013.

Virtualización en la Educación: Laboratorio Portátil de Redes

Barrionuevo Mercedes¹, Gil Cristian, Giribaldi Matias, Suarez Christopher,
Taffermaberry Carlos¹²

¹ Universidad Nacional de San Luis,

Ejército de los Andes 950 - 5700 - San Luis - República Argentina

² Universidad Tecnológica Nacional, Facultad Regional Mendoza

Rodriguez 273, Ciudad Mendoza -M5502AJE - República Argentina

{mbarrio, ctaffer}@unsl.edu.ar

Abstract. El principal objetivo de la virtualización es la reproducción completa de una red de computadoras físicas en software, pudiendo trabajar con dispositivos y servicios de red lógicos. Ayudar a la metodología de enseñanza tradicional mediante la experimentación usando un laboratorio virtual y portátil permitirá realizar pruebas en cualquier momento y lugar. LPR (Laboratorio Portátil de Redes) es una herramienta que permite realizar un ambiente virtualizado de redes de computadoras permitiendo conectar, configurar y comunicar máquinas clientes y servidores. Su objetivo es permitirle al alumno un ambiente apto para verificar el correcto funcionamiento de los servicios de red que se necesiten. En este trabajo se presentan las consideraciones principales tanto del diseño como de la implementación de LPR, además de los resultados obtenidos.

Palabras claves: Virtualización. Servicios de Red. Software Libre. Redes de Computadoras. Educación.

1 Introducción

En la actualidad, nuestra sociedad experimenta Internet a través del uso del protocolo World Wide Web, los correos electrónicos, programas que permiten compartir información, etc., sin la necesidad de conocer cómo es que estas aplicaciones funcionan realmente. Sin embargo, a medida que avanzamos en el estudio de las redes de datos se vuelve cada vez más importante conocer cómo un servicio de red es capaz de transmitir e interpretar los mensajes enviados y recibidos a través de la red.

Poder configurar y probar el correcto funcionamiento de los servicios más comunes utilizados en Internet es fundamental para todo estudiante de Tecnologías de la Información y la Comunicación (TIC). Realizar diversas pruebas en un ambiente seguro, donde se pueda volver a una configuración inicial y estable, es un objetivo de toda cátedra dedicada a la puesta a punto de los servicios de red.

Permitir a diversas aplicaciones ejecutarse en una red virtual exactamente igual que en una red física es un logro de la virtualización de redes. Por otra parte, realizar diversas pruebas de configuración y ejecución de comandos o de aplicaciones

pudiendo guardar el estado actual y recuperar el trabajo realizado luego de un tiempo es una de las ventajas del uso de diversos entornos de trabajo virtualizados.

En base a lo mencionado anteriormente, se propone la creación de un Laboratorio Portátil de Redes, una herramienta fundamental para el desarrollo de las prácticas en las materias Servicios en SO de Redes y toda aquella materia de la Tecnicatura Universitaria en Redes de Computadoras (TUR) de la Universidad Nacional de San Luis que requiera su utilización, buscando ayudar a los profesores y estudiantes en el proceso de enseñanza-aprendizaje. LPR surgió como trabajo final de aplicación de los alumnos autores de este documento.

El resto de este documento está organizado de la siguiente manera: la próxima sección describe los conceptos teóricos involucrados en el desarrollo del prototipo. La sección 3 especifica aspectos básicos del diseño de la herramienta propuesta, detallando el contexto donde será utilizada y las problemáticas abordadas. En la sección 4 se muestran los ensayos y pruebas experimentales. A continuación se especifican los criterios utilizados en la selección de la tecnología a usar y la evaluación del prototipo. Finalmente, se detallan las conclusiones y trabajos futuros.

2 Marco Teórico

Esta sección introduce diferentes conceptos utilizados en el desarrollo de nuestro trabajo, destacándose la virtualización, tipos e implementaciones. A continuación, se enumeran los principales de ellos.

2.1 Virtualización

Normalmente, cuando se utiliza una computadora, el Sistema Operativo (SO) es instalado y ejecutado directamente sobre el hardware y de esta forma se aprovecha de forma completa su potencial. Sin embargo, existe un tipo de programa llamado software de virtualización, o máquina virtual, encargado de emular un hardware determinado, aprovechando los recursos reales de la computadora y, sobre este software, es posible instalar un SO como si se tratase de una computadora real. El SO huésped funciona, a grandes rasgos, como si fuera un SO principal, por lo que todas sus funciones y características están disponibles, convirtiéndolo en una herramienta ideal para realizar pruebas [1].

Cuando hablamos de virtualización hacemos referencia al proceso de reemplazar dispositivos físicos por dispositivos virtuales, disponibles mediante el uso de un software. Se pueden virtualizar servidores, estaciones de trabajo, redes y aplicaciones. Para ello, el software de virtualización administra los recursos físicos de esa máquina: memoria, CPU, almacenamiento y ancho de banda de la red, entre los aspectos más relevantes.

Junto con la disminución del uso de equipos físicos, la virtualización trae como beneficios la reducción de costos de mantenimiento y consumo energético. Esto deriva una consolidación de servidores, optimizando el uso del espacio físico.

2.2 Tipos de Virtualizaciones

Cuando se trata de la virtualización, existen varias maneras de lograr el mismo resultado a través de diferentes niveles de abstracción, siendo los más utilizados [2, 3]:

- *Paravirtualización:* Este mecanismo agrega un conjunto especial de instrucciones (llamado Hypercalls) reemplazando las instrucciones del set de instrucciones referidas a la arquitectura de la máquina real. En la arquitectura x86, por ejemplo, el hipervisor se ejecuta justo por encima del hardware físico (Ring 0) para que los SO huéspedes se ejecuten en niveles superiores. El núcleo de los SO huéspedes necesita ser modificado para poder utilizar los Hypercalls. Por otro lado, tiene como ventaja una sobrecarga de virtualización baja.
- *Nivel de SO:* Esta técnica virtualiza el servidor físico en el nivel del SO. Aquí, el SO anfitrión es un kernel modificado que permite la ejecución de múltiples contenedores aislados también conocidos como Virtual Private Servers (VPS) o servidores virtualizados. Cada contenedor es una instancia que comparte el mismo kernel del SO anfitrión. Tiene como ventaja una baja sobrecarga, y sus implementaciones son ampliamente utilizadas. El principal inconveniente es que no admite virtualizar otros tipos de SO.
- *Virtualización completa:* Esta técnica se utiliza para la emulación de una arquitectura de procesador sobre la arquitectura de otro. Permite ejecutar SO huéspedes no modificados, emulando cada instrucción del primero por medio de la traducción de código. La ventaja de la emulación completa del procesador es su portabilidad multiplataforma de aplicaciones y de SO. La desventaja es la sobrecarga causada por la emulación en software del conjunto completo de instrucciones.
- *Asistida por hardware:* Este tipo de virtualización implementa un nuevo anillo (Ring -1) con un modo de privilegio más alto en la arquitectura del procesador normal de cuatro anillos. Las extensiones de CPU para el soporte de virtualización permiten ejecutar los SO huéspedes sin modificación para que trabajen en Ring 0, mientras que el Hipervisor se ejecuta en Ring -1. Esto permite soportar la virtualización sin necesidad del uso de *Virtualización Completa* o *Paravirtualización*. La ventaja de esta técnica es la reducción de sobrecarga causada por la emulación de software.

2.3 Implementaciones de virtualización

Para poder implementar ambientes de virtualización existen diversas tecnologías. En esta subsección se hace una revisión de algunas tecnologías existentes:

- *KVM (Kernel based Virtual Machine)* [4]. Es una herramienta de software libre la cual permite la virtualización asistida por hardware, para plataformas Intel y AMD. Está incluida por defecto a partir del Kernel 2.6.20 de Linux, permitiendo así una rápida implementación. Esta es soportada por la comunidad y tiene una gran cantidad de herramientas de administración que pueden ser usadas en conjunto.
- *Xen* [5]. Es un hipervisor desarrollado con software libre para virtualización. La virtualización puede llevarse a cabo usando Paravirtualización o Asistencia por hardware. Tiene un gran soporte en la comunidad de desarrolladores y su implementación está madura, pues lleva más de 10 años en uso.
- *LXC (Linux Container)* [6]. Es una interface en espacio usuario para virtualización a nivel SO, y lo logra por medio de una potente API (Application Program Interface) y herramientas simples, permitiendo a los usuarios de Linux crear y manejar contenedores. El desarrollo de esta interfaz fue realizado utilizando software libre y el soporte está incluido actualmente en la mayoría de las distribuciones de Linux. Su desarrollo data del año 2008.
- *Docker* [7]. Es una herramienta de software libre que virtualiza en Nivel SO, permitiendo empaquetar entornos y aplicaciones que posteriormente se pueden desplegar en otro SO con esta tecnología. Docker accede a la virtualización del kernel de Linux a través de la biblioteca *libcontainer* [8], o indirectamente a través de LXC. El proyecto inició en el año 2013.
- *Proxmox* [9]. Es una plataforma de código abierto que no permite virtualizar por sí misma. Utiliza el hipervisor KVM para realizar virtualización asistida por hardware y LXC para hacer virtualización en nivel de SO. Dispone de una página web de administración muy completa, permitiendo realizar toda la administración necesaria, y está basada en la distribución Debian, disponible desde el año 2009.
- *VirtualBox* [10]. Es un software de virtualización que utiliza técnicas de asistencia de hardware y virtualización completa. El software es ofrecido bajo una licencia de software privativo, pero en el año 2007, después de años de desarrollo, surgió la versión Open Source Edition, VirtualBox OSE bajo la licencia GPL 2.
- *QEMU* [11]. Es una herramienta que permite virtualización y emulación de la arquitectura de hardware. Cuando es usado para virtualizar, QEMU utiliza la técnica de Virtualización Completa, mientras que cuando es usado como emulador de hardware, puede ejecutar programas para otra arquitectura. Puede utilizarse en conjunto con el Hipervisor Xen o KVM. La primera versión estable del software data del año 2006, por lo cual su desarrollo está maduro.

Si bien existen otros software de virtualización, solo se consideraron los anteriores por cuestiones relacionadas a licencias libres.

3 Diseño de un Laboratorio Portátil de Redes - LPR

Este trabajo tiene como objetivo general crear un ambiente donde se pueda construir una red de computadoras usando un método de virtualización. A continuación se

detalla el contexto donde va a ser utilizado, los problemas y motivaciones, los objetivos perseguidos y las tecnologías utilizadas para la creación del prototipo LPR.

3.1 Contexto

Un servicio de red es una aplicación que provee información a distintos clientes que la soliciten. Generalmente los servicios de red son instalados en uno o más servidores para compartir información y recursos a computadoras clientes.

El trabajo desarrollado tiene como objetivo ser utilizado en la materia Servicios de Sistemas Operativos correspondiente al 3° año de la Tecnicatura Universitaria en Redes. Esta materia tiene un alto porcentaje práctico dado que se basa principalmente en la configuración, puesta en marcha y prueba de los servicios.

Los servicios de red de interés para la materia, son: Protocolo de Configuración Dinámica de Host (DHCP), Correo electrónico (SMTP), Servicio de Nombre de Dominio (DNS), Protocolo de Transferencia de Hipertexto (HTTP), Proxy, Firewall y Redes Privadas Virtuales (VPN) y Server Message Block (SMB).

3.2 Problemática y motivación

Dada la naturaleza práctica de la materia, en su primer dictado durante el año 2009, se presentaron varios inconvenientes para llevar a cabo de manera eficiente dicha práctica en los laboratorios de la Facultad de Ciencias Físico Matemáticas y Naturales (FCFMyN) de la UNSL. Algunos de ellos se detallan a continuación:

- Limitación en cantidad de redes para realizar los prácticos de encaminamiento en capa de Red. Se necesitaban al menos dos redes físicas distintas o un dispositivo activo con soporte para crear distintas Vlan's pero no se contaba con ellos.
- Limitación en hardware con respecto a placas de red para computadoras multi-homed. Las máquinas que tenían el rol de encaminadores, necesitaban al menos dos placas de red, y no se contaban con ellas.
- Limitación en cantidad de computadoras, las cuales permitieran consumir servicios desde muchos clientes en forma simultánea.
- Inexistencia de software libre para instalar algunos servicios en el SO propietario instalado en las computadoras.
- Falta de permisos de administración en los SO de las computadoras del laboratorio para instalar y/o modificar los servicios a ensayar.
- Propagación de servicios de prueba a toda la red de la Facultad, como DHCP y DNS, entregando direcciones de red, puertas de enlace y resolución de nombres ficticias.
- Finalmente, si los alumnos intentaban recrear los mismos prácticos en sus hogares, se encontraban con las mismas o mayores limitaciones, debido a que generalmente contaban con una sola computadora.

Con el afán de mejorar las prácticas de laboratorio, y en particular el proceso de enseñanza-aprendizaje, surge la motivación para la realización de LPR.

3.3 Trabajos previos: Distribución ADIOS

Una alternativa a algunos de los problemas presentados en el punto 3.2 fue buscar entre diversos softwares la posibilidad de virtualizar escenarios de pruebas.

Se evaluaron distintas alternativas de simulación, como Cisco Packet Tracer [12], CLOONIX [13] o herramientas de virtualización como VNUML (Virtual Network User Mode Linux) [14] de la Universidad Politécnica de Madrid, o ADIOS [15] de la CQ University de Australia. Por funcionalidad y uso de escasos recursos de hardware se seleccionó ADIOS, la cual es una distribución basada en Fedora Release 8 live CD que utiliza User Mode Linux (UML) [16] para crear máquinas virtuales, tiene algunos servicios de los necesitados en la cátedra ya instalados, y una topología de red muy flexible [17], como se puede observar en la figura 1.

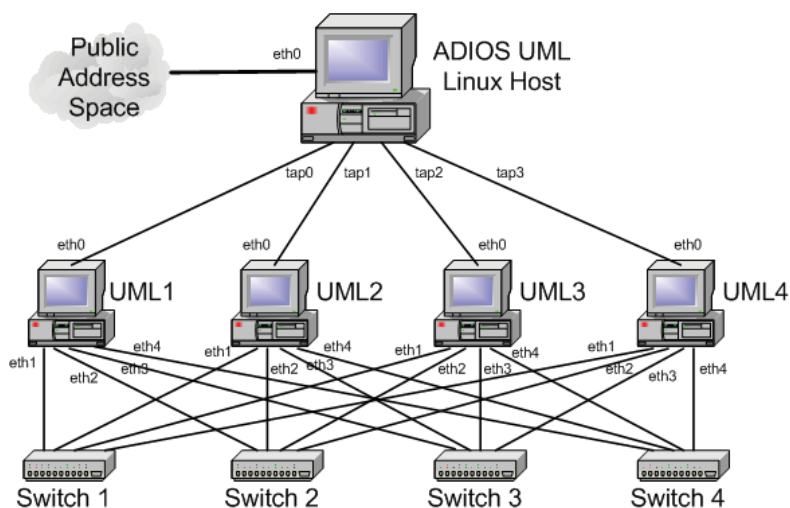


Figura 1. Detalle de la interconexión de las placas de red de las VM de ADIOS

Como la distribución es libre, fue posible hacer varias modificaciones a lo largo de los años, adaptándola a lo necesitado por la cátedra. Ellas son:

- Desasignación de direcciones de red preconfiguradas, e interfaces activas, para que los alumnos activen y configuren las necesarias en cada práctico.
- Instalación del servicio OpenLDAP, y clientes telnet, elinks, mutt y thunderbird.
- Solución de problemas con direcciones de IPv6 asignadas mediante EUI-64, debido a direcciones mac mal generadas.
- Creación de scripts para almacenar los cambios en los archivos de configuración e historial de comandos ejecutados en cada máquina virtual.
- Modificación de accesibilidad, como: tamaños de letra en las terminales y creación de una GUI para iniciar o detener las máquinas virtuales.

En la actualidad, la distribución (con un tamaño de 650 Mb) es totalmente funcional para realizar todas las prácticas de la cátedra, permitiendo utilizarla desde un liveCD. Por lo tanto, sin instalar software alguno, los alumnos pueden realizar los ensayos en

el laboratorio de la Facultad, o en su casa, si disponen de una computadora con lectora de CD y al menos 512 Mb de memoria RAM.

Sin embargo, surgieron nuevos problemas originados por la actualización del hardware en los laboratorios de la FCFMyN hace dos años. Se observó que el kernel de ADIOS (2.6.24) no reconoce algunos nuevos procesadores multi-núcleo, por lo que no funciona en ellos.

Junto con esta nueva limitación, la versión de los servicios sobre los que está desarrollando ADIOS tiene diez años, muchos de los servicios actualmente presentan nuevas características y son deseables de enseñar en la cátedra. Se intentó actualizar en ADIOS la versión de los distintos servicios utilizados, pero por problemas de dependencia de otros paquetes y del kernel mismo, fue imposible realizarlo.

3.4 Objetivos Propuestos:

Los objetivos y alcances específicos de LPR se detallan a continuación:

- Desarrollar una herramienta de software libre.
- Desplegar una red de máquinas virtualizada en una única computadora física, sin tener que instalar software alguno en la misma.
- Trabajar con al menos 6 nodos con distintos tipos de SO.
- Interconectar de diversas maneras los nodos creados para lograr distintas topologías de red en capa de enlace.
- Almacenar toda la configuración realizada en usos futuros.
- Crear una imagen de arranque en dispositivos tales como DVD o USB.
- Optimización de uso de recursos para que pueda ser utilizado en computadoras con limitaciones en el hardware.

3.5 Tecnologías propuestas para el desarrollo:

Para la implementación de los distintos nodos se propone la utilización de tecnologías de virtualización, lo que permite abstraer y aislar los distintos nodos entre sí. Entre ellas se destacan KVM, XEN, LXC, VirtualBox, Qemu, Uml.

La conectividad de capa de enlace entre nodos está dada por el uso de interfaces virtuales (vnics) y switch virtuales, como Open Vswitch. Un método alternativo a considerar es el uso de Vlans.

En cuanto a la implementación de servicios, se plantea la utilización de software GPL, el cual permite mayor flexibilidad a la hora de realizar configuraciones.

El almacenamiento en las máquinas virtuales se propone a través de la utilización de tecnologías de Snapshots, aplicadas en cada máquina virtual (dependiendo del hipervisor a usar), o mediante el uso de tecnologías de almacenamiento como LVM.

Con respecto a la optimización de recursos en las máquinas virtuales se propone el uso de una interface de tipo CLI (Character Line Interface), como el Bourne Shell.

4 Ensayos y pruebas

En esta sección se presentan las pruebas realizadas en cada una de las virtualizaciones. Para ello se utilizaron dos computadoras, con el siguiente hardware:

- Computadora 1: procesador AMD Athlon (tm) 64 X2 Dual Core con 2 Gb de Memoria Ram, almacenamiento en disco SAMSUNG HD080HJ de 8g Gb. Motherboard ASUS M2N-MX SE Plus.
- Computadora 2: procesador AMD Athlon 64 X2 Dual core 5000+ con 2Gb. de Memoria Ram, almacenamiento en disco de 160Gb. Motherboard ASUS M2N-MX SE Plus.

Por otra parte, los SO utilizados fueron Ubuntu 12.04 LTS, Debian 8 y Debian 9.

4.1 Tareas realizadas en cada ambiente de virtualización:

- Estudio de las tecnologías LXC, VNX, CLOONIX y Snapshots.
- Virtualización de redes LAN y WAN, probando la conexión entre los distintos hosts.
- Interconexión entre dos, cuatro y seis host virtuales.
- Verificación del consumo de recursos de las máquinas virtuales.
- Instalación, verificación y prueba de los servicios utilizados en la materia.
- Acceso e interacción a un escritorio remoto de una máquina virtual usando la herramienta de software libre VNCserver.

Siguiendo el esquema de red presentado por ADIOS, fue utilizado el soporte de Linux para Bridging y, de acuerdo a la documentación de Docker al respecto [18,19], se logró configurar exitosamente un ambiente con iguales características.

4.2 Comparativa de tecnologías:

Luego de la instalación de todos los requisitos para el correcto funcionamiento de cada una de las virtualizaciones se realizó una comparativa entre cada una de ellas. La Tabla 1 muestra en detalle la comparación realizada.

Características	KVM	VIRTUAL BOX	QEMU	LXC
<i>Espacio en Disco</i>	6 Gb por máquina virtual	8 Gb por máquina virtual	2 Gb por máquina virtual	488 Mb por contenedor
<i>Espacio en memoria RAM</i>	300 Mb por máquina virtual	512 Mb por máquina virtual	512 Mb por máquina virtual	14 Mb por contenedor
<i>Interconexión de host</i>	SI	SI	SI	SI

Tabla 1. Comparativa entre las virtualizaciones realizadas

5 Criterios de selección de tecnología y Evaluación del prototipo

Como uno de los objetivos planteados es que la práctica de laboratorio pueda ser reproducida por los alumnos en sus hogares, y como no todas las computadoras cuentan con soporte de hardware para virtualización, las técnicas que utilizan asistencia por hardware deben ser descartadas y, en consecuencia, KVM no sería utilizado.

Además, no es necesario emular procesadores con otra arquitectura en las prácticas planteadas, por lo cual el uso de QEMU para virtualizar no aporta ninguna mejora.

Finalmente, a partir de la comparación realizada en la Tabla 1, y teniendo en mente que debe poder ser ejecutado en hardware con pocos recursos, es que se decidió realizar el prototipo a evaluar utilizando la herramienta DOCKER.

El prototipo desarrollado está en proceso de evaluación. Si bien ha tenido una evaluación preliminar por parte de los integrantes de la cátedra, la evaluación real y exhaustiva se llevará a cabo cuando se presente el próximo año a los alumnos de la materia Servicios en SO de Redes. De ahí se podrán analizar las debilidades y fortalezas presentes en LPR.

Sin embargo, se pudo observar el correcto armado de diversas topologías de red, la correcta comunicación entre los distintos host y se pudieron probar todos los servicios dados en la materia. El prototipo se encuentra disponible para ser descargado y utilizado de forma libre y gratuita [20].

7 Conclusiones y futuros trabajos

LPR logró cumplir con la mayoría de los objetivos propuestos. Se considera que el trabajo realizado es de un alto grado de importancia en el ámbito de desarrollo las prácticas de la asignatura Servicios en SO de Redes, siendo una herramienta fundamental en el proceso de aprendizaje de los alumnos. Al ser un desarrollo de software libre, se permite el uso por otra cátedra que lo requiera o necesite, pudiéndolo modificar en base a otros requisitos específicos.

Por otro lado, se está analizando la posibilidad de poder adicionar como SO huésped Microsoft Windows, para poder desarrollar el práctico de SMB y ensayarlos sobre diferentes SO.

Se desea, además, que LPR pueda ser ejecutado desde un medio removible como un USB, para poder almacenar las nuevas configuraciones dado que fue el único objetivo todavía no alcanzado.

References

1. Chowdhury, N. M. K., & Boutaba, R. (2010). A survey of network virtualization. *Computer Networks*, 54(5), 862-876.
2. Fernando Rodríguez-Haro, Felix Freitag, Leandro Navarro, A summary of virtualization techniques, *Procedia Technology*, Volume 3, 2012, Pages 267-272, ISSN 2212-0173

3. A. Souvik Pal , B. Prasant Kumar Pattnaik, Classification of Virtualization Environment for Cloud Computing, Indian Journal of Science and Technology Vol: 6 Issue: 1 January 2013 ISSN:0974-6846
4. Kernel Virtual Machine. web site: <https://www.linux-kvm.org/>
5. XEN Project. web site: <https://www.xenproject.org/>
6. LXC. Linux Containers. web site: <https://linuxcontainers.org/>
7. Docker Enterprise Edition. web site: <https://www.docker.com/>
8. Introducing Execution Drivers and Libcontainer. <http://blog.docker.io/2014/03/docker-0-9-introducing-execution-drivers-and-libcontainer/>
9. Proxmox. <https://www.proxmox.com/>
10. Virtualbox. <https://www.virtualbox.org/>
11. Qemu Process emulator. <http://www.qemu.org/>
12. Packet tracer. <http://www.packettracernetwork.com/>
13. Cloonix Routing and Network Simulation. <http://www.brianlinkletter.com/tag/cloonix/>
14. VNUML. Virtual Network UML https://www.dit.upm.es/vnumlwiki/index.php/Main_Page
15. N.Richter et al., ADIOS web site: <http://os.cqu.edu.au/adios>
16. The User Mode Linux Home-page. web site: <http://user-mode-linux.sourceforge.net/>
17. Adios y networking . Giacchini, M., INFN-LNL, L., & Richtes, I. N. LivEPICS: EPICS Official Distribution Made in LNL.
18. Marmol, V., Jnagal, R., & Hockin, T. (2015). Networking in containers and container clusters. Proceedings of netdev 0.1, February.
19. Docker Networking. <https://docs.docker.com/engine/userguide/networking/>
20. Mediafire. <http://www.mediafire.com/file/52tx4ra8j1pk1ov/livecd.iso>

Visualización del proceso colaborativo como metaconocimiento. Descripción de una estrategia de *mirroring* y sus resultados

Zangara, Ma. Alejandra¹; Sanz, Cecilia^{1, 2}

¹ Instituto de Investigación en Informática LIDI
Facultad de Informática, Universidad Nacional de La Plata
La Plata, Argentina

² Investigador Asociado de la Comisión de Investigaciones Científicas de la Prov. de Buenos Aires
azangara@lidi.info.unlp.edu.ar
csanz@lidi.info.unlp.edu.ar

Abstract. Existe consenso acerca de que las estrategias y el proceso de construcción grupal, constituyen un conocimiento significativo para quienes trabajan colaborativamente. La decisión acerca de qué indicadores seleccionar, y visualizar para el acompañamiento de grupos colaborativos por parte de los docentes, resulta estratégica. También reviste importancia definir cuándo el grupo accede a la información sobre su colaboración: si luego de terminar el proceso, o durante el mismo. En este trabajo se propone y presenta una estrategia de seguimiento de trabajo colaborativo bajo la forma de *mirroring*, que ha sido llevada adelante en una experiencia educativa de postgrado durante 2016. Los resultados preliminares alcanzados parecen confirmar la idea de que el grupo se beneficia a partir del conocimiento sobre cómo se está desarrollando el proceso colaborativo y, aún más, indican que se profundiza la conciencia que cada integrante tiene de su propia tarea y de la de sus compañeros.

Keywords: Indicadores de trabajo colaborativo, Seguimiento del trabajo colaborativo, Estrategia de *mirroring*, Metaconocimiento y trabajo colaborativo.

1 Qué, cuándo y cómo visualizar la colaboración

Hay coincidencia acerca de la importancia que brinda al trabajo grupal y el conocimiento del desempeño por parte de todos los involucrados en el grupo [1], [2], [3], [4]. En [5], los autores describen tres tipos de información que pueden considerarse cruciales para el aprendizaje colaborativo:

- Información pública sobre lo que los integrantes del grupo efectivamente hacen, mencionada también en [6].
- Información cognitiva acerca del *background* de conocimientos previos y/o habilidades de autorregulación para el aprendizaje de cada miembro (también reconocida por [7] y [8] e
- Información social sobre el funcionamiento del grupo, según la percepción de los colaboradores [9].

En [10] se realiza un interesante análisis, acerca de la importancia que tiene, para cada persona en su vida cotidiana, las decisiones y experiencias de los demás. En este trabajo se indica que ya en 1999, se introdujo la idea de que los sistemas informáticos soportan estas actividades, y pueden hacerlas más visibles para ayudar a entender estas relaciones. Se introduce así el concepto de *social translucence* como una propiedad de los sistemas informáticos, que puede ayudar a la comunicación, mostrando aspectos cuantitativos simples de la participación de los usuarios en una tarea compartida. Este enfoque se basa en tres propiedades, presentadas también en [10]. La primera, visibilidad, se refiere a la idea de que los usuarios pueden conocer la información social que se presenta en forma de cifras o gráficos; la segunda, la conciencia, considera el impacto de la información de la actividad de los otros sobre la actividad de cada miembro y las reglas colectivas de trabajo y convivencia; por último, la responsabilidad se refiere a los procesos de autorregulación individual que pueden ocurrir como resultado de la conciencia del usuario sobre sus propias acciones o las de los demás. Estas ideas pioneras fueron continuadas por diversos autores en la construcción de sistemas de visibilización del desempeño de grupos de trabajo: tanto pequeños grupos, como grupos sociales y hasta redes.

En relación con grupos pequeños, se retoma el trabajo de [11], quien cita a [3] y [12], y avanza sobre el concepto de *mirroring*. Lo define como “sistemas que reflejan las interacciones de un grupo” (pág. 119). Estos sistemas, muestran indicadores de interacción que deben ser definidos en función del trabajo y composición del grupo. Existen otros trabajos, como [13], que no sólo ofrecen una descripción de los eventos de interacción, sino que los presentan según indicadores predefinidos y en comparación con estándares esperados. Estos indicadores representan el estado de la interacción, junto a un conjunto de valores y parámetros deseados o esperados. Las visualizaciones pueden ser de diferentes tipos: gráficos de barra y circulares, mapas de temas de un foro (indicando dispersión de tópicos y tareas) e, incluso, desarrollo colectivo de mapas conceptuales [4]. Según la hipótesis de Dimitracopoulou en [13], y en concordancia con los trabajos citados al inicio de este trabajo, esta información podría favorecer tanto el trabajo del grupo en sí, como el de los coordinadores que lo supervisan y orientan. Las estructuras de visualización, con representaciones adecuadas, pueden ayudar a los estudiantes en el desarrollo metacognitivo, y en la regulación de la actividad de colaboración. En cada herramienta que se utilice en la tarea grupal, se deben tomar decisiones respecto de qué información o indicador se muestra, cómo y cuándo. La posibilidad de mostrar desempeño y avance está en estrecha relación con el escenario de la actividad colaborativa: presencial o a distancia y, además, con el manejo del tiempo: sincrónico o asincrónico. Una vez que el grupo comenzó su trabajo colaborativo, es complejo determinar momentos dentro de la actividad para mostrar su avance. En la literatura se encuentra con más frecuencia la visualización de indicadores *ex post facto*, luego de que el trabajo grupal finaliza. Sin embargo, es posible y deseable que las herramientas definan ventanas de tiempo (según Manuel Castells en [14]) para hacer un corte (abstracto) en el trabajo y mostrar su rumbo y relación con lo esperado.

A partir de aquí el artículo se organiza de la siguiente manera: la sección 2, presenta algunos antecedentes de herramientas de seguimiento de la colaboración, la sección 3 describe la estrategia de *mirroring* propuesta y su aplicación, la sección 4 presenta los resultados y la 5 las conclusiones del trabajo.

2 Herramientas de visualización del trabajo colaborativo

En esta sección se clasifican y describen algunas herramientas de seguimiento del trabajo colaborativo para ejemplificar de qué manera soportan esta actividad, qué indicadores muestran y de qué forma realizan las visualizaciones. Estos antecedentes se han utilizado para la propuesta de *mirroring* que se describe en la próxima sección. Las categorías que en las que se las ha agrupado son las tres siguientes:

- Herramientas de espejo (*Mirroring Tools*): esta categoría incluye herramientas que reflejan automáticamente la actividad de los integrantes de un grupo de trabajo. Son visualizaciones gráficas de las acciones de cada uno en cada herramienta.
- Herramientas metacognitivas (*Metacognitive Tools*): esta categoría incluye a la anterior de mirroring y, además, muestra información acerca del desvío del desarrollo de los indicadores respecto de lo deseado. Ejemplos podrían ser: marcada heterogeneidad en la participación de los integrantes del grupo, dispersión de temas, etc.
- Herramientas de guía (*Guiding Systems*): esta categoría incluye a las dos anteriores y agrega un espacio para la guía e intervención del docente.

2.1 Big Five

Son herramientas de visualización que se corresponden con la teoría de grupos Big Five. En [15], se presenta un conjunto de herramientas de visualización específicas del trabajo colaborativo en entornos informáticos que se vinculan con el marco teórico presentado por Albert Bandura [16], dentro de la teoría de la construcción social (*Social Cognitive Theory*, SCT). Este marco identifica las cinco capacidades (*big five*) que definen el trabajo grupal: 1. capacidad de simbolización; 2. capacidad de prefiguración y planificación; 3. aprendizaje vicario; 4. autorregulación; 5. autorreflexión. Un ejemplo de herramientas (del tipo *Metacognitive Tools*) es el Radar de actividad, que consiste en un círculo, que representa el rango de participación de cada miembro en función de un estándar de comparación, que puede ser el promedio del grupo en el pasado, un estándar predefinido o la actividad de un integrante del grupo. Este estándar aparece en el punto medio del círculo (ver Fig.1 A.).

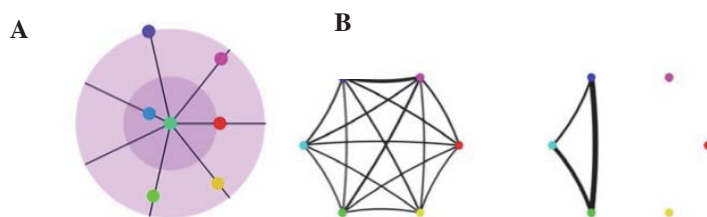


Fig. 1. A. Radar de participación, según *Big Five* y B. Redes de interacción en dos herramientas, con diferente densidad y relaciones (Tomados de [15])

Otra representación está basada en el Análisis de Redes sociales (ARS) y cuantifica las relaciones entre los actores, con el objeto de crear redes gráficas que representen esas relaciones como un todo. Hay tres elementos básicos que conforman una red: actores, límites y relaciones. Los actores están representados con puntos y las relaciones con líneas que los unen (Fig. 1 B.).

2.2 Drew. Dialogical Reasoning Educational Web tool

Según [12] se trata de una herramienta web que muestra gráficamente el mapa de tópicos que se genera en un sistema de argumentación. Se enmarca en el proyecto SCALE de la Comunidad Europea, enfocado al aprendizaje colaborativo para la argumentación a través de internet en la enseñanza secundaria. Sus herramientas están diseñadas para ayudar a los estudiantes a desarrollar, refinar y ampliar su conocimiento argumentativo en un campo determinado. El profesor tiene un rol definido en el software para visualizar lo que sucede, agregar tareas y participantes. Es por ello que se la considera una herramienta dentro de la categoría Guiding Systems.

2.3 iBee. Bulletin Board Enrollee Envisioner

Creado y presentado por [28], iBee es un software bajo el modelo de tablón de anuncios, que funciona como un *plug in* en un entorno virtual de enseñanza y aprendizaje. Sus principales características incluyen: 1.la visualización de la relación entre palabras clave y los alumnos en tiempo real; 2.la visualización de una trayectoria de conversación en un tiempo determinado; 3.la visualización de los últimos niveles de participación de los alumnos y de la frecuencia de uso de palabras clave, y 4.la ubicación de los mensajes según palabras clave, representados con la metáfora de flores y abejas, para que el estudiante pueda acceder a ellas mediante un clic. Puede considerarse a iBee como una herramienta de *mirroring*, acorde a las categorías antes descriptas.

3 Estrategia de *mirroring* implementada

En esta sección se describe la estrategia de visualización tipo *mirroring* que se diseñó e implementó en una actividad colaborativa en un curso de postgrado de la Facultad de Informática de la UNLP. El curso fue el Seminario de “Educación a Distancia” de la Maestría en Tecnología Informática aplicada en Educación. En la cohorte 2016, se trabajó con 11 alumnos. Para esta tarea, se dividió a los participantes en 2 grupos: uno de 5 personas y otro de 6 personas. La acotada cantidad de alumnos y la existencia de sólo dos grupos de trabajo, favoreció el desarrollo de esta estrategia, ya que es necesario revisar en forma detallada todas las actividades individuales y grupales por cada una de las herramientas dispuestas.

3.1 Descripción de la estrategia de *mirroring*

En este seminario, en la octava semana, se lleva adelante una e-actividad de escritura colaborativa, en la que los participantes reciben, en una primera etapa, una consigna individual (que sus compañeros no conocen). En la segunda etapa deben integrar las producciones que desarrollaron a partir de las consignas individuales en un libro, en forma colaborativa. Para esta tarea cuentan con 6 semanas de trabajo, en las que se llevó adelante la estrategia de visualización tipo *mirroring* que se describe a partir de aquí. La primera actividad fue comunicar a los participantes sobre la implementación de esta estrategia, indicando el tipo de trabajo que se llevaría adelante y enfatizando el uso de las herramientas puestas a disposición del grupo para hacer posible el seguimiento *online*. Con esta comunicación inicial se adjuntó un documento con el mapa de indicadores sobre los cuales se focalizaría el seguimiento y el cronograma de presentación de la información, según las etapas de trabajo colaborativo consideradas en la consigna. Los indicadores planificados para llevar adelante la estrategia de *mirroring* propuesta se dividen en individuales y grupales (ver Tabla 2). Los individuales contemplan, desde el punto de vista cuantitativo, la cantidad de mensajes intercambiados con sus compañeros en cada etapa de la resolución de la consigna y la cantidad de mensajes en cada herramienta con la que se trabaja. Desde el punto de vista cualitativo, se clasificaron los mensajes según su contenido en: mensajes orientados a la organización, mensajes orientados al manejo afectivo/motivacional del grupo, mensajes orientados al contenido de la tarea. Los indicadores grupales analizados, se relacionaron con el concepto de interdependencia y se contempló la realización de mapas de tópicos, siguiendo a [4].

Tabla 2. Indicadores del trabajo individual y grupal de *mirroring*

Dimensiones/ Categorías de indicadores	Indicadores
DIMENSION: APORTE INDIVIDUAL DE CADA MIEMBRO	
Aporte de cada miembro al trabajo colaborativo	Análisis cuantitativo <ul style="list-style-type: none"> • Cantidad de mensajes intercambiados con sus compañeros en cada etapa de la resolución de la consigna. • Cantidad de mensajes en cada herramienta Análisis cualitativo <ul style="list-style-type: none"> • Tipos de mensajes según categorías <ul style="list-style-type: none"> - Organización - Contenido - Afectividad
DIMENSION: CONSTRUCCION COLABORATIVA PROPIAMENTE DICHA	
Interdependencia (Concepto grupal)	Mapa conceptual de tópicos. Temas y duración

Para materializar la información del proceso de trabajo colaborativo se utilizó un sitio de *GoogleSites*, dado que los estudiantes ya estaban familiarizados con ella y resultaba de fácil acceso a la información de *mirroring*.

Un tema de suma importancia en el desarrollo de este tipo de estrategias es determinar de qué manera se realizará la visualización de la información. En este caso el formato de presentación de la información según los indicadores de seguimiento fueron: el gráfico de barras que muestra la cantidad y calidad de mensajes según su tipo (contenido, organización, afectividad) por cada integrante del grupo y un mapa de temas o tópicos, indicando tiempo de duración de cada uno, en semanas. Cada tópico fue identificado con un color y se indica en cada uno la o las personas que trabajaron (con sus iniciales). Por otra parte, se eligió un grafo como el de la Fig. 2, que muestra la relación entre tópicos e integrantes del grupo. Esta representación gráfica facilita el análisis de las participaciones e intercambios. Para esta propuesta, los grafos se han construido de acuerdo con las siguientes reglas: a) están conformados por nodos: participantes y tópicos; b) los nodos correspondientes a personas son etiquetados con las iniciales del nombre y los correspondientes a tópicos llevan el nombre del tema; c) todos los nodos se representan con círculos coloreados; d) las aristas que unen a cada persona con un tópico indican la relación del primero con el segundo; e) el tamaño de los nodos aumenta conforme se incrementa la cantidad de aristas que salen o llegan a él y g) el grosor de las aristas aumenta conforme se incrementa la participación de una persona con un tópico. Allí puede observarse qué tan lejos del mapa de tópicos está cada integrante del grupo y visibilizar la significatividad de la participación de cada uno.

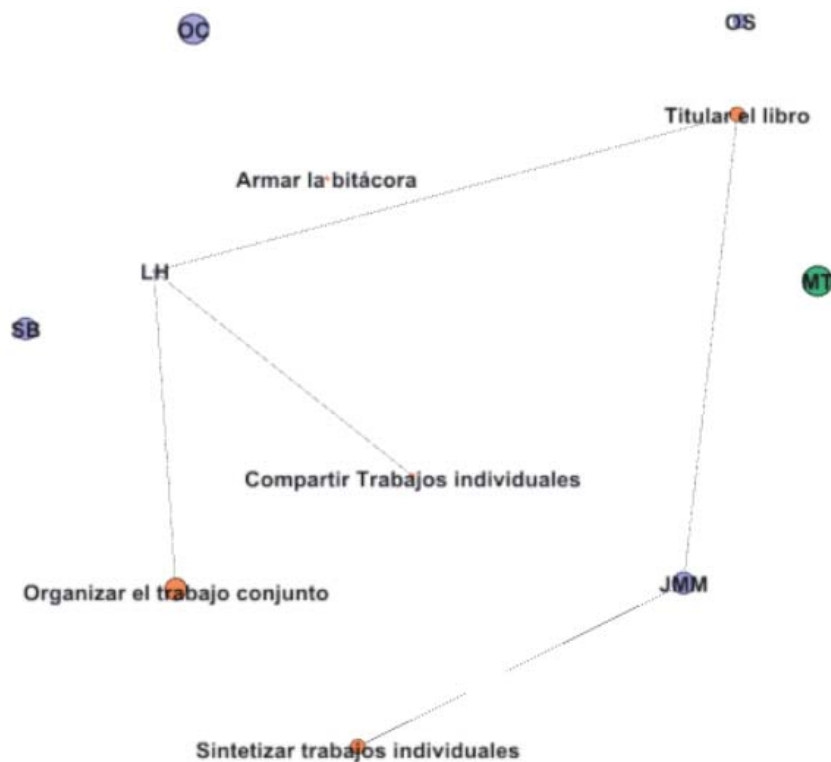


Fig. 2. Grafo de relaciones en la información de *mirroring*

3.2 Indagación del impacto de *mirroring* en el trabajo colaborativo

Luego de las semanas de publicación de la información, en la semana inmediatamente posterior, se administró un instrumento de indagación individual y privada a los integrantes de los grupos acerca de la experiencia de trabajo colaborativo vivenciada y el impacto de la estrategia de *mirroring*. El instrumento estuvo compuesto por 5 secciones: 1. Datos personales, 2. Trabajo Individual, 3. Uso de herramientas, 4. Trabajo grupal y 5. Impacto de la estrategia de *mirroring*. En este trabajo se pone el foco en la indagación realizada a través de la sección 5, que buscó recuperar: la frecuencia de consulta de la información, la utilidad de la información, la atención a los indicadores individuales y/o grupales, la atención a la información del propio grupo versus la del otro grupo, la utilidad de la información según el formato, decisiones que se modificaron en función de la información del *mirroring*, y opinión general sobre cómo influyó esta estrategia en el trabajo grupal.

4 Resultados obtenidos en la experiencia con *mirroring*

En este apartado se comentan los resultados más significativos obtenidos a partir de la indagación presentada en el apartado anterior, y administrada a través de una encuesta *online*.

En cuanto a la **frecuencia de visualización** de la información en el sitio de *GoogleDocs*, se preguntó: ¿Cuántas veces visualizó el sitio de información del proceso colaborativo? Las opciones de respuesta establecían una escala de frecuencia de 1 a 5, siendo 1 nunca y 5 diariamente. La Fig.3 muestra los resultados, que indican **que un 30% de los estudiantes accedían diariamente y un 44% indicó una frecuencia en la escala entre 3 y 4**. Al mismo tiempo se les consultó las razones por las que accedían. Se encontraron respuestas como las siguientes: “sentí curiosidad del tipo de información que se estaba manejando”, “me interesé por saber cómo se mostraba y *qué datos se visualizaban sobre las participaciones de los integrantes del grupo*” y “*me pareció importante saber si lo que se observaba en el trabajo colaborativo de mi grupo estaba reflejado en el sistema*”.

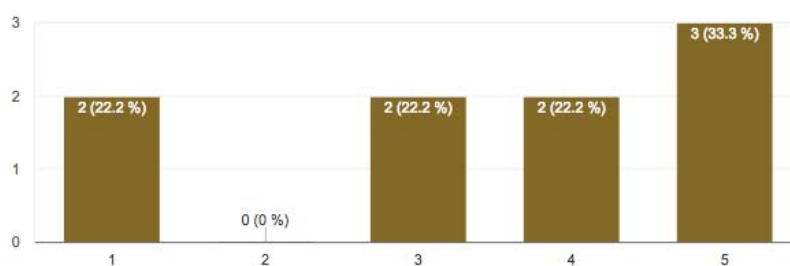


Fig. 3. Frecuencia de acceso al sitio de información de *mirroring*

El ítem respecto al **tipo de información y formato** resultó de interés, ya que ponía en juego los indicadores seleccionados y su potencialidad para comprender, comunicar y hasta mejorar el trabajo colaborativo. La pregunta fue: “¿Qué tipo de información le resultó de más utilidad?”. Las opciones de respuesta indicaban todos los formatos en los que la información de seguimiento había sido formalizada en el sitio web: textos, valores numéricos, tablas de datos, gráfico de barras, grafos de trabajo grupal, imágenes, etc. Las respuestas obtenidas se presentan en la Fig.4.

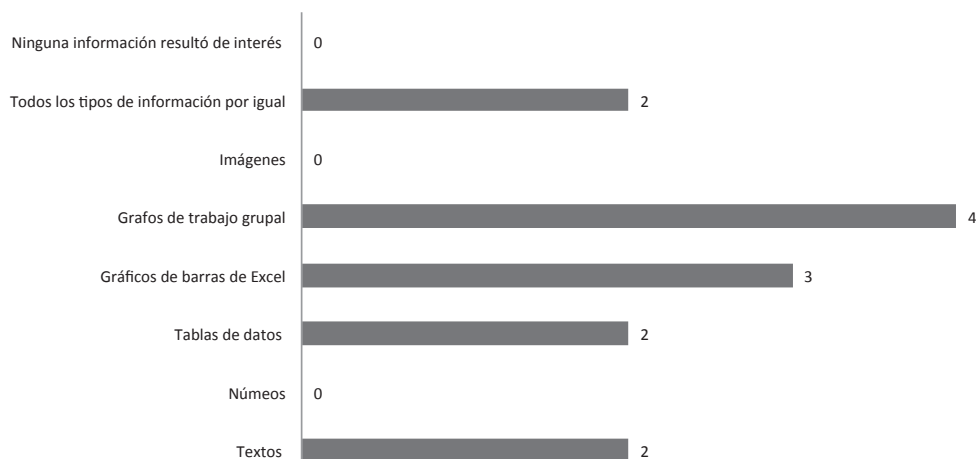


Fig. 4. Acceso a la información según formato

En este ítem podía indicarse más de una opción de respuesta, ya que los tipos de información no resultaban excluyentes. Comenzando por la información que resultó de mayor interés (4 personas) deben resaltarse los **grafos de presentación de la actividad tanto individual como grupal**. En segundo lugar, aparecen los **gráficos de barra que mostraron desempeño individual**: cantidad de mensajes por categoría. Antes de estos gráficos, se incluían las tablas de datos que les habían dado origen. Esta información también fue valorada por los estudiantes. Finalmente, el **valor del texto como elemento de integración, explicación y contextualización** apareció como valioso. En cuanto a las razones de su visualización, se esgrimieron: “*Los gráficos representan los datos en forma clara y precisa*”; “*Personalmente, prefiero analizar gráficos a números*”; “*Prefiero el grafo porque puedo ver la conexión entre los integrantes con las acciones y los tipos de mensajes*”. También se mencionó la tabla de datos: “*Me sirvió el desglose de forma individual de las participaciones y la tabla de datos y sus subdivisiones*”. Se valoró suficientemente el texto y la integración de diferentes tipos de información en el sitio: “*Todos los informes fueron leídos con atención. Los gráficos y tablas y grafos son más fáciles de interpretar, pero siempre viene muy bien la parte de textos*” y “*Cada elemento enriqueció de manera distinta las estadísticas del proceso y fueron complementarios*”.

Para apuntar a la hipótesis inicial, se consultó: **¿Qué decisiones se modificaron en función de la información mostrada en el sitio?** (se dividen las opciones en tres aspectos: afectividad, contenido, organización). Las respuestas fueron se muestran en la Fig.5.



Fig.5. Decisiones tomadas a partir del *mirroring*

En la Fig.5, se observa que las **decisiones de organización** fueron las más impactadas por el conocimiento a partir de la estrategia de *mirroring*. Dentro de las decisiones de organización aparecen dos componentes fundamentales: el manejo del tiempo y la participación equilibrada de todos los integrantes del grupo. Respecto de la participación de todos, pudo seguirse a través del sitio el devenir de esa participación a través del componente de información individual. En cuanto a las **decisiones de contenido**, todas fueron indicadas como impactadas por esta estrategia. Las decisiones en el campo de la **afectividad** impactadas por la información del *mirroring* fueron las de motivación: el conocer qué había hecho cada participante en su dimensión individual y grupal ayudó a dirigir los mensajes en función de conseguir un mayor involucramiento y actividad de todos. Lo que indicaron los estudiantes en referencia a temas de organización fue lo siguiente: “*Me parece que lo principal fue organizar el tiempo de los integrantes para finalizar cuanto antes con las responsabilidades individuales*”; “*Permitió organizarnos teniendo en cuenta el tiempo disponible a fin de poder terminar el trabajo en los tiempos previstos. Además se debió tener en cuenta el elaborar una síntesis del tema de una integrante que había abandonado el curso*”. También sirvió como refuerzo del metaconocimiento acerca de la *performance* individual: “*Al ver los datos, aprecié en real medida lo baja que había sido mi participación en algunos momentos*”.

5 Conclusiones

La idea rectora del trabajo radica en la importancia que reviste la información del proceso de trabajo colaborativo como componente metacognitivo del desempeño, tanto individual como colectivo. Conocer de qué forma trabaja cada individuo y sus compañeros se vislumbra como un aporte de valor, tanto al proceso como al resultado del trabajo de un grupo. Este conocimiento, de característica metacognitiva, resulta de importancia tanto para los integrantes del grupo como para el coordinador, quien debe orientarlos, e intervenir en el caso de que sea necesario. En este artículo se describió una estrategia de *mirroring* que se diseñó e implementó con un grupo de estudiantes de postgrado, como espacio de experimentación metacognitiva. Los resultados obtenidos refuerzan la idea de [13] acerca de la visibilidad, conciencia y responsabilidad que promueve entre los miembros de un grupo el conocimiento de su desempeño. Entre los hallazgos que pudieron observarse en la experiencia, se encuentran los siguientes: a) se modificó la cantidad de mensajes asociados a la organización, ni bien aparecieron los datos sobre tipos de mensajes en el sitio de *mirroring*, el grupo comenzó a comunicarse en términos de planificación y organización de su tarea; b) la mayoría de los participantes monitoreó la información que el sitio mostraba sobre su tarea individual: cada uno se aseguró de que el sitio relevara exactamente lo que había hecho. Incluso, enviaron explicaciones de por qué no habían podido realizar una tarea específica en forma privada. A partir de esta visualización, se observó que cada integrante del grupo estuvo atento a lo que hizo (y no hizo) y a lo que sus compañeros (de su grupo y del otro grupo) produjeron durante las 6 semanas de trabajo. En este sentido, se cree relevante la estrategia implementada ya que reafirman varias de las hipótesis recuperadas en la revisión de antecedentes y abre nuevos caminos de indagación ya que presenta nuevos hallazgos sobre qué tipos de visualización han impactado más en el proceso. Como trabajo futuro se continuarán realizando experiencias que permitan ampliar la cantidad de participantes y tener así resultados más concluyentes.

Referencias

1. Avouris, N., Komis, V., Margaritis, M., & Fidas, C. (2004). ModellingSpace: A tool for synchronous collaborative problem solving. In Proceedings of Int. Conf. on Educational Multimedia & Telecommunications, pp. 381-386.
2. Avouris, N., Margaritis, M., & Komis, V. (2004). Modelling interaction during small-group synchronous problem-solving activities: The Synergo approach. In Proc. of ITS 2004 workshop on designing computational models of collaborative learning interaction, pp13-18.
3. Jermann, P., Soller, A., & Muehlenbrock, M. (2001). From mirroring to guiding: A review of the state of art technology for supporting collaborative learning. In European Conference on Computer-Supported Collaborative Learning EuroCSCL-2001, pp. 324-331.
4. Martínez Maldonado, R. (2014). Analysing, visualising and supporting collaborative learning using interactive tabletops. Doctoral dissertation, The University of Sydney, Australia.
5. Bodemer, D., & Dehler, J. (2011). Group awareness in CSCL environments. Computers in Human Behavior, 27(3), pp. 1043-1045.

6. Janssen, J., Erkens, G., & Kirschner, P. A. (2011). Group awareness tools: It's what you do with it that matters. *Computers in Human Behavior*, 27(3), pp. 1046-1058.
7. Dehler, J., Bodemer, D., Buder, J., & Hesse, F. W. (2011). Guiding knowledge communication in CSCL via group knowledge awareness. *Computers in Human Behavior*, 27(3), 1068-1078.
8. Sangin, M., Molinari, G., Nüssli, M. A., & Dillenbourg, P. (2011). Facilitating peer knowledge modeling: Effects of a knowledge awareness tool on collaborative learning outcomes and processes. *Computers in Human Behavior*, 27(3), 1059-1067.
9. Phielix, C., Prins, F. J., & Kirschner, P. A. (2010). Awareness of group performance in a CSCL-environment: Effects of peer feedback and reflection. *Computers in Human Behavior*, 26(2), 151-16
10. Erickson, T., Kellogg, W. (2000) Social translucence: an approach to designing systems that support social processes. *ACM Transactions on Computer-Human Interaction (TOCHI)* - Special issue on human-computer interaction in the new millennium, Part 1: Volume 7 Issue 1, March 2000. Publisher: ACM
11. Dimitracopoulou, A. (2005). Designing collaborative learning systems: current trends & future research agenda. In *Proceedings of the 2005 conference on Computer support for collaborative learning: learning 2005: the next 10 years!*, pp. 115-124. International Society of the Learning Sciences.
12. Serpaggi, X., Baker, M., Quignard, M., Lund, K., & Séjourné, A. Corbel, A., Jaillon, P., (2003, April). DREW: Un outil Internet pour créer des situations d'apprentissage coopérant. In Desmoulins, Marquet & Bouhineau (Eds.), *EIAH2003 Environnements Informatiques pour l'Apprentissage Humain, Actes de la conférence EIAH*, pp. 109-113.
13. Dimitrakopoulou, A., Petrou, A., Martínez, A., Marcos, J.M., Kollias, V., et al. (2006) State of the art of interaction analysis for Metacognitive Support & Diagnosis. (D31.1.1) EU Sixth Framework Programme priority 2, Information society technology. Network of Exc.
14. Campbell, S. W. (2007). Manuel Castells, Mireia Fernandez-Ardevol, Jack Linchuan Qiu, and Araba Sey: "Mobile Communication and Society: A Global Perspective". *International Journal of Communication*, 1(1), 7.
15. Kay, J., Maisonneuve, N., Yacef, K., & Reimann, P. (2006). The big five and visualisations of team work activity. In *Intelligent tutoring systems*, pp. 197-206. Springer Berlin Heidelberg
16. Bandura, A. (1986). *Social foundations of thought and action*. USA: Prentice Hall.
17. Mochizuki, T., Kato, H., Yaegashi, K., Nagata, T., Nishimori, T., Hisamatsu, S. I., & Suzuki, M. (2005, May). Promotion of self-assessment for learners in online discussion using the visualization software. In *Proceedings of th 2005 conference on Computer support for collaborative learning: learning 2005: the next 10 years!*, pp. 440-449. International Society of the Learning Sciences.

XV Workshop Computación Gráfica, Imágenes y Visualización (WCGIV)

Coordinadores

Castro, Silvia - Guerrero, Roberto - Bría, Oscar

Comité Científico

Abásolo, María José (Argentina)

Baldassarri, Sandra (España)

Collazos Ordóñez, César Alberto (Colombia)

D'Amato, Juan Pablo (Argentina)

Ganuza, María Luján (Argentina)

Gazcón, Nicolás (Argentina)

Gröller, Eduard (Austria)

Larese, Mónica (Argentina)

Larrea, Martín (Argentina)

Lecumberry, Federico (Uruguay)

Manresa-Yee, Cristina (España)

Meschino, Gustavo (Argentina)

Pina, Alfredo (España)

Simari, Patricio (USA)

Urribarri, Dana (Argentina)

A tutorial on the implementations of linear image filters in CPU and GPU

Alvaro Pardo
apardo@ucu.edu.uy

Facultad de Ingeniería y Tecnologías
Universidad Católica del Uruguay

Abstract. This article presents an overview of the implementation of linear image filters in CPU and GPU. The main goal is to present a self contained discussion of different implementations and their background using tools from digital signal processing. First, using signal processing tools, we discuss different algorithms and estimate their computational cost. Then, we discuss the implementation of these filters in CPU and GPU. It is very common to find in the literature that GPUs can easily reduce computational times in many algorithms (straightforward implementations). In this work we show that GPU implementations not always reduce the computational time but also not all algorithms are suited for GPUs. We believe this is a review that can help researchers and students working in this area. Although the experimental results are not meant to show which is the best implementation (in terms of running time), the main results can be extrapolated to CPUs and GPUs of different capabilities.

1 Introduction to Linear Filtering

Image filtering is one of the most studied problems in the image processing community. Image smoothing, sharpening, feature detection and edge detection are some of the applications of image filtering. In the literature we can find two broad categories of image filters: linear and non linear. More recently, non local methods attracted the attention of researchers in the area. In fact, several of the state of the art algorithms are both non local and non linear (see [4] for more details). In this tutorial we will focus on the analysis and implementation, both in CPU and GPU, of linear filtering methods. The approach will be strongly connected to the theory of linear systems and digital signal processing. We refer the interested reader to [9] and [1] for further details on these areas. First we recall that a filter, or system, that takes an input image to produce and output one, is said to be linear if for all linear combinations of inputs produce a linear combination of outputs with the same weighting coefficients. Before analyzing linear image filter using tools from linear systems we will describe linear image filters in their most basic form using sliding windows (convolution masks).

We start with a simple linear averaging filter in which each pixel $x = (i, j)$ of the output image is computed as the average of all pixels in a 3×3 window centered at x in the input image. Processing the whole image can be expressed

with a sliding window algorithm. Given the pixel x the average filter can be implemented moving a 3×3 window with weights $1/9$ across the input image. Mathematically this can be formulated as:

$$g(i, j) = \sum_{i'=-1}^1 \sum_{j'=-1}^1 w(i', j') f(i + i', j + j') \quad (1)$$

where $f(.,.)$ and $g(.,.)$ are the input and output images and $w(.,.)$ is the window containing the filter weights. In the previous example of linear averaging $w(m, n) = 1/9$ for all (m, n) . Changing the values of $w(m, n)$ different filters can be obtained. Inspecting equation (1) we can see that is very similar to a two dimensional convolution. Recalling the theory of linear systems we know that the output of a linear and invariant system can be obtained convolving the input f with the impulse response of the system h : $g = f * h$ ¹. The impulse response of an image filter can be obtained as the output of the filter when the input image is a discrete impulse. If \mathcal{N} is a neighborhood of the same size of the sliding window centered at pixel (i, j) and $h(.,.)$ is the impulse response of the filter, equation (1) can be rewritten as a discrete convolution:

$$g(i, j) = \sum_{(m,n) \in \mathcal{N}} h(i - m, j - n) f(m, n). \quad (2)$$

The main difference between equations (1) and (2) is the range of indexes (i, j) and (m, n) . Both formulations convey useful information; the first one is more suited for interpretation while the second one enables us to connect linear image filtering with convolution and the frequency response of the filter.

The equation (2) can be reformulated interchanging the role of h and f ; instead of moving h across f we move f and leave h fixed. To do that we center h around the origin and extend it filling it with zeros outside the original window range and extend the input image outside the original range $[0, M - 1] \times [0, N - 1]$. In this way the equation (2) turns into: $g(i, j) = \sum_{(m,n)} h(m, n) f(i - m, j - n)$. In the next section we will use this formulation to obtain the frequency response of linear image filters.

1.1 Z Transform and Frequency Response

The Z transform is a very useful tool in the context of linear and invariant systems, signal processing and discrete control theory [1]. To justify the Z transform we will first deduce it starting from the convolution product. If we consider an input image $f(i, j) = z_x^i z_y^j$ with z_x and z_y arbitrary complex numbers, the output signal is: $z_x^i z_y^j \sum_m \sum_n h(m, n) z_x^{-m} z_y^{-n}$. This simple result shows that $z_x^i z_y^j$ are eigenfunctions of linear and invariant filters with corresponding eigenvalues $H(z_x, z_y) = \sum_m \sum_n h(m, n) z_x^{-m} z_y^{-n}$. This expression is known as the Z transform of $h(m, n)$ or transfer function of the filter. One of the most important properties of the Z transform states that the convolution of two signals is the

¹ The filter impulse response h is sometimes referred as filter kernel.

product of their respective Z transforms, see [1] for details. Hence, if f and g are the input and output signals related by $g = f * h$, their relationship in the Z space is: $G(z_x, z_y) = H(z_x, z_y).F(z_x, z_y)$. If we evaluate the Z transform in the unit sphere we obtain the Fourier transform. The Fourier transform of h is $H(\theta_x, \theta_y) = \sum_{(m,n)} h(m, n) \exp(-j(\theta_x m + \theta_y n))$, the frequency response of the filter. In the following section we will use the Z transform to study image filters and propose alternative formulations for some of them. The interested reader can obtain more information about the Z transform in [1].

2 Implementation of linear image filters

In this section we discuss the implementation of linear image filters using the tools presented in previous sections. We will describe the implementation details and address the computational complexity of each approach. One of the goals of the following analysis is to determine which are the best implementations given the filter characteristics (window size, symmetry, etc.). First we show how to implement the filters in their traditional sequential form used for CPU algorithms. Later on, we study the parallel versions of the same algorithms suited to GPU architectures.

2.1 Convolution

The implementation of equations (1) and (2) is straightforward. Basically, the idea is to visit every pixel in the image and apply the corresponding equations. Typically, the sliding window approach, equation (1), is the first option since is very easy to understand and code. The trickiest part of the implementation is the management of the border conditions. That is, how to process pixels close to the image borders where part of the filter window falls out of the image.

Convolution computational cost To conclude the description of this method we will estimate the number of operations needed to implement it. To simplify the estimation we will assume that the image size is $N \times N$ and the window filter size is $(2W+1) \times (2W+1)$. It can be easily seen that each pixel demands $(2W+1)^2$ operations and therefore the total number of operations is of order $N^2(2W+1)^2$. To avoid confusions we distinguish computational cost from computational time.

2.2 Separable convolution

A filter is said to be separable if its kernel can be broken into two one-dimensional vectors that multiplied give the original filter response: $w(i, j) = u(i)v(j)$. The convolution of an image with a separable kernel can be implemented with two one-dimensional convolutions. First, each row in the image is convolved with v , then the result is processed across columns convolving it with u . The mathematical justification can be easily obtained substituting $w(i, j) = u(i)v(j)$ into (2): $g(i, j) = \sum_m u(m) (\sum_n v(n) f(i - m, j - n))$.

Separable convolution computational cost Convolution with separable kernels requires $N(N(2W + 1)) + N(N(2W + 1)) = 2N^2(2W + 1)$ operations. The reduction in the number of operations is $(2W + 1)/2$ compared to the traditional two-dimensional convolution. Therefore, the larger the kernel the better speed-up can be obtained with this approach.

2.3 Special case: Box Filtering

The box filter is an average filter with uniform weights; the output at pixel $x = (i, j)$ is the average of all pixels in the filtering window:

$$g(i, j) = \sum_{i'=-W}^W \sum_{j'=-W}^W \frac{1}{(2W + 1)^2} f(i + i', j + j')$$

The beauty of this filter is that it can be implemented using integral images to reduce the number of operations. The integral image of an image f at pixel (i, j) , denoted as $Sf(i, j)$, is the sum of all elements in the rectangular region with upper-left and lower-right vertices $(0, 0)$ and (i, j) : $Sf(i, j) = \sum_{i' \leq i, j' \leq j} f(i', j')$. Given the pixel (i, j) , the output of the box filter can be obtained using the integral image as follows: sum the pixels in the square region defined by the points $(i - W, j - W)$ and $(i + W, j + W)$ and divide it by the number of pixels in the window (recall that the filter window is $[-W, W] \times [-W, W]$). This can be easily implemented with integral images [3]:

$$g(i, j) = \frac{Sf(i + W, j + W) - Sf(i + W, j - W) - Sf(i - W, j + W) + Sf(i - W, j - W)}{(2W + 1)^2} \quad (3)$$

The computation of the box filter implies two steps: the computation of the integral image and, after that, the computation of the output using equation (3). This formulation is especially useful when we need to filter the image at different scales, i.e. with different filter sizes, because the first step can be re-utilized and regardless the filter size, the second step has always the same cost in terms of operations. In order to estimate the cost, in terms of operations, we have to estimate the cost of computing the integral image and the actual filter, from equation (3). It can be easily seen that the computation of the integral image requires N^2 operations. On the other hand, the filtering requires only four operations per pixel so the total number of operations to apply the filter is $4N^2$. Therefore, the total number of operations for the case of the box filter using integral images is $5N^2$. As said before, this does not depend on the filter size. This is why integral images are very attractive to filter the same image at different scales (the work of Viola and Jones popularized this idea [10]).

Moving Average Filter Moving Average Filters are in fact an implementation of the Box Filter. Using the two-dimensional Z transform it can be shown that the relationship between the Z transforms of input and output images is:

$$G(z_x, z_y) = \frac{F(z_x, z_y)}{(2W + 1)^2} \sum_{i=-W}^W \sum_{j=-W}^W z_x^i z_y^j = \frac{F(z_x, z_y)}{(2W + 1)^2} \frac{z_x^{W+1} - z_x^{-W}}{1 - z_x} \frac{z_y^{W+1} - z_y^{-W}}{1 - z_y}$$

Taking the inverse Z transform the previous equation gives:

$$g(i+1, j+1) = g(i+1, j) + g(i, j+1) - g(i, j) + \frac{(f(i-W, j-W) - f(i+W+1, j-W) - f(i-W, j+W+1) + f(i+W+1, j+W+1))}{(2W+1)^2}.$$

The computation cost to process an $N \times N$ image in this case is $7N^2$. Comparing to the separable alternative there is no big difference in this case in terms of number of operations.

3 Introduction to GPU programming using CUDA

In this section we review the main concepts behind GPUs and the parallel implementation of algorithms using this technology. In particular we will use CUDA. For a more detailed presentation we refer to [8].

GPUs are highly parallel processors with many cores and the ability to run multiple threads that provide high performance computing. The architecture of the GPUs, traditionally optimized for graphic applications, has some limitations; less cache and flow control limitations. GPUs provide advantages in applications where the same computations can be applied in parallel to many data elements. However, memory transfers from main memory to device memory (GPU) have to be considered. A GPU implementation pays off if its computation cost is higher than memory access cost. To process data with an algorithm implemented in the GPU the data must be transferred from main memory to the device, process it in the device and transfer back to main memory. Therefore, the computation cost must be higher enough to pay the overhead introduced by memory transfers. CUDA (Compute Unified Device Architecture) is a programming language by nVidia that allows programming the GPUs abstracting the code from the actual hardware details (OpenCL is another option). Provides the user a high level interface so that he can take advantage of the capabilities of GPUs without having to directly handle the hardware. The CUDA programming model allows the user to use GPU capabilities from a simple interface similar to C language (C language extension). CUDA proposes three abstractions: a hierarchy of thread groups, shared memory and synchronization [6]. These abstractions provide an easy way to understand and handle parallelization. These abstractions are designed so the actual implementation does not need to know the details of the hardware (number of cores, etc.) (see Figure 1). The idea is to divide the problem in blocks of threads. Then each block of threads works cooperatively to solve the problem. In this way scalability is easily achieved, see Figure 1. To understand image filtering implementations on GPUs, and to make this article self-contained, we first review the basics using simple examples of vectors and matrices addition (This section is based on [6]).

CUDA Kernels A kernel is a function that runs N times in parallel on N different threads. In the following code a kernel is used to sum in parallel to vectors of dimension N using N threads.

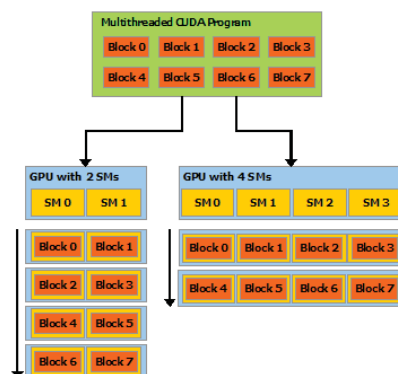


Fig. 1. From [6]: A multithreaded program is partitioned into blocks of threads that execute independently from each other, so that a GPU with more multiprocessors will automatically execute the program in less time than a GPU with fewer multiprocessors.

```
// Kernel definition
--global-- void VecAdd(float* A, float* B, float* C){
    int i = threadIdx.x;
    C[i] = A[i] + B[i];
}
// Kernel invocation with N threads
VecAdd<<<1, N>>>(A, B, C);
```

CUDA threads are three dimensional vectors which enable processing blocks up to dimensions three. The following example shows how to add two matrices.

```
// Kernel definition
--global-- void MatAdd(float A[N][N],
                      float B[N][N], float C[N][N]){
    int i = threadIdx.x;
    int j = threadIdx.y;
    C[i][j] = A[i][j] + B[i][j];
}
// Kernel with one block of NxNx1 threads
int numBlocks = 1;
dim3 threadsPerBlock(N, N);
MatAdd<<<numBlocks, threadsPerBlock>>>(A, B, C);
```

Since the number of threads is bounded (in actual GPUs by 1024) and all threads of a block reside in the same core, when dealing with large vectors or matrices, the problem must be organized into several blocks. CUDA blocks can be organized into one, two or three dimensional grids. In this way, the problem can be organized into a number of blocks per grid and threads per block. This allows for flexibility to organize the computations. The following code shows how to add two matrices organizing the computation into blocks of size 16×16 . The matrices are divided with a tiling of 16×16 . Since there is no guarantee that

N is multiple of 16, inside the kernel we must verify that the pixel (i, j) resided inside the matrices. The choice of blocks of size 16×16 can be modified to take advantage of the GPU capabilities.

```
// Kernel definition
--global-- void MatAdd(float A[N][N],
                      float B[N][N], float C[N][N]){
    int i = blockIdx.x * blockDim.x + threadIdx.x;
    int j = blockIdx.y * blockDim.y + threadIdx.y;
    if (i < N && j < N)
        C[i][j] = A[i][j] + B[i][j];
}
// Kernel invocation
dim3 threadsPerBlock(16, 16);
dim3 numBlocks(N / threadsPerBlock.x, N / threadsPerBlock.y);
MatAdd<<<numBlocks, threadsPerBlock>>>(A, B, C);
```

4 Image Filtering in GPUs

This section discussed the GPU implementation using CUDA of the image filters presented above. We will use the basic notions of GPU programming with CUDA introduced in previous section.

4.1 Convolution

The code below is a direct implementation of 3×3 linear image filter. The code is very similar to the one in C used for the CPU. The main difference is that in this case the pixel indices i and j are obtained from the grid and block organization of the computation on the device. This implementation follows the same philosophy of the code seen before to add two matrices. The next snippet of code shows how to organize the memory allocation and kernel invocation.

```
// Kernel definition
--global-- void filter(float *f, float* g, int rows, int cols){
    int i = blockIdx.x * blockDim.x + threadIdx.x;
    int j = blockIdx.y * blockDim.y + threadIdx.y;
    ...
    if ( i >= w && j >= w && i < cols - w && j < cols - w ){
        float h[3][3] = {...};
        float sum = 0;
        for (int ii = -w; ii <= w; ii++){
            for (int jj = -w; jj <= w; jj++){
                float fij = f[ (i+ii)*rows + (j+jj) ];
                sum += h[ii+w][jj+w] * fij;
            }
        }
        // store output. sumh is the sum of all weights h[][]
        g[i*cols + j] = sum/sumh;
        ...
    }
```

```

// HOST memory
float *f = new float [arraySize];
float *g = new float [arraySize];
// DEVICE memory (GPU)
cudaMalloc((void*)&dev_f, size * sizeof(float));
cudaMalloc((void*)&dev_g, size * sizeof(float));
// Copy image f from DEVICE to HOST
cudaMemcpy2D(dev_f, pitch, f, size*sizeof(float)*fils,
             cudaMemcpyHostToDevice);
// Kernel invocation
dim3 block(16, 16);
dim3 grid(fils/16, cols/16);
filter<<<grid,block>>>(dev_f, dev_g, fils, cols);
// Copy result from DEVICE to HOST
cudaMemcpy(g, dev_g, size*sizeof(float), cudaMemcpyDeviceToHost);

```

The first two columns of Table 1 show the results of executing the CPU and GPU version of the sliding window (convolution) method. For small images, the overhead time of memory transfers is higher than the computation cost and therefore the GPU implementation does not give any speed up. For larger images the GPU is an alternative to speed up linear image filtering (the speed ups factors are shown in parenthesis). Although the breakpoint of when GPU outperforms CPU can depend on the hardware (CPU & GPU), the main result holds valid; for small images and filters of low computational demands (small windows) GPUs are not faster than CPUs due to the memory transfer overheads.

CUDA Texture Memory Texture memory is a read-only memory that can be used to improve performance. Optimizing memory access in the GPU provides benefits in terms of computational time [8]. Texture memory is one of the most basic improvements that can be added to the code of the image filter. The only modification in the kernel code is the access to pixel data $f[i][j]$ using $\text{float } f_{ij} = \text{tex2D}(\text{texf}, i+ii, j+jj)$ where texf is a texture connected to array f . In Table 1 we can see that the use of texture memory reduces the running time. Once again, we observe that the differences appear for large images.

4.2 Separable Convolution

The GPU implementation of a separable filter needs two kernels; one to filter by rows and the other by columns. Separable convolution can provide speedups around 3 times². According to [7] the use of texture memory and other memory optimizations an additional speedup of factor 2 can be obtained (see [7] for details).

² In <https://blog.kevinlin.info/nvidia-cuda-gpu-computing-and-computer-vision/> there is a detailed analysis of the separable implementation

5 Results and Discussion

CPU versus GPU The first result is that when dealing with small images GPU does not provide advantages over CPU (see columns 1 and 2 from Table 1. The actual size of the image where one implementation outperforms the other depends on hardware features. However, the observation holds valid and, as we said before, is due to memory transfers from host to device and backwards. Furthermore, on the CPU side there is still room for improvements using, for example, parallelization with multicores. Therefore, the GPU implementation payoff for large images on when additional operations will be performed in the GPU with the same data. That is, when other processes will be applied to the same image. In this case, the data transfer cost is shared among several process and makes GPU more attractive. Since in many areas we are seeing an increasing use of high definition images (HDTV, Ultra HDTV), we can expect to have to process large images and therefore GPUs are obviously a good alternative. This is the case of mobile platforms which include a GPU to handle image and video data.

Algorithms Now we discuss the impact of the algorithms that reduce the computation cost. First we reviewed separable convolution which is a case of interest since many traditional image filters are separable (Gaussian filters, Sobel filters for edge detection, etc.). From the data in Table 1 we can observe and speedup of $\times 1.5$ for a filter of size 3×3 . This factor agrees with the estimation in Section 2.2. In this case $W = 1$ so the theoretical computation cost reduction is $3/2$. As we mentioned in Section 4.2 GPU implementation of separable convolution gives an additional speed up (see [7]). To illustrate the benefits of applying the correct algorithms to decrease the computational cost and improve running times, we discussed Box Filters in Section 2.3. Box filters are a special case of linear image filters with many real applications due to their reduced computational cost [9]. The last column of Table 1 shows the obtained running times for a CPU implementation. We must be careful when directly comparing this implementation with the others since this is a special filter (with uniform weights). If we assume that all algorithms implement the same box filter, using a uniform kernel, we can see that the implementation of the box filter (MAF) outperforms all other algorithms. Hence, if the application allows a box image filter then the MAF is a simple and computational efficient algorithm (there is no need for a GPU implementation). Finally, if we need a multiscale version of the box filter, the use of integral images is a good solution. In [5,2] the authors compare GPU and CPU implementation of integral images.

6 Conclusions

In this paper we presented an overview of linear image filtering, its basic results based on the theory of linear and invariant systems, and different algorithms to implement the filter. We reviewed different algorithms to reduce the computational cost and discussed their CPU and GPU implementations. We discussed

pros and cons of algorithms and their implementations. Based on the results presented in in Table 1 we can see that image size must be considered to select the most suited implementation. This paper was intended to understand the basics behind linear image filtering using CPU and GPUs. In real applications, libraries such as NPP (nVidia Performance Primitives) or ArrayFire to name two, must be considered.

N	CPU (1)	GPU (2)	GPU texture (3)	CPU sep. (4)	MAF (5)
512	10	65 (x0.15)	65 (x0.15)	7 (x1.42)	1
1024	42	76 (x0.55)	75 (x0.55)	25 (x1.68)	4
2048	165	105 (x1.58)	91 (x1.83)	105 (x1.58)	17
4096	660	202 (x3.27)	162 (x4.07)	440 (x1.50)	66

Table 1. Running times in msec for a 3×3 filter. GeForce GT 430 (96 cores, 1400 Mhz). Intel i7-2600 3.4 GHz, 16 GB RAM, Windows 7 64 bits. (1) Standard CPU. (2) Direct GPU. (3) Direct GPU using texture memory. (4) CPU separable convolution. (5) CPU implementation of MAF.

References

1. V Oppenheim Alan, W Schafer Ronald, and RB John. *Discrete-time signal processing*. 1989.
2. Berkin Bilgic, Berthold KP Horn, and Ichiro Masaki. Efficient integral image computation on the gpu. In *Intelligent Vehicles Symposium (IV), 2010 IEEE*, pages 528–533. IEEE, 2010.
3. Scott Krig. *Computer Vision Metrics: Textbook Edition*. Springer, 2016.
4. Peyman Milanfar. A tour of modern image filtering: New insights and methods, both practical and theoretical. *IEEE Signal Processing Magazine*, 30(1):106–128, 2013.
5. Diego Nehab, André Maximo, Rodolfo S Lima, and Hugues Hoppe. Gpu-efficient recursive filtering and summed-area tables. *ACM Transactions on Graphics (TOG)*, 30(6):176, 2011.
6. nVidia. *Cuda c programming guide*, 2017.
7. Victor Podlozhnyuk. Image convolution with cuda. *NVIDIA Corporation white paper, June, 2007*(3), 2007.
8. Jason Sanders and Edward Kandrot. *CUDA by Example: An Introduction to General-Purpose GPU Programming, Portable Documents*. Addison-Wesley Professional, 2010.
9. Steven W Smith et al. *The scientist and engineer’s guide to digital signal processing*. California Technical Pub. San Diego, 1997.
10. Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I–I. IEEE, 2001.

An Open-Source Spanish Video Game as a Case Study for the Development of an Interface for Users with Visual Impairment

Estefanía Güimil, Yesica Sacristán, Dana K. Urribarri, Martín L. Larrea

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, Buenos Aires, Argentina
estefaniaguimil@yahoo.com.ar, yesi.sacristan@gmail.com,
dku@cs.uns.edu.ar, mll@cs.uns.edu.ar

Abstract. With the growth of information provided through computer systems, it is fundamental and legally necessary, that everyone have equal access to the information in order to live fairly. For the last few years there have been a positive trend in the software industry to include people with disabilities as their target users. As a consequence of this, access for visually impaired users to software is improving. This brings new challenges from the Human Computer Interaction viewpoint. This paper describes the design, development and evaluation of an open source video game for visually impaired users in Spanish. This game was conceived as a case study for the Human Computer Interaction theory applied to those with visual problems.

Keywords: human computer interaction, video games, visually impaired

1 Introduction

Human Computer Interaction (HCI) is the field of study that involves the design, implementation and use of computer technology, emphasizing the interfaces between humans and computers ([6]). Every time a human uses a computer, there is a dialog between them. This dialog is the interaction, and what the computer presents to the user is the interface. There are many ways in which a human can interact with a computer; and the interface between them is crucial to facilitating this interaction. Nowadays the prevalent user interface is the graphical user interface (GUI). Desktop applications, web pages, web apps, mobile applications, and general electronic devices make use of some type of GUI. Our perception of the world comes, mainly, from our ability to see it. But, how do we perceive the world when our vision is disabled or heavily reduce? More important to us, how can a user interact with a computer when a GUI is not an option because the user is visually impaired? Designing human computer interactions for visually impaired users is a great challenge. Without the sight sense, other senses must

be used as input channels for the human, hearing being the main one. Video games are playing a large and growing part in the leisure activities of the public. In Argentina, the video game industry has a US\$ 30 millions total annual turnover, with 95% of its local production for the export market¹. As their popularity increases, the exclusion of visually impaired people from these activities is becoming more of an issue. In this work we present the development of a video game for the visually impaired users. In particular, the development of a special interface for this type of users and the results from our users evaluations. Our work allows disabled users to enjoy a video game as an individual activity or with the companion of their relatives as a family activity. In the remaining parts of this article, we first review some basic concepts of HCI and visual impairment. Section 3 provides some background on previous works. Then, in Section 4, we describe our proposal for a video game for the visually impaired users; followed by implementation details and usability results in Section 5 and Section 6. We conclude in Section 7 with a brief discussion on limitations and advantages of our proposal and future work.

2 Background

2.1 Human Computer Interaction

Human Computer Interaction (HCI) can be defined in many possible ways; from [7] in 1992 “*HCI is a discipline concerned with the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them*” to [5] in 2017 “*HCI is the intersection of the cultural, the social, the cognitive, and the aesthetic with computing and information technology. It encompasses a huge range of issues, theories, technologies, designs, tools, environments, and human experiences in knowledge work, recreation and leisure activity, teaching and learning, and the potpourri of everyday life*”.

Although its definition has evolved over the years, in its core HCI aims to support people’s performance and experiences by optimizing human-system interactions. An important application of HCI is the design and evaluation of interactive computing technologies across different domains to ensure their usability. The main approach for achieving system usability in HCI field is user-centered design.

2.2 Visual Impairment

Worldwide 285 million people are estimated to be visually impaired, with 39 million blind and 246 with low vision. About 90% of the world’s visually impaired live in low-income settings. According to the World Health Organization, there are 4 levels of visual function:

¹ <http://adva.com.ar/en/acerca/>

- normal vision
- moderate visual impairment
- severe visual impairment
- blindness

Moderate visual impairment combined with severe visual impairment are grouped under the term *low vision*: low vision taken together with blindness represents all visual impairment.

Today, interactive system usage is predominantly based on graphical interaction, where the visual presentation of information is essential for input, hand-eye coordination when using a computer mouse, and output, seeing the information on a screen. This can create difficulties for visually impaired users [14], both at an individual level when interacting with a system, and also when collaborating with other users.

When making alternative non-visual presentations for blind computer users, touch and hearing are the most obvious candidates for presenting information. With today's interaction technology, the combination of tactile and auditory information is the approach commonly available for blind users.

2.3 Video Games

Video games are a form of software and thus an obvious object of study in HCI. Interaction with video games differs from the usual understanding of HCI, because people play video games rather than use them. They are an extremely influential form of software ([1]). They earn enormous amounts of money, generate heated controversy and debate, and their players pour huge amounts of time and effort into them: they are permeating everyday life. This level of popular importance has led to a steadily increasing interest from the academic community in understanding how games work, what they do, and what they could do.

Video games have great positive potential in addition to their entertainment value. There has been considerable success ([4]) using games to address a specific problem or to teach a certain skill. They can clearly consume the attention of children and adolescents. They have the capacity to engage children in learning experiences: just by watching the children it becomes very clear that they prefer this type of approach to learning.

3 Previous Work

What follows is a list of video games that were design for the visually impaired and are reviewed in the context of our work. The first two cases are the ones that most resemble our proposal. The main questions that we ask for each case are, *Is it a free game?*, *Is it an open source game?* and *Is it available in Spanish?*

Quacky's Quest ([11]) is a video game designed by an eleven-year-old boy for his blind grandmother. He designed and implemented the game by using the free starter version of design app called GameMaker ([9]). In it, the user plays as a

duck where he weaves through a series of mazes to find a Golden Egg. Sound cues help to find the way. If the player picks up gems, he hears a cash register sound. If he hits a wall, a deep unpleasant noise is played. If the user goes the wrong way down a passage, a spider noises is heard. If he goes too far down a passage, he set off dynamite. The game is not currently available for download. The only information available about it is on web pages that talk about it or a video filmed by its creator Dylan Viale. The game doesn't have any menu or tutorial to introduce its functionality. Although these are drawbacks to the game, it is important to remark that is a very impressive work for an eleven-year-old boy.

The Explorer and the Mystery of the Diamond Scarab ([13]) is a Nintendo Wii Game for visually impaired and fully sighted children. In this game the user plays as Ben the Archaeologist who is searching for the Temple of the Diamond Scarab. Once he is in the temple he meets Tiri, an enchanting Egyptian princess. Going through a maze of underground passages they look for ancient Egyptian treasures and confront various challenges. The game is played on a Balance Board, a standard Wii accessory. With this the user can move through the maze, find treasures and enter challenging situations. The game feedback is provided by different sounds that are played based on the user actions. The game was developed with an User Center Design approach, where several visually impaired children were involved in its design and testing. The game is only available in Duch or English. Because is design for the Wii console, it's not open source neither free to play.

Zomblind ([2]) is a Spanish video game designed by Antonio Fernández, a young researcher from Universidad de Granada. In this game, the player must survive a zombie apocalypses by using a smart phone as a gun. Unlike previous examples, this game does not have a GUI. All feedback is sound based and all inputs is done by the smart phone sensors. It was only available for Android phones and its source code is publicly available. Unfortunately its latest update is from 2013 and is not longer available in Android's store.

AudioDisc [8] is an ability game created in Spanish by Javier Mairena. It designed for visually impaired and fully sighted people, because it has a GUI and is also sound based. It is a two-player game where each one must receive and throw a disk trying to be faster than the opponent; the opponent can be the computer. Before throwing the disk, the player can choose a direction using the keyboard. Different sounds indicate different throwing directions: a high-pitched tone indicates upward, a medium-pitched tone indicates toward the middle, and a low-pitched tone indicates downward. This sound allows the opponent to recognize where the disk is going. It also uses stereo sound to indicate position in the screen. A voice-over narrator gives instruction to select between the different options of the game. The game was originally available for several Operating Systems, now it is possible to download only a Windows version and an Adobe Flash one. Moreover, the last update for the games is from 2008. Even the game is a free software, there is no source code available for download.

Terraformers [12] was a successful attempt to create an environment in which the same information is accessible via realistic 3D graphics and via audio. With

a complex set of sounds and voice menus, the player is able to know where are the different objects (doors, walls, etc) in the scene and recognize them. It includes a high-contrast mode for low-vision gamers and a no-graphics mode for blind gamers. The plot of the game takes place in the future. Rebel robots not only captured the professor but also dismantled the computer that the professor needs to defeat the rebels. The goal is to find the pieces of the computer and free the professor. Terraformers is now a free Windows game, although at the beginning it was a commercial game. However, the last update of the game is from 2003. The source code of the game is not currently available and the game is in English.

Open Field Echo Sounder [3] is a payed application for Smartphones which allows blind children to play outdoors in open field. The gamer uses the GPS and sounds through headphones to find virtual objects. There is very few information available about the game online, its source code was not found and the information available is in English.

Although there are games for visually impaired users in Spanish ([2, 8]), the majority of the games are in English. In relation with the availability of the games, several are free-downloadable games but out-of-dated or not open-source ones. Some founded games require special hardware, like a game console ([13]) or a smartphone ([2, 3]). In both cases, a special hardware means an extra cost for the user.

4 Our proposal

The goal in creating this game arises from the real need of a particular user who is blind and does not have games according to his needs. Currently he plays with a SEGA console, and he adapts to the video games since they are not designed for his characteristics. The only interaction he has with these types of games is through sounds. However, sounds are an additional effect that complements the visual interface of the game and are not specifically designed to guide the player. On the other hand, the user does not read Braille so we can not use this resource. We approach this enterprise with an user-center design strategy. After two meetings with the user we detected the following functional and non-functional requirements:

- Game's instruction must be available at all time during the game.
- Game's instruction must be delivered by using a clear Spanish voice.
- At any time during the game, the user can exit the game by pressing the ESC key.
- The game must run on a Core i3, 2GB of ram PC with Windows 7 OS. Which is the PC the user has.
- The game should not require any installation. It should be a single portable file in order to facilitate its movement from PC to PC.

Based on this situation, we propose to develop a free open-source video game design for a visually impaired user. The proposal is a simple game with clear

audio-based instructions that are easy to learn and remember. This results in a game that can be played by a visually impaired person without the help of a sighted one. In this scenario, better user experience implies more independence for visually impaired people. Based on a follow-up interview with the user and his family, we design several user cases for the game (Figure 1). The intention is to make an Arcade adventure game based on the player's skill. The game consists of a maze where the main character (the player) has to collect objects while avoiding the villain who wants to catch him.

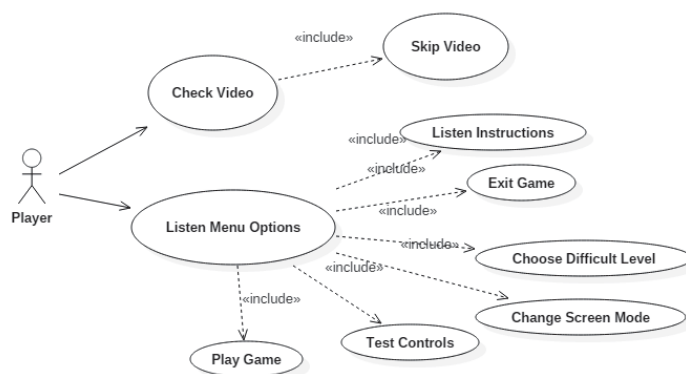


Fig. 1. Use cases detected for the game.

The scenario for the game is a farm, therefore the main character is a farmer named Neri who must gather animals while is pursued by a wolf. The goal is to go through the maze and gather all the animals. If the wolf reaches the farmer and he has animals in his basket, they are released to random positions; if he has no animals, the character dies and the game ends. The player wins when he has managed to collect all the animals. To facilitate the user orientation in the maze, the configuration of the maze is not random. There are a fixed number of mazes that differs in the location of the walls.

Sounds are used to guide the player in the search of the animals. Each animal emits its typical sound, but the player will only be able to listen to the nearest one when it is within a certain range. The further away the animal is to the farmer the lower it sounds. We also used sound to identify the walls on the maze. Every time the user moves and hits a wall, a specific sound is played.

On each maze there are 13 animals that are placed each time at random locations. If the player manages to collect all the animals the game gives the user two possibilities, to choose a scenario other than the current one or to repeat the same maze. This design decision was made considering that the player may prefer to always play in the same scenario and thus be able to memorize the

position of the walls and facilitate orientation. In total there are 10 different mazes that differ in the location of the walls.

We include three aids to help the user in the game, those are *GPS*, *Super Hearing* and *Shield*. The *GPS* aid is always available and when requested it will tell the user how to reach the nearest animal. All instructions in the game are in Spanish, but as an open-source project they can be easily change to other languages. *Super Hearing* allows the user to hear animals from a greater distance and *Shield* protects the player from the wolf. These last two aids only last 20 seconds.

The game provides three levels of difficulty, *easy*, *normal* and *hard*. The difference between each level is only on the behavior of the wolf. On the *easy* level the wolf moves randomly, on the *normal* level the wolf moves to the user and finally, on the *hard* level, a second wolf appears. On this last level one wolf behaves like a wolf of the *easy* level and the second one, like the wolf of the *normal* level.

Although the game was designed for visually impaired users, we decided that it should also be played by sighted users. In this way, the game can be seen as a family activity. For this reason, the game has a GUI which will be detailed in the next section. It is possible to blackout the entire GUI in order to give any user the experience of a visually impaired one, this is known as *blind game mode*.

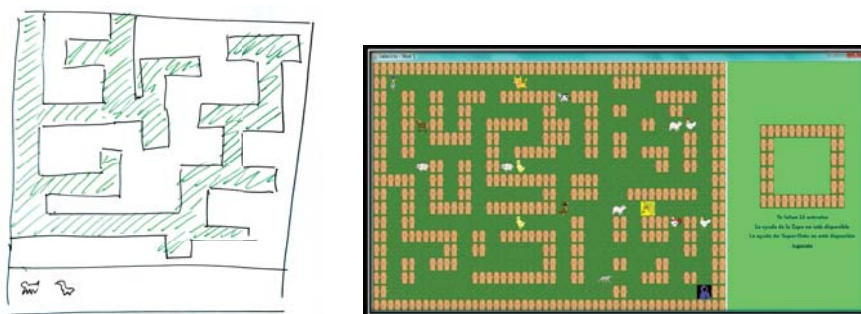
5 Implementation

The language chosen for the development of this game was Java, which allowed us to develop a multiplatform game with as few implementation dependencies as possible. In order to support the use of a joystick or gamepad, the JInput [10] library was included. All the images used in the game were created by hand, scanned and painted digitally. All audio messages were recorded by the authors. The game can be downloaded for free². Its source code is publicly available for download³.

Three sprites were used in the game. One for the main character movements, another for the wolf movement and the last one for the rest of the graphics. Figure 2(b) shows a screenshot of the game. Figure 2(a) shows the first prototype of the GUI on paper. This prototype was used to test the interfaces with the sighted users. On this first version the general layout of the game was vertical, with the maze on the top part and on the bottom the area where the found animal were stock. On interviewing the sighted users we noticed that a horizontal approach would make more efficient use of space on a computer monitor. On the final version of the GUI, the maze is on the left and the stock area on the right. The game contains ten level, where each one is stored separately in a text files. Each line of the file represents a row of the matrix that is a level. The files contain only "p" or "s" characters, the "p" represents a wall and the "s" represents floor.

² <http://cs.uns.edu.ar/~m11/lagranjadeneri/>

³ http://www.cs.uns.edu.ar/~m11/lagranjadeneri/files/neri_source_code.zip



(a) First prototype for the GUI in paper. (b) Screenshot of the final GUI of the game. All visual elements were created by the authors.

Fig. 2. Evolution of the GUI from first paper prototype to final result.

6 Usability test

For the usability test, 4 participants (3 sighted and 1 blind) were recruited for the study. Their ages were between 20 and 55 and they had different educational backgrounds. The blind user was 26 years old at the time of the test and has a cerebral injury from birth. Such a small number of users is insufficient to provide any statistically meaningful results but some conclusions can still be achieved from the experiment results.

A brief training was given to the participants to familiarize them with the game and the audio system. They were introduced to the game story, goals and game play. The sighted participants were also introduced to the *blind game mode*. All users performed training without problem, although the blind user required more attention.

We considered that evaluating the video game for the sighted users as a conventional video game with graphical interfaces will not be meaningful, so we decided that they played in *blind game mode* only.

6.1 User Experience Feedback

#	Task	Success/Failure
1	Test the controls	
2	Listen to instructions	
3	Exit the game	

Table 1. Tasks evaluated in the test. For each task it was marked if the user could accomplish it successfully or not.

Each participant had to accomplish three basic tasks while using the game (see Table 1). And after experiencing the game, they answered a follow-up us-

ability review. In particular, the usability review for blind users consisted in a survey (see Table 2) and the one for sighted users consisted in an interview.

The blind participant usually plays console games that are not prepared for his characteristics. At first, it was difficult for him to get accustomed to the game interface. However, once he played several times, the interface got easier to use.

The sighted participants were asked to play the game in blind mode to obtain more feedback about the understanding of spoken menus and sounds. In this mode the gamers were confused and intrigued about their location in the screen. Moreover, they did not pay enough attention neither to the spoken menus nor to the sounds of the game. This lack of attention resulted in more required time to comprehend the usage and mechanics of the game and then, to be able to actually play.

#	Question	yes/no
1	Was the first thing the user tried to play?	No
2	Did the user listen to all options from the menu?	Yes
3	Did the user test every difficult levels?	No
4	Did the user listen to all the instructions?	Yes
5	Did the user understand that the wolf is bad?	Yes
6	Did the user make the correct association between sound and animals?	Yes
7	Was the user distracted by the sounds?	No
8	Did the user finish the game?	Yes
9	Was the user still well oriented after hitting a wall?	Yes
10	Once the user won, did he want to change the map?	Yes
11	When the user listen an animal, did he try to catch it?	No
12	Did the user use the GPS?	Yes
13	Did the user attend to the orders of the GPS?	Yes
14	Did the user abuse of the usage of the GPS?	Yes
15	If the user did not use the GPS, was it because he did not want to?	No
16	If the user did not use the GPS, was it because he forgot about it?	Yes
17	Did the user follow the instructions?	Yes
18	Did the user prefer to use the keyboard?	No
19	Did the user prefer to use the joystick?	Yes
20	Did the user remember the keys of buttons from the device?	Yes

Table 2. List of yes-no question asked to the user after using the game.

7 Conclusion and future work

In a world where technology is part of every person's daily life, where from the moment we get up to when we go to bed electronic devices join our routine, access to them should be possible for visual impaired users. If used in the right way, video games can inspire learning and improve skills. A good video game is challenging, entertaining, and complicated. Video games can also be a source of

family entertainment, with parents, children, and grandparents all participating on it.

In this work we presented the development of an Spanish open-source video game for visual impaired users. As we presented in Section 3 open-source video games for this type of user are rare, specially in Spanish. With this game, visual impaired users both children and adults, can enjoy a video game created for them. The game can also be seen as a social and family activity, because of the game's GUI all users can share the experience. As the usability test proved, the game was very well received. This result is a simple game but with a wide target audience. As for future work, we would like to include more audio options for the instructions, like the possibility to have a familiar voice giving the instruction (like a parent, friend or relative) and conduct more usability testing.

References

1. Pippin Barr, James Noble, and Robert Biddle. Video game values: Human-computer interaction and games. *Interacting with Computers*, 19(2):180–195, 2007.
2. Antonio Fernández. Zomblind, 2012. URL: <http://web.deantares.es/zomblind/>.
3. Mark Frauenfelder. open-field-echo-sounder, 2014. URL: <http://boingboing.net/2014/10/03/open-field-echo-sounder-game.html>.
4. Mark Griffiths. The educational benefits of videogames. *Education and Health*, 20(3):47–51, 2002.
5. Jonathan Grudin. From tool to partner: The evolution of human-computer interaction. *Synthesis Lectures on Human-Centered Interaction*, 10(1):i–183, 2017.
6. Martin G Helander. *Handbook of human-computer interaction*. Elsevier, 2014.
7. Thomas T Hewett, Ronald Baecker, Stuart Card, Tom Carey, Jean Gasen, Marilyn Mantei, Gary Perlman, Gary Strong, and William Verplank. *ACM SIGCHI curricula for human-computer interaction*. ACM, 1992.
8. Javier Mairena. Audiodisc, 2008. URL: <http://www.javiermairena.net/>.
9. Mark Overmars. Teaching computer science through game design. *Computer*, 37(4):81–83, 2004.
10. Unknown. Jinput, 2016. URL: <https://github.com/jinput/jinput>.
11. Dylan Viale. Quacky's quest, 2012. URL: <https://www.youtube.com/watch?v=1YV21D0JjCY>.
12. T Westin. Game accessibility case study: Terraformers – a real-time 3d graphic game. In *In Proc. of the The Fifth International Conference on Disability, Virtual Reality and Associated Technologies*, pages 95–100, 2004. URL: <http://www.terraformers.nu>.
13. R Willems, C Pinkster, L Kuiper-Hoyng, et al. The explorer and the mystery of the diamond scarab: A serious wii game. 2012.
14. Fredrik Winberg. *Contextualizing Accessibility: Interaction for Blind Computer Users*. PhD thesis, KTH, 2008.

Fusión de información de geometría e intensidad para segmentación de imágenes TOF

Luciano Lorenti, Javier Giacomantone, Oscar Bria, Armando De Giusti

Instituto de Investigación en Informática (III-LIDI),
Facultad de Informática - Universidad Nacional de La Plata - Argentina.
La Plata, Buenos Aires, Argentina.
{llorenti,jog,onb,degiusti}@lidi.info.unlp.edu.ar

Resumen. Las cámaras de tiempo de vuelo (TOF) generan dos imágenes simultáneas, una de intensidad y una de rango. Esto permite abordar problemas de segmentación donde la información de intensidad o de rango separadamente es insuficiente para extraer los objetos de interés de la escena 3D. A su vez, la información de rango permite obtener una aproximación del vector normal de cada punto de las superficies capturadas. En este artículo se presenta un método de segmentación espectral, que combina la información de intensidad, de rango y las orientaciones de los vectores normales para mejorar los resultados de la segmentación. Los agrupamientos obtenidos suponen una estructura subyacente común entre todas las fuentes de información, llamadas vistas. Se utilizan técnicas de clustering espectral co-regularizado para obtener agrupamientos que sean consistentes de acuerdo a todas las vistas. La evaluación del método propuesto fue realizado sobre imágenes reales. El rendimiento obtenido al combinar las tres fuentes de información presenta mejoras en los agrupamientos resultantes.

Palabras claves: Segmentación, Imágenes de Rango, Cámaras de Tiempo de Vuelo, Agrupamiento Espectral.

1. Introducción

El problema de segmentación de imágenes es uno de los principales problemas en el campo de la visión automática. Su objetivo consiste en la extracción de los elementos que constituyen una imagen [3][15]. Para poder lograrlo, estos métodos agrupan píxeles de acuerdo a algún criterio de semejanza. Tradicionalmente, el problema de segmentación de imágenes es abordado utilizando la información de color o intensidad de los objetos presentes en la escena. Los progresos recientes en segmentación de imágenes han mostrado que incorporar la profundidad de los objetos como característica adicional mejora la precisión de los métodos de segmentación [9]. A su vez, a partir de una nube de puntos es posible obtener los vectores normales locales de las superficies. Estos vectores permiten discriminar con mayor precisión los objetos presentes en la escena [7] [6].

Desarrollos recientes en *hardware* permiten estimar la geometría de la escena y posibilitan la utilización de nuevos enfoques para segmentar imágenes. Con esta perspectiva el problema de segmentación puede ser formulado como la búsqueda de formas efectivas para particionar adecuadamente un conjunto de muestras con información de intensidad, distancia e información acerca de la geometría de los objetos presentes en la escena.

En este trabajo utilizamos una cámara de tiempo de vuelo, *Time of Flight* (TOF), que nos permite obtener imágenes de rango y de intensidad simultáneamente, la cámara utilizada es la MESA SR 4000 [4]. La SR 4000 es una cámara activa, utiliza su propia fuente de iluminación mediante una matriz de diodos emisores de luz infrarroja modulada en amplitud. Los sensores de la cámara detectan la luz reflejada en los objetos iluminados y la cámara genera dos imágenes. La imagen de intensidad es proporcional a la amplitud de la onda reflejada y la imagen de rango o distancia es generada a partir de la diferencia de fase entre la onda emitida y reflejada en cada elemento de la imagen [2].

El método propuesto, en una primera etapa, obtiene la información geométrica de la escena a partir de la nube de puntos organizada que provee la cámara TOF mediante el cálculo de los vectores normales locales a las superficies. Luego combina la información de intensidad, de distancia y geométrica para mejorar la calidad de la segmentación utilizando técnicas de agrupamiento espectral co-regularizado [8]. Las técnicas de agrupamiento espectral han mostrado resultados prometedores en el campo de la segmentación de imágenes [14] [5]. Estas técnicas requieren la construcción de un grafo de afinidad entre los píxeles de las imágenes y la resolución de un problema generalizado de autovalores. El método utiliza un mecanismo que permite la construcción del grafo de afinidad que se ajusta a las características particulares de cada imagen. Además, con el objetivo de reducir la demanda computacional obtiene una solución aproximada del problema generalizado de autovalores.

El método propuesto es evaluado comparando 4 métricas de evaluación supervisadas [13] sobre un conjunto de imágenes reales.

El artículo está organizado del siguiente modo, en la sección 2 se muestra la técnica utilizada para la obtención del vector normal a las superficies, en la sección 3 se presenta una revisión de los conceptos de agrupamiento espectral utilizados en el método propuesto. En la sección 4 se expone el método. En la sección 5 se presentan resultados experimentales. Finalmente en la sección 6 se presentan las conclusiones.

2. Estimación del vector normal

Una de las características más importantes para la interpretación de datos de rango es el vector normal de la superficie. Sin embargo, ya que no se puede medir de forma directa, tiene que ser estimado para cada elemento de la nube de puntos. Los nuevos sensores de profundidad desarrollados obtienen nubes de puntos organizadas, en donde los puntos de la nube 3D son muestreados a partir

de una grilla regular 2D. Esto permite la utilización de imágenes integrales para optimizar los tiempos de procesamiento [7].

Una imagen integral I_o correspondiente a una imagen O se define como la suma de todos los elementos que se encuentran dentro de un área rectangular entre $(0,0)$ y (m,n) : $I_O(m,n) = \sum_{i=0}^m \sum_{j=0}^n O(i,j)$. Es posible calcular una imagen integral con una sola pasada por la imagen.

El valor promedio de una región puede ser calculado como:

$$S(I_O, m, n, r) = \frac{1}{4r^2} (I_O(m+r, n+r) - I_O(m-r, n+r) - I_O(m+r, n-r) + I_O(m-r, n-r))$$

En donde (m,n) es el centro de la región y r el radio interno de la región rectangular.

Un modo tradicional de estimar el vector normal a la superficie \vec{n}_p en un punto p en la ubicación en la imagen $(m,n)^T$ es calcular el vector 3D $\vec{v}_{p,h}$ entre el vecino izquierdo y derecho; calcular el vector $\vec{v}_{p,v}$ entre el vecino superior e inferior de p y luego calcular el producto externo entre los dos vectores: $\vec{n}_p = \vec{v}_{p,h} \times \vec{v}_{p,v}$. A partir de la utilización de imágenes integrales es posible calcular el vector normal para cada pixel de la nube de puntos organizada de forma eficiente teniendo en cuenta que:

$$\begin{array}{l} \vec{v}_{p,h,x} = \frac{P_x(m+r,n) - P_x(m-r,n)}{2} \\ \vec{v}_{p,h,y} = \frac{P_y(m+r,n) - P_y(m-r,n)}{2} \\ \vec{v}_{p,h,z} = \frac{S(I_{P_z}, m+1, n, r-1) - S(I_{P_z}, m-1, n, r-1)}{2} \end{array} \left| \begin{array}{l} \vec{v}_{p,v,x} = \frac{P_x(m, n+r) - P_x(m, n-r)}{2} \\ \vec{v}_{p,v,y} = \frac{P_y(m, n+r) - P_y(m, n-r)}{2} \\ \vec{v}_{p,v,z} = \frac{S(I_{P_z}, m, n+1, r-1) - S(I_{P_z}, m, n-1, r-1)}{2} \end{array} \right.$$

Donde P_x, P_y y P_z son mapas de las coordenadas x, y y z de la nube de puntos organizada, I_{P_z} es la imagen integral del componente z de la nube de puntos y r el radio de suavizado a utilizar.

3. Agrupamiento espectral

Dado un conjunto de patrones $X = \{x_1, x_2, \dots, x_n\} \in \mathbb{R}^m$, y una función de semejanza $d : \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}$, es posible construir una matriz de afinidad W tal que $W(i, j) = d(x_i, x_j)$. Los algoritmos de agrupamiento espectral obtienen una representación de los datos en un espacio de dimensión inferior resolviendo el siguiente problema de optimización:

$$\begin{aligned} \max_{U \in \mathbb{R}^{n \times k}} \quad & Tr(U^T L U) \\ \text{s.a.} \quad & U^T U = I \end{aligned} \quad (1)$$

donde $L = D^{-\frac{1}{2}} W D^{-\frac{1}{2}}$ es la matriz laplaciana de W de acuerdo a [12] y D es una matriz diagonal con la suma de las filas de W ubicadas en su diagonal principal. Una vez obtenido U sus filas son consideradas como las nuevas coordenadas de los patrones. En esta nueva representación es más sencillo aplicar un algoritmo de clustering tradicional [14].

Es posible obtener una aproximación a las coordenadas en este nuevo espacio calculando las afinidades de un pequeño conjunto de píxeles y aproximando las afinidades restantes.

Sea $A \subset X$ un subconjunto de patrones muestreados y $B = V - A$, el resto de los patrones no muestreados. W_A es la matriz de semejanza derivada de los datos de A y L_A es la matriz laplaciana de W_A . W_B y L_B son las matrices correspondientes de las afinidades de los puntos de A y B . Es posible definir a L como:

$$W = \begin{bmatrix} W_A & W_B \\ W_B^T & W_C \end{bmatrix} \quad L = \begin{bmatrix} L_A & L_B \\ L_B^T & L_C \end{bmatrix}$$

Es posible obtener una aproximación de W , denominada \hat{W} , solamente a partir de A y B :

$$\hat{W} = \bar{U} \Lambda \bar{U}^T = \begin{bmatrix} A & B \\ B^T & B^T A^{-1} B \end{bmatrix}$$

Con el objetivo de obtener los autovectores de la matriz laplaciana aproximada, $\hat{L} = \hat{D}^{\frac{1}{2}} \hat{W} \hat{D}^{\frac{1}{2}}$, es necesario calcular \hat{L}_A y \hat{L}_B :

$$L_{Aij}^{\hat{}} = \frac{W_{Aij}}{\sqrt{\hat{d}_i \hat{d}_j}} \quad L_{Bij}^{\hat{}} = \frac{W_{Bij}}{\sqrt{\hat{d}_i \hat{d}_{j+|A|}}} \quad (2)$$

donde $\hat{d} = \hat{W} \mathbf{1}$. Si \hat{L}_A es positiva definida, es posible hallar los autovectores ortogonales aproximados en un solo paso. Sea $S = \hat{L}_A + \hat{L}_A^{-\frac{1}{2}} \hat{L}_B \hat{L}_B^T \hat{L}_A^{-\frac{1}{2}}$ y su diagonalización $S = U_S \Lambda_S U_S^T$, Fowkles et al. [5] demostraron que si la matriz V se define como

$$V = \begin{bmatrix} \hat{L}_A \\ \hat{L}_B^T \end{bmatrix} \hat{L}_A^{-\frac{1}{2}} U_S \Lambda_S^{-\frac{1}{2}} \quad (3)$$

\hat{L} es diagonalizada por V y por Λ_S y $V^T V = I$

3.1. Co-regularización

Cuando el conjunto de datos tiene más de una representación, a cada una de ellas se la denomina vistas. En el contexto de agrupamiento espectral las técnicas de co-regularización intentan fomentar la semejanza de los ejemplos en la nueva representación generada a partir de los autovectores de cada una de las vistas.

Sean $X^{(v)} = \{x_1^{(v)}, x_2^{(v)}, \dots, x_m^{(v)}\}$ los ejemplos para la vista v y $L^{(v)}$ la matriz laplaciana creada a partir de X para la vista v . Definimos $U^{(v)}$ a la matriz formada por los primeros k autovectores correspondientes como la matriz $L^{(v)}$ de acuerdo con (1). En [8] fue propuesto un criterio que mide el desacuerdo entre dos representaciones:

$$D(U^{(v)}, U^{(w)}) = \left\| \frac{K_{U^{(v)}}}{\|K_{U^{(v)}}\|_F} - \frac{K_{U^{(w)}}}{\|K_{U^{(w)}}\|_F} \right\|_F^2$$

Donde $K_{U^{(v)}}$ es la matriz de semejanza generada a partir de los patrones en la nueva representación $U^{(v)}$ y $\|\cdot\|_F$ es la norma Frobenius. Si se utiliza como medida de semejanza el producto interno entre los vectores se obtiene $K_{U^{(v)}} = U^{(v)}U^{(v)T}$. Ignorando las constantes aditivas y de escalado, la ecuación anterior puede ser formulada de la siguiente manera:

$$D(U^{(v)}, U^{(w)}) = -Tr \left(U^{(v)}U^{(v)T}U^{(w)}U^{(w)T} \right) \quad (4)$$

El objetivo es minimizar el desacuerdo entre las representaciones obtenidas a partir de cada una de las vistas. Por lo tanto, si se disponen de m vistas, se obtiene el siguiente problema de optimización que combina los objetivos de agrupamiento espectral individuales y el objetivo que determina el desacuerdo entre las representaciones:

$$\max_{\substack{U^{(i)} \in R^{n \times k}, \\ 1 \leq i \leq m}} \sum_{v=1}^m Tr \left(U^{(v)T}L^{(v)}U^{(v)} \right) + \lambda \sum_{\substack{1 \leq v, w \leq m \\ v \neq w}} Tr \left(U^{(w)T}L^{(w)}U^{(w)} \right) \quad (5)$$

$$\text{s.a.} \quad U^{(v)T}U^{(v)} = I \quad \forall 1 \leq v \leq m$$

El parámetro λ balancea el objetivo de agrupamiento espectral y el de desacuerdo entre las representaciones. El problema de optimización conjunta puede ser resuelto utilizando maximización alternante. Dados $U^{(w)}, 1 \leq w \leq m$ dado se obtiene el siguiente problema de optimización para $U^{(v)}, v \neq w$:

$$\max_{U^{(v)} \in R^{n \times k}} Tr \left(U^{(v)T} \left(L^{(v)} + \lambda \sum_{\substack{1 \leq w \leq m \\ v \neq w}} U^{(w)}U^{(w)T} \right) U^{(v)} \right) \quad (6)$$

$$\text{s.a.} \quad U^{(v)T}U^{(v)} = I$$

Lo que resulta en un algoritmo de clustering tradicional con la matriz laplaciana modificada $L^{(v)} + \lambda \sum_{\substack{1 \leq w \leq m \\ v \neq w}} U^{(w)}U^{(w)T}$

4. Método propuesto

A partir de la imagen de intensidad I y la imagen de rango R provistas por la cámara de tiempo de vuelo se obtiene el mapa con los vectores normales de cada punto de la superficie, N , según lo descrito en la sección 2.

Para determinar la semejanza W_{ij} entre cada elemento de una imagen $\text{Img} \in \{I, R, N\}$ se utiliza una función que combina la distancia de los pixeles en el plano de la imagen y la semejanza entre sus valores:

$$W(\text{Img})_{ij} = \exp \left(\frac{-\|\text{pos}_i - \text{pos}_j\|_2^2}{2(sx)^2} - \frac{d(\text{Img}(i), \text{Img}(j))^2}{2(sy)^2} \right)$$

En donde pos_i es la ubicación espacial (x, y) del pixel i -ésimo; $\text{Img}(i)$ son los elementos i -ésimos de la imagen. El parámetro sx determina la importancia otorgada a la ubicación espacial en la función de semejanza y sy determina la importancia otorgada a la diferencia entre los valores de cada pixel. d es una función de distancia entre los elementos de la imagen.

En lugar de seleccionar un solo parámetro sy para toda la imagen en [16] proponen calcular un parámetro de escalado local para cada punto teniendo en cuenta las estadísticas locales de su vecindad. La escala local para un punto i de una imagen P utilizando una distancia d se define como $\max d(P(i), P(j)) \forall j \in N(i)$, en donde $N(i)$ son todos los vecinos dentro de un radio de r pixeles.

Sean p, r dos elementos de la imagen $I, d_I(p, r) = |p - r|$. La misma función se utiliza para la imagen de rango. Si p, r son dos elementos de N , $d_N(p, r) = \mathbf{p}^T \mathbf{r}$.

El método propuesto consiste en las siguientes etapas:

1. A partir de I, R y N se obtienen las matrices laplacianas aproximadas \hat{L}_1 y \hat{L}_2, \hat{L}_3 , respectivamente según lo descrito en (2). Utilizando para cada imagen su función de semejanza correspondientes: $W(I), W(R)$ y $W(N)$.
2. Se obtiene \hat{V}_2, \hat{V}_3 los autovectores aproximados de \hat{L}_2 y \hat{L}_3 calculados de acuerdo a (3)
3. Se resuelve el problema de optimización 6 para \hat{V}_1 dadas \hat{V}_2 y \hat{V}_3 .
4. La optimización es llevada a cabo de forma cíclica por todas las vistas manteniendo fijas las obtenidas previamente.
5. Se evalúa $\sum_{i=1}^3 \sum_{j=1}^3 D(V_i, V_j)$. Si el desacuerdo disminuye ir a 4.
6. Se aplica un algoritmo de agrupamiento sobre \hat{V}_1

5. Resultados experimentales

El rendimiento del algoritmo de segmentación propuesto fue evaluado sobre 13 imágenes capturadas utilizando la cámara de tiempo de vuelo MESA Swiss-Ranger SR4000 [4]. La cámara de tiempo de vuelo proporciona dos imágenes: una imagen de amplitud y una imagen de rango ambas de 144×176 pixeles.

Se utilizan las siguientes métricas evaluación supervisadas para determinar la calidad de la segmentación obtenida con el método propuesto: la medida de precisión-exhaustividad para objetos y partes, F_{op} , propuesto en [13], la medida de precisión-exhaustividad para contornos [10], F_b , la Cobertura de Segmentación [1], SegCov y la variación de información [11], VoI . En este contexto una segmentación obtenida es mejor cuanto menor es su VoI y cuando mayor es su SegCov, F_b y su F_{op} .

Se utilizó un muestreo de 150 píxeles para construir las matrices de afinidad y se obtuvieron los 6 autovectores correspondientes a los autovalores de mayor magnitud para generar el espacio en donde se aplica k-medias como método de agrupamiento. Como término de co-regularización se utilizó $\lambda = 0,001$ según lo sugerido en [8].

Con el objetivo de determinar el parámetro sx , se evaluó la métrica F_{op} para el conjunto de las imágenes TOF. La figura 1a muestra el promedio de los resultados para cada imagen con respecto a la variación de la influencia de la ubicación espacial en la función de semejanza. Teniendo en cuenta los resultados se estableció $sx = 35$. La figura 1b muestra la influencia del tamaño de la ventana a considerar al momento de seleccionar el escalado local de sy . Se utilizó $r = 5$.

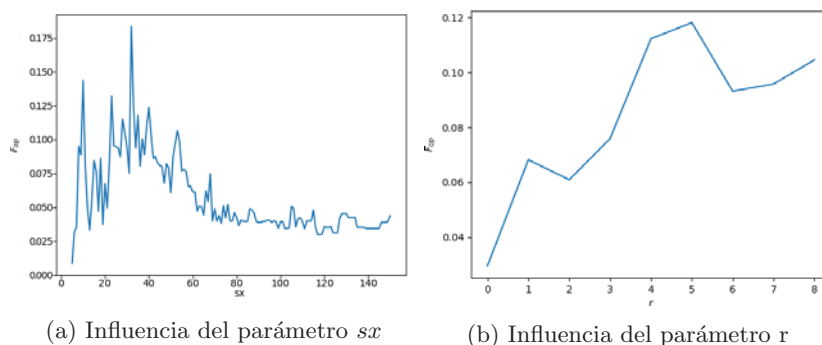


Figura 1: Influencia de los parámetros sx y r con respecto a la métrica de precisión-exhaustividad de objetos y partes

La tabla 1 muestra el análisis de rendimiento del método propuesto en comparación con utilizar el método de Nystrom tradicional [5] sobre la imagen de intensidad, de rango, y normal por separado, Co-Regularización entre intensidad y rango [9] y el método propuesto. Los resultados representan el promedio de 10 ejecuciones sobre las 13 imágenes del conjunto de datos. Es posible observar que el método propuesto mejora las métricas F_{op} , F_b y VoI , indicando que los agrupamientos obtenidos tienen mayor coincidencia con los agrupamientos del agrupamiento de referencia. La medida SegCov es mayor para las imágenes de rango e intensidad. Esto puede deberse a que los agrupamientos obtenidos a partir de estas imágenes tienden a agrupar al fondo junto con los objetos. Es-

ta métrica de evaluación no penaliza en gran medida las segmentaciones poco precisas.

	F_b	VoI	F_{op}	Seg. Cov.
Intensidad	0.07935	0.23629	0.32422	2.80678
Rango	0.05077	0.22466	0.27684	3.05854
Normal	0.05283	0.17645	0.32911	2.95473
IR Co-Reg	0.08162	0.23582	0.32879	2.70953
Método Propuesto	0.10558	0.16610	0.35124	2.79449

Tabla 1: Evaluación de rendimiento del método propuesto

La figura 2 presenta resultados experimentales del método propuesto aplicado a dos capturas del conjunto de datos. Es posible observar de forma cualitativa que los segmentos obtenidos a partir de la imagen recobran los objetos presentes en la escena.

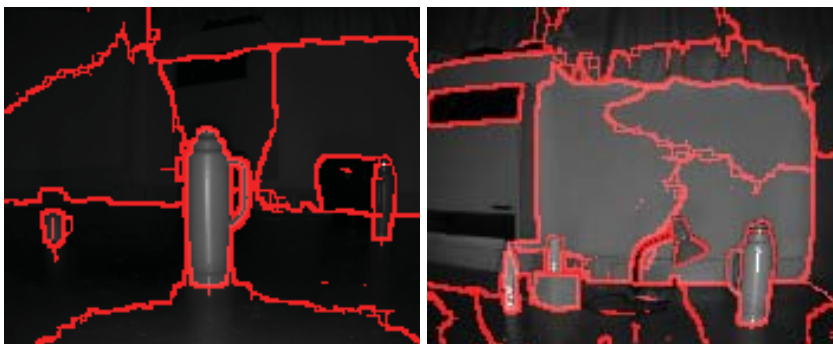


Figura 2: Segmentación realizada utilizando el método propuesto

6. Conclusiones

En este trabajo presentamos un método de agrupamiento aplicado a segmentación de imágenes capturadas con cámaras de tiempo vuelo. Los datos presentan resultados preliminares satisfactorios. El algoritmo extrae información de la geometría de la escena y los combina con los datos de intensidad y rango mejorando los resultados de la segmentación. El rendimiento resultante al utilizar las orientaciones de los vectores normales en el contexto de aprendizaje semi-supervisado

presenta mejoras en los casos probados de acuerdo a las métricas utilizadas. Una etapa futura de este trabajo prevé la obtención de los autovectores aproximados con métodos más eficientes y menos sensibles al ruido, la utilización de otras técnicas de agrupamiento en la etapa final del método y la incorporación de información adicional a la función de semejanza.

Referencias

1. Arbelaez, P., Maire, M., Fowlkes, C., Malik, J.: Contour detection and hierarchical image segmentation. *IEEE transactions on pattern analysis and machine intelligence* 33(5), 898–916 (2011)
2. Blanc, N., Oggier, T., Gruener, G., Weingarten, J., Codourey, A., Seitz, P.: Miniaturized smart cameras for 3d-imaging in real-time [mobile robot applications]. In: *Sensors, 2004. Proceedings of IEEE*. pp. 471–474 vol.1 (Oct 2004)
3. Canny, J.: A computational approach to edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on PAMI-8(6)*, 679–698 (Nov 1986)
4. Cazorla, M., Viejo, D., Pomares, C.: Study of the sr 4000 camera. In: *XI Workshop de Agentes FÁsicos (2004)*
5. Fowlkes, C., Belongie, S., Chung, F., Malik, J.: Spectral grouping using the nyström method. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26(2), 214–225 (February 2004)
6. Holz, D., Behnke, S.: Fast range image segmentation and smoothing using approximate surface reconstruction and region growing. *Intelligent autonomous systems* 12 pp. 61–73 (2013)
7. Holzer, S., Rusu, R.B., Dixon, M., Gedikli, S., Navab, N.: Adaptive neighborhood selection for real-time surface normal estimation from organized point cloud data using integral images. In: *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*. pp. 2684–2689. IEEE (2012)
8. Kumar, A., Rai, P., Daume, H.: Co-regularized multi-view spectral clustering. In: *Shawe-Taylor, J., Zemel, R., Bartlett, P., Pereira, F., Weinberger, K. (eds.) Advances in Neural Information Processing Systems 24*, pp. 1413–1421. Curran Associates, Inc. (2011), <http://papers.nips.cc/paper/4360-co-regularized-multi-view-spectral-clustering.pdf>
9. Lorenti, L., Giacomantone, J.: Time of flight image segmentation through co-regularized spectral clustering. In: *XX Congreso Argentino de Ciencias de la Computación (CACIC 2014) (2014)*
10. Martin, D.R., Fowlkes, C.C., Malik, J.: Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE transactions on pattern analysis and machine intelligence* 26(5), 530–549 (2004)
11. Meila, M.: Comparing clusterings: an axiomatic view. In: *Proceedings of the 22nd international conference on Machine learning*. pp. 577–584. ACM (2005)
12. Ng, A.Y., Jordan, M.I., Weiss, Y.: On spectral clustering: Analysis and an algorithm. In: *ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS*. pp. 849–856. MIT Press (2001)
13. Pont-Tuset, J., Marques, F.: Measures and meta-measures for the supervised evaluation of image segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 2131–2138 (2013)
14. Shi, J., Malik, J.: Normalized cuts and image segmentation. In: *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*. pp. 731–737 (Jun 1997)

15. Wu, Z. y Leahy, R.: An optimal graph theoretic approach to data clustering: theory and its application to image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 15(11), 1101–1113 (Nov 1993)
16. Zelnik-Manor, L., Perona, P.: Self-tuning spectral clustering. In: *Advances in neural information processing systems*. pp. 1601–1608 (2005)

Immersive Platform for Neuroscience Experimental Studies

Florencia Rodríguez¹, Marcos Lazo¹, Virginia Cifuentes¹, Juan Pablo D'Amato¹, Manuel Serodio², Fabricio Ballarini², Pedro Bekinschtein³, and Cristian García Bauza¹

¹ PLADEMA, Universidad Nacional del Centro, Tandil, Buenos Aires, Argentina
{mfrodri, mlazo, cifuente, jpdamato, crgarcia}@exa.unicen.edu.ar,
<http://www.pladema.net>

² Instituto de Biología Celular y Neurociencias, Facultad de Medicina, UBA, Buenos Aires, Argentina
{manuelserodio.arg, faballarini}@gmail.com
<http://www.ibcn.fmed.uba.ar>

³ Instituto de Neurociencia Cognitiva y Traslacional, CONICET-INECO, Universidad Favaloro, Buenos Aires, Argentina
pbekinschtein@fmed.uba.ar
<http://www.ibcn.fmed.uba.ar>

Abstract. In this work, we present a computational toolkit to perform neuroscience experiments with people inside virtual immersive environment like a CAVE (Computer Assisted Virtual Environment). Several experiments can be designed, performed and measured for customized virtual scenes. Each participant is exposed to experiences in which they search and collect objects within a three-dimensional space, where they have to use spatial strategies to solve the task. The purpose is to study cognitive functions such as perception, attention, cognition and short term memory.

These tools can reduce the time of experiments formulations and helps to discover mechanisms of attention and spatial and episodic memory. All information about each participant trajectory, the virtual scene configuration, the location of objects in these scenes, and answers to particular questions regarding the experience is collected and presented in.

Keywords: Virtual Reality, CAVE, Data Visualization, Neuroscience

1 Introduction

Computer Assisted Virtual Environment technology or CAVE aims to improve the sense of immersion using immersive virtual environments [14]. This methodology has great acceptance in the field of scientific visualization, engineering, defense and aeronautics, and others [21, 10, 7].

CAVE is implanted in a room, usually cube shape whose walls project computer-generated images. These virtual environments allow working in multiple operating situations similar to the real ones world, and evaluate other atypical situa-

tions or unforeseen risks, always using a combination of visually real scenes and tangible elements, such as knobs, joysticks or controls [6].

Different study protocols about perception, attention, cognition and memory can be constructed using virtual reality techniques. For example, VREX [28] provides a procedural generation of (interconnected) rooms that can be automatically furnished with a click of a button. Researchers can combine different rooms and environments to perform end-to-end experiments including different testing situations and data collection.

Most neuropsychological research has focused on increasing our understanding of component cognitive processes, such as attention, executive functions, memory, and language spatial abilities.

One source of variability in memory performance is hypothesized to be tied to changes in the input to the dentate gyrus (DG). Rodent studies have demonstrated that the DG has a particular role in spatial pattern separation [16, 20]. Stark [27] defines the pattern separation such as the ability to distinguish among similar experiences. Hunsaker [18] placed rats with localized DG lesions in an environment with two objects spaced 60 cm apart. When the animals were later placed in the same environment with the same objects now placed 40 cm apart, DG-lesioned animals (unlike control animals) did not re-explore the objects or environment. These data suggest that the DG-lesioned rats were not able to discriminate between the training and test environments.

Since object-based pattern separation appears to be modulated by the DG in humans, we wondered if these findings could be extended to spatial pattern separation in a RV environment. So, we extended these studies to the human, using a task designed to assess spatial pattern separation behavior in a CAVE [3]. In our work we want to present the cave as a virtual reality tool that is used for neuroscience experiments. The idea is to extend to human pattern-separation experiments previously done on rats [15, 30].

This paper presents a software platform deployed on a CAVE environment for carrying out experiments in neuroscience. The idea is to observe and measure how adults responds to 3D stimulus. In the future, the main purpose is to apply this software in neurogenesis understanding [9, 8], studying the short term memory based on objects geo-location correlated physical activities.

2 Methodology

2.1 Experiments on CAVE

As it was previously named CAVE consists of a closed room wherein computer-generated virtual environments are projected generating the feeling of immersion of the user. Particularly, CAVE installed in PLADEMA Institute is formed by a three meters metal cube-shaped structure, whose inner walls are coated with special fabrics that allow for rear-projection (Fig.1). Four projectors are used to display the image on the front and side walls and floor. Each of the projectors is connected to a dedicated PC responsible for generating the image.

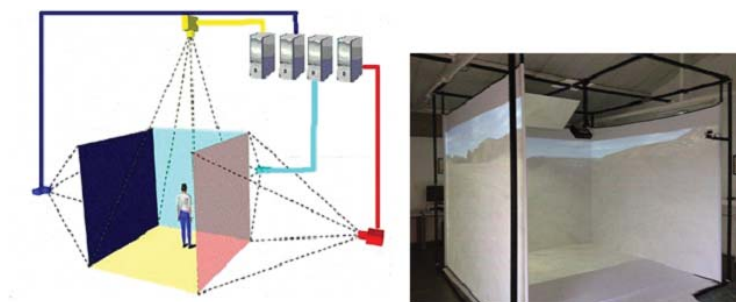


Fig. 1. CAVE installed in PLADEMA Institute.

The projecting scene in CAVE is selected from a set of predefined landscapes. Following, different components are introduced to the experimental scene as the experimental subject, being the pennants, the questions, or the zonings are some examples. A homely joystick is used to move on the scene and interact with applications.

Two types of actions are allowed for the participant:

1. **Collect the pennant.** It is the statement that defines the action whereby the subject will have to search a pennant located in a particular position of the scene.
2. **Locate the pennant.** It is the instruction whereby a pennant is assigned to a subject that should place it in a position in the scene, intermediate or between to the two flags that were collected before.

2.2 Editor of experiments

To perform the testing exercises, an edition tool, that configures individual experiences and presents an immersive virtual environment for each participant, was developed.

With the editor, the number of pairs, the location and the color of pennants and the configurations that the neuroscientists want to evaluate are added: place the pairs of pennants in a circular shape around the center of the arena (Fig.9) or select between different conditions (C1 y C2) during the test execution (Fig.10).

As it was previously stated, the research scene is represented by a wide and flat terrain that the user can tour entirely. There is a mark on the floor indicating the center of the arena to be used a reference point for the beginning of the exercise and distance measurement. The pennants are distributed in a uniformly distance of the center of the arena, according to the scene configuration.

The software has a set of 12 scenarios. The predefined scene shows a quarry like the one shown in Fig.2. The rest of the scenarios have the same place but add shrubs, grass, drums, squares and other elements. The software user can choose between scenes according to the number and variety of objects he or she wishes to view.

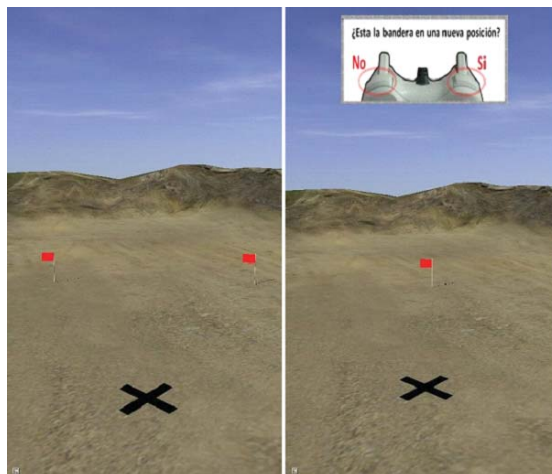


Fig. 2. Beginning of the exercise (left).Execution of the condition 2 (right).

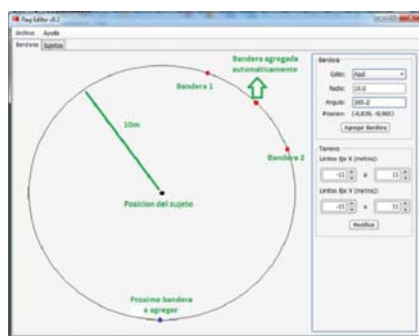


Fig. 3. Scene Editor Interface: location of pennants.

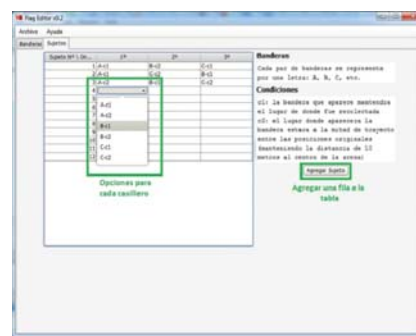


Fig. 4. Interface Scene Editor: selection of patterns.

In addition to the module in charge of the design and visualization of scenarios, the tool has a module that is in charge of the logic of the execution of the experiment. This means the sequence with which the elements of the scene (flags and boxes with questions) appear and disappear according to the movements of the participant.

We are currently working to develop a tool that allows us to configure a greater number of parameters in order to be able to perform an endless number of experiments.

3 Formulation of experiences

At the beginning of experimentation, the participant is directed to the center of the virtual scene and makes a practice test executing an instruction set, that

was previously given to accomplish the test objectives. Among others, the actions that perform an experimental subject on the immersive virtual scene (o setting) are displacements to either explore, collect or drop different objects. All these actions are performed using a standard joystick.

3.1 Experiments based on pattern separation

In these experiments, the participant appears in an open space with weak environmental sound. To move or collect pennants (until three pairs: A, B, and C) a joystick is used. The test places the pennants, circularly, at approximately 10 meters.

The partaker must collect three pairs of pennants that form angles of 50° , 80° and 120° with each other relative to the center stage. Such order corresponds with the pairs A, B y C respectively. Table 1 presents the studied alternatives posed to each participant.

Alternative	1st test	2nd test	3rd test
A1	50°	80°	120°
A2	80°	120°	50°
A3	120°	50°	80°
A4	50°	120°	80°
A5	80°	50°	120°
A6	120°	80°	50°

Table 1. Alternatives of separation of objects.

The partaker proceeds with each pair of pennants as follows: collects the first pennant, returns to the center of the arena, and repeats the same actions with the second pennant. Then, the system locates a new pennant within the scene. Now, the subject answers if the pennant is on the same position as any of the ones previously collected (see option C1 on Fig. 5), or midway between the two original positions of the pair of pennants (see option C2 on Fig.5), while keeping the distance of 10 meters to the center of the arena. There are no variations in the remaining experiences.

In Table 1 the different alternatives (A_i) can be observed, according with the order in which the pennants appear according to the angle. Thus, for alternative A_1 , the two pennants will first appear separated by an angle of 50° ; after the participant completes the 1st test, the flags will appear separated by an angle of 80° ; and for the third test the pennants will be separated by an angle of 120° .

Protocols (or paradigms) are defined to consider in each simulated experiences. They establish alternative location of objects within the virtual environment, according to the order of appearance of the pairs of pennants (pair A, B and C), and the location of appearance of the new pennant (options C1 or C2). The protocols considered are as follows.

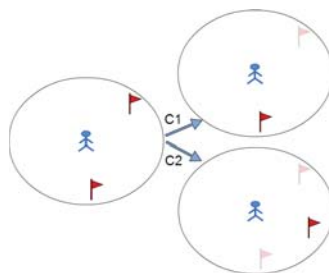


Fig. 5. Posibles ubicaciones de los banderines al responder la pregunta.

To evaluate the memorization of the participant, within dialog integrated boxes into the scene, some questions are formulated. Each question has its correct answer and it is not informed to the subject, it only serves to the neuroscientist to validate the answer or evaluate the experiment with this particular subject.

Finally, while the participant is performing the virtual experience within the CAVE, their movements are automatically recorded: the position of the pennants that has to collect (or move), the distance between them at the beginning and end of each activity, among others data.

3.2 Characterization of the participant behavior

To characterize the overall behavior of the subjects, different agreed metrics are represented and described below.

- **Total effective time:** the time the subject take to perform the test.
- **Individual routes:** participant's positions are recorded along the experiment. Considering the amount of evidence, tours are organized in order. The output is a set of coordinates and grouped data by subject and route. This representation allows observing how people have moved on scene.
- **Response time:** Between each test, the person takes some time to respond to the slogan. This time is calculated as the average delay time to answer the three questions.
- **Characterization of the route:** One idea is to characterize the movement of participant in relation to the objective; particularly if they move (or not) in a beeline from the scene center to each pennant. To indicate quantitatively how a person moves, it is calculated as the average distance to the ideal straight (calculated using linear regression). To characterize the way, it is said that if the average distance is low, the route tends to be linear; whereas if it is high, it presents nonlinear forms.
- **Success rate:** It is estimated from the number of right correct questions over the sum between the correct and incorrect questions.

4 Results

Eight people have involved, two female and the others males, between 24 and 45 years old among which there were students and graduates of the Engineering Systems career of the Exact Sciences Faculty of UNICEN.

Participants performed the different exercises on an immersive virtual environment. After analyzing the experiments, the information was processed in order to characterize the behavior of the participants. To do this, the application builds the effective visualization of the routes from the information collected in CAVE.

The partaker is located in the scene center. From this point, the partaker begins to move to collect the pennants. Each route is projected on the Cartesian plane as shown in Fig.6. The route path display for each participant is constructed using the JFreeChart⁴ library. In each route (going and forth to the starting point) collects a pennant; each iteration consists of two trajectories and a question about the location of the new pennant. The colors of the graphics grouped the tests performed on each iteration.

Although the set of samples is dramatically reduced, from Table 2, it follows that S0, S2 and S5 have correctly answered all the questions. However, they consumed very different times. This may be because of the expertise at using that each participant has the joystick, or the alternative separations of the pennants, S0 presents excessively abrupt direction changes in the followed trajectory while S2 and S5 remain almost rectilinear movements (see Fig.6).

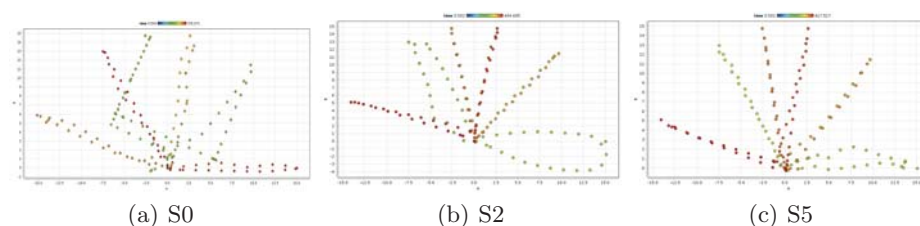


Fig. 6. Participant displacements in the virtual environment.

If we analyze the data answers times (see Table 2 and Fig.7), we find that in most cases the subject correctly answers the first question. Incorrect answers correspond to subjects that performed the alternatives A1 and A2. Therefore, the separation of the pennants on those who have answered the question is 50° and 80° . The same applies to the second question, the subject that answers wrongly performed the A1 and A2 alternatives, but the angle separation between pennants is 80° and 120° . The third question was answered wrongly by 2 of 8 individuals, who perform again the A1 and A2 alternatives with the pennants separated by 50° and 120° angles.

⁴ <http://www.jfree.org/jfreechart/>

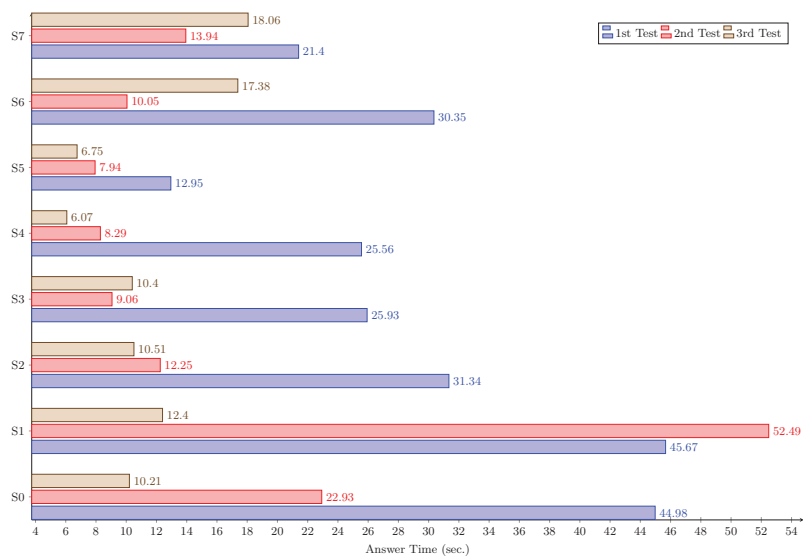


Fig. 7. Answer Time for each participant

5 Conclusions

A kit of computational tools was developed for neuroscience experiments in virtual environments to process the information that comes from actions performed by each participant and, finally, elaborate pertinent conclusions. Specifically, the scenes are projected onto the CAVE where participants explore, gather or locate different objects. The tool design can be extended: add new stages to create new metrics, apply filters to discard irrelevant information or quantify ratios between different variables of interest through their correlation. For example, one of the ideas was to measure the correlation between the time response and the difficulty of the exercise. Another idea is to analyze the movement within the scene when

Id	Aleternative	Time means (sec.)	Answers time (sec.)			Success		
			1st	2nd	3rd	1st	2nd	3rd
S0	A1	26,04	44,98	22,93	10,21	1	1	1
S1	A2	36,85	45,67	52,49	12,40	1	0	0
S2	A6	18,04	31,34	12,25	10,51	1	1	1
S3	A1	15,13	25,93	9,06	10,40	1	0	1
S4	A2	13,31	25,56	8,29	6,07	1	0	1
S5	A6	9,22	12,95	7,94	6,75	1	1	1
S6	A1	19,26	30,35	10,05	17,38	0	1	0
S7	A2	17,80	21,40	13,94	18,06	0	1	1

Table 2. Shows the delay responding times and the successes.

the person has no history of controls use as the joystick, or experience in 3D environments.

On the other hand, it is planned to extend the application to provide support to display data in different graphic types under different configurations, being able to record demographic data of the subjects to establish a relationship between the data produced by the CAVE and the demographic data of the subject.

References

1. Simex: Simulador de maniobras y operaciones para el manejo de excavadoras hidráulicas en entornos de trabajo relacionados a la minería y los trabajos pesados, 12 2013. Fundación UOCRA, Instituto de Estadística y Registro de la Industria de la Construcción (IERIC) y Cámara Argentina de la Construcción. Financiado por el Ministerio de Trabajo, Empleo y Seguridad Social de la Nación.
2. Subtes: Sistema simulador de conducción de tren subterráneo para la formación de personal, 2015. Subterráneos de Bs. As. Sociedad del Estado (SBASE), Ciudad Autónoma de Bs. As.
3. AIMONE, J. B., DENG, W., AND GAGE, F. H. Adult neurogenesis: integrating theories and separating functions. *Trends in cognitive sciences* 14, 7 (2010), 325–337.
4. ARONOV, D., AND TANK, D. W. Engagement of neural circuits underlying 2d spatial navigation in a rodent virtual reality system. *Neuron* 84, 2 (2014), 442–456.
5. ARRIBAS-BEL, D. Accidental, open and everywhere: Emerging data sources for the understanding of cities. *Applied Geography* 49 (2014), 45–53.
6. AYAZ, H., ALLEN, S. L., PLATEK, S. M., AND ONARAL, B. Maze suite 1.0: A complete set of tools to prepare, present, and analyze navigational and spatial cognitive neuroscience experiments. *Behavior research methods* 40, 1 (2008), 353–359.
7. BAUZA, C. G., D'AMATO, J., GARIGLIO, A., ABÁSULO, M. J., VÉNERE, M., MANRESA-YEE, C., AND MAS-SANSÓ, R. A tennis training application using 3d gesture recognition. In *International Conference on Articulated Motion and Deformable Objects* (2012), Springer, pp. 239–249.
8. BEKINSCHTEIN, P., KENT, B. A., OOMEN, C. A., CLEMENSON, G. D., GAGE, F. H., SAKSIDA, L. M., AND BUSSEY, T. J. Brain-derived neurotrophic factor interacts with adult-born immature cells in the dentate gyrus during consolidation of overlapping memories. *Hippocampus* 24, 8 (2014), 905–911.
9. BEKINSCHTEIN, P., OOMEN, C. A., SAKSIDA, L. M., AND BUSSEY, T. J. Effects of environmental enrichment and voluntary exercise on neurogenesis, learning and memory, and pattern separation: Bdnf as a critical variable? In *Seminars in cell & developmental biology* (2011), vol. 22, Elsevier, pp. 536–542.
10. BORONI, G., BAUZA, C. G., D'AMATO, J., AND LAZO, M. Siper-virtual reality simulator of periscope. *World Applied Sciences Journal* 18, 6 (2012), 813–817.
11. BORONI, G., AND VÉNERE, M. Un simulador distribuido para entrenamiento de operarios. In *VIII Congreso Argentino de Ciencias de la Computación* (2002).
12. CHAWLA, M., GUZOWSKI, J., RAMIREZ-AMAYA, V., LIPA, P., HOFFMAN, K., MARRIOTT, L., WORLEY, P., MCNAUGHTON, B., AND BARNES, C. Sparse, environmentally selective expression of arc rna in the upper blade of the rodent fascia dentata by brief spatial experience. *Hippocampus* 15, 5 (2005), 579–586.

13. EWING, R., AND CLEMENTE, O. *Measuring urban design: Metrics for livable places*. Island Press, 2013.
14. FERRINGTON, G., AND LOGE, K. Virtual reality: A new learning environment. *Computing Teacher* 19, 7 (1992), 16–19.
15. GILBERT, P. E., AND KESNER, R. P. The role of the dorsal ca3 hippocampal subregion in spatial working memory and pattern separation. *Behavioural brain research* 169, 1 (2006), 142–149.
16. GILBERT, P. E., KESNER, R. P., AND LEE, I. Dissociating hippocampal subregions: A double dissociation between dentate gyrus and ca1. *Hippocampus* 11, 6 (2001), 626–636.
17. GUZOWSKI, J. F., KNIERIM, J. J., AND MOSER, E. I. Ensemble dynamics of hippocampal regions ca3 and ca1. *Neuron* 44, 4 (2004), 581–584.
18. HUNSAKER, M. R., ROSENBERG, J. S., AND KESNER, R. P. The role of the dentate gyrus, ca3a, b, and ca3c for detecting spatial and environmental novelty. *Hippocampus* 18, 10 (2008), 1064–1073.
19. I, O., M, S., BORONI, G., AND VÉNERE, M. Simulation in real time of radar of horizontal scan. In *Proceedings of First South American Congress on Computational Mechanics* (2002).
20. KESNER, R. P., LEE, I., AND GILBERT, P. A behavioral assessment of hippocampal function based on a subregional analysis. *Reviews in the Neurosciences* 15, 5 (2004), 333–352.
21. KRIJN, M., EMMELKAMP, P. M., OLAFSSON, R. P., AND BIEMOND, R. Virtual reality exposure therapy of anxiety disorders: A review. *Clinical psychology review* 24, 3 (2004), 259–281.
22. LEE, S., AND TALEN, E. Measuring walkability: A note on auditing methods. *Journal of Urban Design* 19, 3 (2014), 368–388.
23. LEUTGEB, S., AND LEUTGEB, J. K. Pattern separation, pattern completion, and new neuronal codes within a continuous ca3 map. *Learning & Memory* 14, 11 (2007), 745–757.
24. MARCOS, L., CRISTIAN, G. B., AND D'AMATO, J. Sistema de simulación y entrenamiento: visualización de un entorno virtual de navegación.
25. RAMIREZ-AMAYA, V., MARRONE, D. F., GAGE, F. H., WORLEY, P. F., AND BARNES, C. A. Integration of new neurons into functional neural networks. *The Journal of neuroscience* 26, 47 (2006), 12237–12241.
26. SCHULTHEIS, M. T., HIMELSTEIN, J., AND RIZZO, A. A. Virtual reality and neuropsychology: upgrading the current tools. *The Journal of head trauma rehabilitation* 17, 5 (2002), 378–394.
27. STARK, S. M., YASSA, M. A., AND STARK, C. E. Individual differences in spatial pattern separation performance associated with healthy aging in humans. *Learning & Memory* 17, 6 (2010), 284–288.
28. VASSER, M., KÄNGSEPP, M., MAGOMEDKERIMOV, M., KILVITS, K., STAFINJAK, V., KIVISIK, T., VICENTE, R., AND ARU, J. Vrex: an open-source toolbox for creating 3d virtual reality experiments. *BMC psychology* 5, 1 (2017), 4.
29. VÉNERE, M., CIFUENTES, M. V., D'AMATO, J. P., AND GARCÍA BAUZA, C. Editor de escenarios para aplicaciones de realidad virtual. *34^a JAIIO, Jornadas Argentinas de Informática e Investigación Operativa* (2005).
30. WILSON, I. A., GALLAGHER, M., EICHENBAUM, H., AND TANILA, H. Neurocognitive aging: prior memories hinder new hippocampal encoding. *Trends in neurosciences* 29, 12 (2006), 662–670.

Separación de sombras a los objetos detectados con sustracción de fondo en video

Rosana Barbuzza^{1,3}, Leonardo Fernandez Esteberena^{1,2}, Leonardo Dominguez^{1,2}, Alejandro Perez¹, Aldo Rubiales^{1,3} y Juan P. D'Amato^{1,2}

¹ PLADEMA, Universidad Nacional del Centro de la Provincia de Buenos Aires,

² Consejo Nacional de Investigaciones Científicas y Técnicas, CONICET

³ Comisión de Investigaciones Científicas, CICPBA

Resumen. El trabajo presenta los resultados preliminares de un algoritmo de detección y eliminación de sombras, en secuencias de video. Se propone que a partir de la base de la sustracción de fondo con el algoritmo *Visual Background Extraction (ViBE)*, que identifica zonas de movimiento, aplicar un post-procesamiento para separar los píxeles del objeto real y los de sombra. Debido a que las zonas de sombras tienen características parecidas a la de los objetos en movimiento, la separación es una tarea difícil, por lo que los algoritmos usados para esta clasificación producen gran cantidad de falsos positivos. Para atacar este problema, se parte de la premisa de utilizar información del objeto, como el tamaño y la dirección de movimiento, para estimar la posición aproximada de la sombra. Además, se realiza el análisis de las similitudes entre el cuadro actual y el modelo de fondo, con el indicador tradicional de correlación cruzada normalizada para detectar sombras. El algoritmo se puede aplicar para detectar personas o vehículos en aplicaciones para seguridad ciudadana, monitoreo de tráfico, análisis deportivo, entre otros. Los resultados obtenidos en la detección de objetos muestran que es factible recortar la sombra con alta tasa de acierto y con bajo costo computacional, lo cual también permite mejorar etapas de procesamiento posteriores, como el reconocimiento y el seguimiento de los objetos.

Palabras clave: procesamiento de video, detección de objetos, segmentación

1 Introducción

Actualmente, existe un gran interés tanto en la industria como en la investigación en contar con sistemas de video análisis. En el área de seguridad, se ha mostrado que el video digital es una herramienta fundamental tanto para la pericia forense, como para la disuasión del delito. En particular, en Argentina la inseguridad es un tema crítico que requiere su atención inmediata, para lo cual ha crecido la cantidad de centros de monitoreo, con gran cantidad de operadores observando cámaras. Ciertamente, este mecanismo no es eficiente por lo que es importante contar con técnicas de análisis de video que ayuden a los operadores en las diferentes tareas cotidianas. En esta línea, este grupo de investigación

ha presentado varios trabajos, principalmente mostrando la arquitectura de un sistema abierto, distribuido y escalable [1].

La complejidad de estos algoritmos radica en trabajar en ambientes dinámicos, como son las cámaras del exterior, donde sufren las inclemencias del clima. Para contrarrestar estos efectos, se suelen utilizar ventanas de movimiento como se menciona en [2], cuyo objetivo consiste en aplicar una operación lógica, para procesar únicamente los movimientos que se detecten dentro de sus dimensiones y descartar todos aquellos que se produzcan por fuera.

Estos algoritmos de seguimiento finalmente son utilizados en plataformas de video-análisis en virtud de detectar y analizar situaciones de interés, por lo que es crítico tener una alta tasa de confianza en la detección. En este contexto, la sombra afecta considerablemente la percepción de los objetos detectados, tal como se ve en la Figura 1, ya que altera el tamaño, color u orientación del mismo afectando drásticamente una posterior clasificación por tamaño o color; por lo que debe reducirse de alguna manera.



Fig. 1. Imágenes de diferentes casos, donde la sombra cambia el tamaño y forma del objeto.

El presente trabajo se enfoca en poder detectar y eliminar sombras a partir de videos en escala de grises o en el espacio color transformado a escala de gris (ej. `rgb2gray` o HSV correspondiente a la componente Value). Esta consideración también es adecuada ya que muchas cámaras de bajo costo a menudo no proveen una imagen de calidad suficiente, afectada también por la compresión de video. Lo novedoso del método propuesto es que utiliza información del objeto y de su orientación, para determinar *a priori* la posición de la sombra, para luego aplicar en la región externa al objeto un método tradicional que tiene en cuenta las características de textura y color en las zonas de sombras.

El trabajo se organiza de la siguiente manera. En la sección 2 se detalla el estado del arte, y en la sección 3 el método de sustracción de fondo ViBE y las modificaciones realizadas. En las secciones 4 y 5 se realiza la propuesta de detección de sombras y de resultados obtenidos. Finalmente en la sección 6 se presentan las conclusiones y futuros trabajos.

2 Estado del arte

Las técnicas de detección de eventos usados en sistemas de *Video-Surveillance*, se basan principalmente en discriminar rápidamente el movimiento a partir de una video-cámara fija. Trabajos como [3], detallan los algoritmos más comunes para la detección y seguimiento de objetos. En [4] se realiza una comparación interesante entre distintos algoritmos de detección básicos, en donde se concluye que la combinación de éstos puede ser muy útil para disminuir la tasa de falsos positivos, mientras se mantiene el tiempo y la tasa de verdaderos positivos.

Los sustractores de fondo usualmente usados clasifican las sombras como parte de los objetos, lo cual altera el tamaño y forma de los mismos, afectando en consecuencia la eficacia de tales algoritmos [5]. También, dificulta otros procesamiento que necesitan el resultado de la detección de objetos, como por ejemplo, la clasificación, o el seguimiento de la trayectoria y análisis de comportamiento (ej. merodeo, vandalismo, infracciones de tránsito, etc.). Este problema también afecta a técnicas basadas en características [6].

Existen varias investigaciones acerca de la detección y separación de sombras. Algunos algoritmos que trabajan sobre imágenes estáticas, son computacionalmente complejos y no aplicables para análisis de video en tiempo real [7]. A diferencia, otros algoritmos están diseñados especialmente para video en escala de grises o color, y tienen menor costo computacional. Para realizar la clasificación, la mayoría se basa en las características usuales de la región con sombras (zonas más oscuras que el fondo de la escena, color o textura uniforme e invariable, etc.). Aunque estas características permiten crear candidatos para determinar la zona de sombras, no son decisivas, lo cual causa que todos los métodos fallen en la clasificación, ya que estas propiedades suelen aparecer también sobre partes de los objetos [8][9]. En particular, otro grupo de algoritmos incluyen información sobre la geometría de la sombra o modelo de iluminación, algunos especializados solamente en sombra de personas [10].

3 Sustractor de fondo en video

En este trabajo se utiliza el sustractor de fondo ViBE, propuesto en [11], el cual tiene buen comportamiento en entornos variados como suelen ser típicos en videovigilancia. Entre las virtudes de este método se destacan el bajo tiempo de cómputo, las altas tasas de detección y la robustez ante la existencia de ruido, las cuales son necesarias en capturas de cámaras de supervisión utilizadas hoy en día. Igualmente, la propuesta puede llegar a aplicarse a otros algoritmos [3]. En la Figura 2, se muestra la sustracción de fondo realizada con ViBE para un *frame* particular de video de monitoreo urbano. En el ejemplo, se puede ver la imagen original y a su izquierda, las componentes *foreground* detectadas, personas y vehículos. La proyección de la sombra del ómnibus forma parte del objeto, deformando notablemente la silueta del mismo, y el tamaño se incrementa en aproximadamente un 50%.

El algoritmo ViBE utiliza un modelo representativo del fondo basado en píxeles (guarda aleatoriamente muestras del valor de intensidad de cada píxel

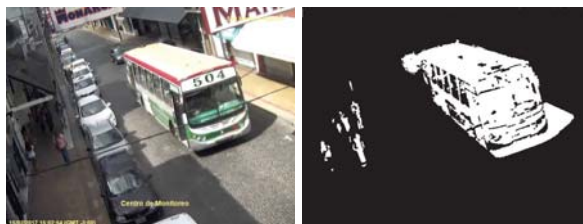


Fig. 2. Imagen del video (*izq.*) y sustracción del fondo con ViBE (*der.*), para un video registrado en cámara de monitoreo.

de la secuencia previa de video). En la Figura 3, se representa la secuencia de pasos del algoritmo ViBE original y los módulos incorporados para mejorar la clasificación. El módulo *Detección*, clasifica las escenas como *background* o *foreground*, calculando la distancia entre cada píxel de una imagen actual respecto de las muestras guardadas en el modelo de fondo.

La actualización del modelo de fondo para cada píxel es aleatoria, con cierta probabilidad de reemplazar uno de los valores guardados, por un nuevo valor de intensidad del mismo píxel en el *frame* actual (Figura 3 módulo *Actualizar modelo*). La aleatoriedad sobre este mecanismo de actualización permite reducir gradualmente con el tiempo, la probabilidad de que las muestras guardadas persistan en el modelo. Es muy importante que solamente aquellos píxeles procesados y clasificados como *background* reemplacen muestras en el modelo de fondo correspondiente. La inserción de píxeles mal clasificados, que pertenecen a objetos en movimiento (*foreground*), o que resultan inciertos en la clasificación, pueden alterar significativamente los resultados de la detección. Esto último es muy importante ya que ha permitido la introducción de mejoras al método ViBE original logrando mayor tasa de detección, por ejemplo para adaptar el algoritmo a los movimientos de escena o fondo dinámico (detectados como falsos positivos), o el rellenado de objetos detectados en forma incompleta (falsos negativos) [12][13]. En este trabajo, se agrega el algoritmo de operación morfológica de cierre de los objetos (dilatación, rellenado de agujeros), para lograr tener componentes conectadas, se seleccionan los objetos con tamaño considerado, y se descartan los pequeños y aislados (Módulo *Contornos y Rellenado*). Luego de este proceso, se aplica el algoritmo de separación de sombras del objeto real (Módulo *Sombras*).

Este módulo propone corregir los píxeles clasificados como *foreground* pero que se corresponden a sombras. Cabe aclarar también, que no resulta adecuado clasificar las sombras como *background*, ya que estas muestras no deberían modificar o alterar el modelo de fondo, cuando actualiza el modelo, para evitar errores posteriores en la clasificación.

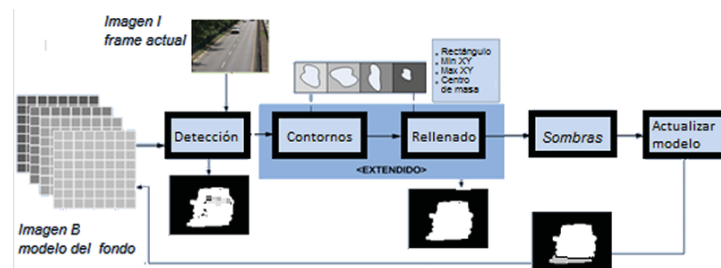


Fig. 3. Algoritmo ViBE y módulos de mejoras implementados

4 Propuesta de detección de sombras

Para la detección y eliminación de sombras por un lado se utiliza la correlación cruzada normalizada (NCC) entre la imagen actual (I) y una imagen representativa del fondo (B), ajustando con propiedades características de la sombra como en [14], pero además se incorpora al método original de NCC, conocimiento elemental sobre los objetos para mejorar la tasa de detección, detallando ambos a continuación.

4.1 Identificación de píxeles candidatos con NCC

El indicador NCC permite identificar imágenes similares con diferente escala de intensidad. Para representar al fondo (B), se utilizó una imagen con las muestras que ViBE selecciona aleatoriamente como representativas del modelo del fondo (Ver Figura 3).

Para cada píxel (i, j) clasificado como *foreground* por ViBE, se consideran los vecinos en una región R cuadrada, con centro en ese píxel. Los píxeles candidatos a clasificar como sombra son aquellos que tienen valor de correlación alto, teniendo en cuenta esta región de vecinos:

$$C(i, j) = \sum_{i=-N}^N \sum_{j=-N}^N (I(i, j) * B(i, j)) \quad (1)$$

donde $I(i, j)$ y $B(i, j)$ corresponde al valor de intensidad del píxel en la imagen I y B, respectivamente, y $(2N+1)^2$ es el tamaño de R . La correlación se normaliza según:

$$NCC(i, j) = \frac{C(i, j)}{\sqrt{mI(i, j)} * \sqrt{mB(i, j)}} \quad (2)$$

donde $mI(i, j)$ y $mB(i, j)$ son los momentos no centrados de orden 2 en la misma región R centrada en el píxel (i, j) , en la imagen I y B, respectivamente.

El valor de referencia para detectar sombras con NCC (Ec. 2) es un valor alto definido en el rango [0.95-0.98] [14], y se usa como paso inicial para detectar posibles candidatos.

Luego, se rectifican píxeles, usando estadística sobre la región R teniendo en cuenta el análisis de la relación entre los valores de intensidad de las imágenes I y B para cada pixel (i, j) :

$$\alpha(i, j) = \frac{I(i, j)}{B(i, j)} \quad (3)$$

Las zonas de sombras deben ajustarse a un rango definido de valores, donde α en la Ec. 3 se define usualmente entre $[0.4-1]$, dada la oscuridad de la sombra sobre el modelo de fondo [14]. También se considera, que la desviación estándar de la relación α calculada en la región R de vecinos, debe tener un valor bajo debido a que la zona de sombra es un área homogénea. Se recomienda un desvío estándar menor a 0.05 en la región R de tamaño 5×5 píxeles, siendo N igual a 2. El problema del método original [14] es que genera falsos negativos cuando el valor de intensidad del objeto coincide con los valores de la sombra.

En la Figura 4 se muestra un *frame* del video Pedestrians [15] (izquierda) y otro *frame* de un video de monitoreo urbano (derecha). En la primer fila de la Figura 4 se muestra el resultado de la sustracción de fondo con ViBE. Se puede observar en el caso de la persona caminando, que la sombra (falsos positivos) no está conectada al cuerpo y en el caso del vehículo la sombra se proyecta en la parte inferior, desformando el objeto. En la segunda fila de la Figura 4, se utilizaron parámetros recomendados para NCC (Ec. 2) mayor a 0.95, y para α en el rango $[0.4-1]$. Se marca la zona de sombras resultantes (en color amarillo). Se puede observar que no siempre la detección de sombra es correcta, como por ejemplo, sobre las piernas de la persona, y en los vidrios del vehículo, generando falsos negativos. El método no trabaja correctamente, principalmente en las escenas donde hay reflejos de luz, o el fondo se parece mucho a la sombra. En estos casos, fue necesario considerar diferentes valores de α y desvío estándar, dependiendo del ruido en la señal de video, y en casos donde la sombra es muy oscura.

4.2 Incorporación de conocimiento

En este trabajo se tuvo en cuenta no aplicar las reglas anteriores para NCC en zonas potenciales del objeto verdadero. En primer lugar, la clasificación de sombras se realiza a nivel de una región rectangular o *blob* detectado (componentes conectadas en la máscara *foreground* detectada por ViBE) y no de cada píxel aislado como en la versión original del método [14]. Se propone entonces, utilizar la información espacial dentro de esta región en la imagen I . En el caso de trabajar con videos de personas y/o vehículos a partir de video con cámara fija, se decidió además incorporar conocimiento de algunas propiedades de los mismos (ej. orientación, forma, tamaño).

En particular, al encerrar a las personas o los vehículos con una elipse, usando el centro de masa de la máscara *foreground*, es muy probable que la proyección de la sombra exceda los límites de la elipse, como en el caso del vehículo en la Figura 4. En otros casos, la proyección de la sombra puede quedar separada en una elipse diferente como en el caso de la persona caminando.



Fig. 4. Sustracción de fondo y detección de sombra. Imagen original (izq.) y Máscara resultante (derecha). Algoritmo ViBE (arriba). Algoritmo detección de sombras con NCC (centro) y Detección de sombras con algoritmo propuesto (abajo)

En cada *blob*, se calcula la distribución direccional para encontrar la elipse que mejor se ajusta a la distribución de píxeles *foreground*. Como resultado de este método, se define una elipse mediante el eje mayor, el eje menor, y la orientación del eje mayor respecto del eje horizontal [16]. Además, fue necesario adaptar estos valores para que cubran la totalidad de los diferentes objetos (tanto personas como vehículos), y no exceder demasiado los límites del *blob*. En forma heurística, resultó adecuado un factor de 1.8 para el eje mayor y de 1.6 para el eje menor.

Como primer intento de no invadir en el recorrido las zonas del objeto verdadero, se tuvo en cuenta que la altura y el ancho del cuerpo humano mantienen cierta proporción, y además que la persona camina erguida. De esta manera, la regla de decisión para determinar que en un *blob* hay una persona, es por el tamaño del área, la relación entre el diámetro mayor y menor de la elipse sea mayor igual a 3, y finalmente que la orientación del eje mayor se mantenga cercana a los 90. En el caso de la Figura 4 (arriba), se puede ver que la diagonal mayor de la elipse es casi vertical para el caso de la persona, y que la diagonal mayor de la elipse que encierra la sombra queda casi horizontal. Para identificar automóviles, se consideró que la orientación puede variar, sin embargo el tamaño del *blob* es mucho mayor que el que contiene personas, bicicletas o motos. En la Figura 4 (abajo), se puede ver que los resultados del algoritmo propuesto, mejoran respecto de algoritmo original de ViBE+ (Figura 4, arriba) y también de ViBE+[14] (Figura 4, centro). La persona que camina no tiene sombras sobre las piernas, y tampoco se observan sombras en los vidrios del vehículo (falsos negativos). Cuando se detecta sombra fuera de la elipse, el algoritmo continúa

sobre los píxeles vecinos, manteniendo la continuidad incluso si sobrepasa los límites de la elipse.

5 Resultados

El algoritmo propuesto fue testeado con los videos reales *Pedestrians* y *Highway* de la base de datos [15] [17], los cuales fueron adquiridos con cámaras estáticas para uso en el contexto de la videovigilancia. Estos videos cuentan con imágenes *groundtrue* para calcular la tasa de aciertos y comparar resultados. De esta forma, la librería de algoritmos de sustracción de fondo clásicos permite comparar con los métodos y contribuciones recientes en el área. En nuestro caso, se analizan los resultados obtenidos con otros métodos estocásticos tradicionales como VIBE+ y Gaussian Mixture Model (GMM)[3].

Para cada método se utilizaron dos métricas evaluación conocidas *Precision*, *Recall* según:

$$Precision = \frac{TP}{(TP+FP)} \quad (4)$$

$$Recall = \frac{TP}{(TP+FN)} \quad (5)$$

siendo FN el número de falsos negativos, FP el número de falsos positivos, TP es el total de positivos y TN el total de negativos. En este caso, el resultado es mejor, cuanto más cercana a 1 es la métrica. También se consideró un indicador sobre errores en la sombra como,

$$TShadow = \frac{nbErrorShadows}{FP} \quad (6)$$

siendo nbErrorShadows el total de Falsos Positivos que se producen en zona de sombras según el *groundtrue*, y FP el número de falsos positivos.

En la Tabla 1, se muestran los valores de las ecuaciones anteriores para aproximadamente 700 imágenes del video *Pedestrians*, y en la Tabla 2 los valores para 1300 imágenes de *Highway*.

Métodos	Recall	Precision	Tshadow
GMM [18]	0.98	0.93	0.22
ViBE+ [11]	0.93	0.96	0.50
ViBE+NCC [14]	0.83	0.99	0.11
Propuesto ViBE+NCC+Información	0.91	0.99	0.07

Table 1. Valores de indicadores para cada método con el video Pedestrians

Se puede observar para ambos videos *Pedestrians* y *Highway*, que con los método ViBE+ o GMM, el porcentaje TShadow es mucho mayor (cercano al

Métodos	Recall	Precision	Tshadow
GMM [18]	0.89	0.91	0.81
ViBE+ [11]	0.84	0.92	0.95
ViBE+NCC [14]	0.58	0.94	0.47
Propuesto ViBE+NCC+Información	0.85	0.94	0.55

Table 2. Valores de indicadores para cada método con el video Highway

doble) que con el método propuesto. El método ViBE+NCC[14] aumenta la cantidad de falsos negativos haciendo que la métrica *Recall* disminuya considerablemente respecto de ViBE+ (en *Pedestrians* de 0.93 a 0.83 y en *Highway* de 0.84 a 0.58). También en ambos videos con el método propuesto, se nota un aumento de la tasa *Precision* respecto de ViBE+, indicado una reducción de la cantidad de falsos positivos. Finalmente, con el método propuesto ViBE+NCC+información, la tasa *Recall* y *Precision* es comparable a la de los métodos ViBE+ y GMM, sin embargo, el porcentaje de mal clasificados en zona de sombras con *TShadows* es notablemente menor.

6 Conclusiones

En este trabajo, se han mostrado los resultados preliminares del método de detección y separación de la sombra, basado en información contextual, los cuales han sido promisorios. Se ha podido procesar los videos, y visualizar la clasificación en tiempo real, ya que algoritmo tiene bajo costo computacional.

Se ha logrado disminuir el error en la clasificación respecto de la aplicación de NCC tradicional, calculando además la orientación y ubicación probable del objeto, la determinación de persona o vehículo.

En futuros trabajos se pretende incorporar la adaptación automática de los umbrales del rango para α , teniendo en cuenta la variación de intensidad de la sombra al procesar el video, ya que el algoritmo tiene inconvenientes en detectar zonas de sombra muy oscuras. El otro desafío es que los parámetros del algoritmo se adapten automáticamente a diferentes momentos del día y al mismo tiempo, comparando cuantitativamente con otros métodos de referencia.

References

1. Dominguez L., Perez A., Rubiales, A., Damato J., and Barbuzza R.: Herramientas para la detección y seguimiento de personas a partir de cámaras de seguridad. In: Proceedings del XXII Congreso Argentino de Ciencias de la Computación, pp. 251–260, (2016).
2. Kruegle H. ,CCTV Surveillance: Video practices and technology, Butterworth-Heinemann, (2014).
3. Legua C. C.,Seguimiento automático de objetos en sistemas con múltiples cámaras (2013).
4. Shaikh S. H. and Saeed K. and Chaki N. ,Moving Object Detection Using Background Subtraction, Springer (2014).

5. Azab, M., Shedeed, H., Hussein, A., A new technique for background modeling and subtraction for motion detection in real-time videos. In: IEEE International Conference on Image Processing (ICIP), pp. 3453-3456 (2010).
6. Raad Ahmed Hadi, Ghazali Sulong, and Loay Edwar George; Vehicle detection and tracking techniques: A concise review, Signal and Image Processing : An International Journal (SIPIJ) Vol.5, No.1, (2014).
7. Fredembach, C., Finlayson, G.,: Simple shadow removal. In: Proceedings of the 18th International Conference on Pattern Recognition (ICPR06), pp. 832–835, (2006)
8. Qin R. and S. Liao and Z. Lei and S. Z. Li: Moving Cast Shadow Removal Based on Local Descriptors. In: 2010 20th International Conference on Pattern Recognition, pp. 1377–1380, (2010)
9. Chen F., Zhu and B., Jing W. and Yuan L. : Removal shadow with background subtraction model ViBE algorithm. In: 2013 2nd International Symposium on Instrumentation and Measurement, Sensor Network and Automation (IMSNA), pp. 264–269,(2013)
10. Bi S., Liang D., Shen X. and Wang Q.: Human Cast Shadow Elimination Method Based on Orientation Information Measures. In: IEEE International Conference on Automation and Logistics, Jinan, pp. 1567-1571, (2007)
11. Barnich O. and Van Droogenbroeck M.: ViBe: A Universal Background Subtraction Algorithm for Video Sequences. In: IEEE Transactions on Image Processing, ISSN 1057-7149, 20 (6), pp. 1709–1724. (2011)
12. Gervasoni L., Damato J., Barbuzza R., Vénere M.: Un método eficiente para la sustracción de fondo en videos usando GPU. In: Mecánica Computacional, Vol 33, pp. 1721–1731, ISSN 1666-6070, AMCA, Buenos Aires, (2014)
13. Barbuzza R., Damato J., Rubiales A. Dominguez L, Perez A., and Venere M.: Un método para la sustracción de fondo en videos inestables. In: Mecánica Computacional, Vol 34, pp 3409–3417, (2016)
14. Jacques J. C. S., Jung C. R. and Musse S. R.: Background Subtraction and Shadow Detection in Grayscale Video Sequences. In: XVIII Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI'05), pp. 189–196. IEEE Computer Society, ISSN 1530-1834, Washington DC, USA (2005)
15. Goyette N., Jodoin P. M., Porikli F., Konrad J. and Ishwar P.: Changedetection.net: A new change detection benchmark dataset. In: 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 20 (6), pp. 1–8, (2012)
16. Lahoz Beltra, R., Ortega J. and Fernandez Montraveta C.: Métodos estadísticos en biología del comportamiento: Editorial Complutense, ISBN 84-7491-512-0, (1994)
17. Sobral A. Vacavant A. A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos. In: Computer Vision and Image Understanding, Vol. 122, Elsevier, pp. 4–21, (2014)
18. Sobral A. Vacavant A. Improved adaptive Gaussian mixture model for background subtraction. In: Proc. Int. Conf. Pattern Recognition, IEEE, Piscataway, NJ, pp. 28–31 (2004)

Virtual Trainer for Physical Activities using a Natural User Interface

Nicolás Jofré, Graciela Rodríguez, Yoselie Alvarado, Jacqueline Fernandez, and Roberto Guerrero

Laboratorio de Computación Gráfica (LCG)
Universidad Nacional de San Luis,
Ejército de los Andes 950
Tel: 02664 420823, San Luis, Argentina
{npasinetti, gbrodriguez, ymalvarado, jmfer, rag}@unsl.edu.ar

Abstract. Despite the known health benefits of regular participation in physical activity people still refuse to practise. Nowadays, Virtual Reality (VR) is a very powerful and compelling computer tool by which humans can interface and interact with computer-generated environments. In this paper, we propose a virtual training system which can be customized for the physical activity level of the user. The system provides real time visual action guides and a performance's feedback of users through a Natural User Interface (NUI). We conduct a brief pilot study to evaluate our virtual trainers in which participants' performance doing physical activities is evaluated via our NUI. Initial results indicate that virtual training through a NUI is motivating and entertaining for any kind of user, in particular for individuals with a low level of physical activity.

Keywords: Virtual Reality (VR), Embodied Conversational Agent (ECA), Natural User Interface (NUI), Virtual Physical Activities, Kinect.

1 Introduction

Virtual Reality technology has become a very popular technology integrating the newest research achievements in the fields of computer graphics, sensor technology, ergonomics and Human-Computer Interaction (HCI) theory. As a research tool, Virtual Reality provides numerous opportunities of what can be done and seen in a virtual world that is not possible in the real world [1, 2].

In the HCI area, Embodied Conversational Characters have emerged as a specific type of multimodal interface, where the system is represented as a person conveying information to human users via multiple modalities such as voice and hand gestures, where the internal representation is modality-independent, both propositional and non-propositional. Embodied Conversational Characters answer questions and perform tasks through interaction in natural language-style dialogues with users contrasting the traditional view of computers [3].

Virtual human research has progressed rapidly over the last 18 years. The first implemented work of this type appeared in 2008 and many works of Embodied Conversational Characters have recently emerged due to interfaces have

great potential as smart assistants, travel agents or investment advisors, among others [4–8].

Verbal and non-verbal behavior of virtual characters has become more and more sophisticated as a result of advances in behavior planning and rendering [9]. VR uses these characters for user-environment interaction, where system must collect gestural, positional, sound and biometric user information through use of sophisticated devices [10]. These devices have incremented their simplicity and user-friendliness over time, improving interaction and giving rise to Natural User Interfaces [11]. These interfaces are used to solve problems in many areas such as medicine, robotics, non-verbal communication, among others.

In last years sedentary lifestyle has become an increasing problem for people's health. It is reported to be indeed the cause of several serious illnesses like obesity, diabetes, hypertension and so on. There are many reasons why people do not perform any physical activity: motivational lack, time constraints, difficulties to start, gym membership fees, equipment costs, among others. Physical activity is defined as any bodily movement produced by skeletal muscles that results in energy expenditure.

In order to provide a more natural interaction between humans and machines, HCI works have brought about the wide focus on body motion recognition. This involves the recognition of physical movement of the head, hands, arms, face or body with the aim of conveying kinesic information [12].

In this context, there are some commercial applications like *Nintendo* or *Xbox Games* which provide different entertainment activities [13, 14]. These applications track user's body movements, such as knees, arms, legs, waist, hip, among others. However, the aim is not focused on obtaining a healthier lifestyle and training is not personalized.

There exist many physical activities research works in HCI [15–19]. From previous studies, the existing work technologies for this purpose are smart phone applications, sensor devices and image processing, with a few solutions implementing virtual characters [20–22]. Those last-mentioned share in common a low-cost virtual training computer application for users who need to undertake regular physical exercise or cognitive tasks at home, with identified specific shortcomings in user's interface, interaction's place and evaluating methods.

In this work we present a training system based on an Embodied Conversational Agent as a real-time simulation approach to train and guide people's physical activities through a NUI.

This paper is arranged as: a design guidelines of our system is introduced in Section 2. Our system is described in Section 3. The system architecture is explained and illustrated in Section 4. The evaluation and results are shown in Section 5. We conclude our paper and propose some future works in Section 6.

2 Design guidelines

In 2005, researchers from different Universities of the United States provided a state of the art of pervasive computing in sport technologies, with the aim to

encourage new research in this emerging area and to describe how technology can be applied to sports [23]. They identified three areas of application - *athletic performance*, *entertainment* and *support for referees* - pointing out some research lines for each, such as studying which sensors are more appropriate to monitor the performance of different kinds of athletes and studying the use of games and special equipment to encourage users to exercise harder.

As was stated in Section 1, our work's proposal is the study result of several works related to virtual trainers [20–22, 24]. From this, two relevant concepts have been identified: *standard physical activities* (considered by a lot of works) and *learning methods* (used by many similar systems).

2.1 Standard Physical Activities

Levels of participation in physical activity remain low across many age groups, and some strategies to increase activity levels throughout the population are needed. Particularly, is possible to perform a generic levels classification on two groups:

1. Average-to-High Level: people with active jobs or that occasionally play sports and people considering physical activity as regular.
2. Low Level: people with office's work or with relaxing daily life activities.

It is necessary and important to identify and use strategies for making physical activities that are both effective and cost-effective. In general, physical activities can be distinguished in **flexibility exercises** (stretch, do yoga), **muscle fitness exercises** (wall climb, use exercise bands), **vigorous sports and recreation** (play sports, hike, play active games), **vigorous aerobics** (ride a bike, jog, jump rope), **moderate physical activity** (walk, play games, do yard work).

At the same time, these activities can be associated to levels of participation in physical activity of people as follows:

- Average-to-High Level: flexibility exercises, muscle fitness exercises, vigorous sports and recreation, and vigorous aerobics.
- Low Level: flexibility exercises and moderate physical activity.

2.2 Learning Method

Virtual-reality-enhanced interactive learning environments are increasingly common. VR brings together a mixture of virtual and real-life scenarios for a wide range of potential possibilities in teaching and learning.

In this context, *Game Based Learning* (GBL) is a learning method leveraging the power of computer games to captivate and engage end users for a specific purpose, such as to develop new knowledge and skills [25]. From past reviewed studies, the minimum components required in all online games for learning are:

back story and story line, game mechanics, rules, immersive graphical environment, interactivity, challenge/competition and risks [26]. Additionally, many educators use this method in combination with the following metaphors for an educational game [27]:

- Acquisition: to transfer information from who owns it (the teacher) to a passive receiver (the student).
- Imitation: focuses on imitation of model behaviors through observing the reactions of others to the facts.
- Experimentation: it is applied in the learning of specific activities, complex or dangerous tasks, since it promotes active and contextualized learning processes, mainly related to practical activities and physical abilities to a great extent.
- Participation: the transmitted content by the teacher is taken as a learning stimulus, which occurs naturally and difficult to predict.
- Discovery: learning by discovery can be an individual or social activity; the crucial point is that it creates new contents through the active student's participation.

3 System Description

The basic idea was to simulate an environment where the user can train in a green space accompanied by a virtual trainer.

For user-system interaction, a sport field scenario was modeled and rendered. Scenario was set up with ambient sound, inanimated objects (benches, lights, trees) and animated objects (people). Particularly, people on stage are avatars. The user can navigate through the park, paddle court and other places allowing social interaction. The trainer is a female animated character called "*Sara*". This was developed to provide a visual action guide in real time and an actions assessment after physical users' exercises.

As was mentioned in Section 2.2, a computer educational game requires certain components and can use several metaphors. While our system allows physical training, it is also an intention to add entertainment to interaction. Additionally, as it has seen in Section 2.1, it is possible to perform a generic level classification and to associate certain activities for each level. Thus, some components and metaphors will be indicated along this section.

According to the referred levels, this work established the following particular activities:

- Average-to-High Level: single-leg squats, mountain climbers, tuck jumps and burpees.
- Low Level: squats, lying hip raises, side lunges and frog jumps.

At the interaction's beginning, *Sara*'s face representation welcomes the user and explains him about system's operation (*Acquisition*). The user have to

choose one participation level and consequently, the associated level's activities from a *Graphic User Interface* (GUI). The act of choosing a level and an activity implies to enter voice data through a microphone. It is assumed that the user will choose a level according to his/her daily life activities.

After selection, system will show *Sara* in full body to start teaching user selected activity's actions. User's actions are captured and compared to standard *Sara*'s actions (see Figure 1) while giving, in real time, a fitness score representing how well user performs actions (*Challenge*), besides providing an activity completeness percentage. *Sara* will then provide interpreted feedback based on obtained fitness score and how user can improve his/her actions (*Imitation* and *Experimentation*).

Considering that each physical exercise is related to a bodily movements set which have limbs' degrees associated (information provided by real trainer stored in a database), system provides a fitness score (percentage of mistakes 0% - 100%) comparing input movements degrees with stored movements degrees.



Fig. 1. System Operation.

The system was developed to work on a computing platform for immersive collaborative 3D virtual world visualization (*Immersive Graphical CAVE-like environment*). This platform allows the user to play an appropriate role (through a character) during system's explanation, physical activity and feedback according with conversational aspects of virtual trainer (which results in a *Verbal Natural User Interface*).

User's actions are captured by a motion sensing input device [28]. The motion device is able to detect the shape of a human body by using an RGB camera and depth sensor (*Skeleton Tracking*). User and bodily sensor must be located inside CAVE-like environment for an enhanced interaction (*interactivity*) and a better recognition of user movements during physical activity (resulting in a *Gestural Natural User Interface*).

4 System Architecture

The implemented training system is an integrated system consisting of three subsystems working in an independent parallel way: a *CAVE-like Subsystem*, a *Conversational Character Subsystem* and a *Motion-NUI Subsystem*. The system comprises the hardware and software necessary to gather the information obtained during the interaction between the user and the training system: via a motion sensing device, sound system, microphone, screen/projection surfaces and projectors, among others.

In the *CAVE-like Subsystem* (immersive environment), the user's movements are captured by the *Motion-NUI Subsystem* which sets in motion the avatar animation according with the user movements and notifies to *Conversational Character Subsystem* a percentage of mistakes done by body area. According to the obtained percentage the character advises users how well they are doing the exercises. Figure 2 shows an overview of training system.

CAVE-like SubSystem will provide the necessary structure for attributes definition, rendering and collaborative multi-visualizations, as well as the needed interactive resources. *Cake-like multi-VRmedia Subsystem* is an approach on a computing platform for immersive collaborative visualization of 3D and dynamic system proposed by [29].

Conversational Character SubSystem is composed by a Conversational Character gifted with human figure animation and verbal communication skills, like natural language processing (by *Automatic Speech Recognition mechanism* (ASR)) and speech recognition and synthesis (by *Text-To-Speech mechanism* (TTS)). This subsystem, named *CAVE-VOX* was presented and evaluated in [7, 30] and it was based on previous work [31]. This particular Conversational Character has a collection of responses relevant to a physical activities topic.

Motion-NUI SubSystem is a NUI based on gestures that corresponds with bodily movements. This subsystem obtains, processes and replies information about user's movements. User data is collected through a motion device, which uses a depth sensor and obtains a 3D input based on user location and posture. The 3D input is utilized for two aspects: user movements analysis and animation data matching. User movements analysis calculates the difference between 3D input and data stored from a predefined data base, and animation data matching calculates the correspondence between 3D input (user body) and user's virtual avatar (virtual body). After animation data matching an avatar animation is set up for reply the user movements.

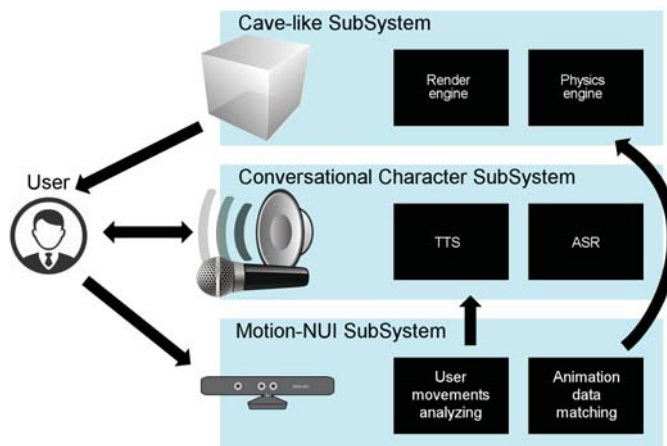


Fig. 2. System Architecture.

5 System Evaluation

A pilot test was applied to a twenty-two participants group between 16 and 60 years old. At first, the system functions were explained and demonstrated to the users undertaking the testing. Every user was then required to explore every function and try it out by him/herself.

As was mentioned in Section 3, participants must be classified in some level. From the total participants, 8 chose the Average-to-High Level and 14 the Low Level.

During the test, each participant is required twice to execute some corresponding level's activity. Each activity consisted of three sets of exercises: a set of 15 repetitions (first set), a set of 30 repetitions (second set) and a set of 40 repetitions (third set).

Each execution was evaluated in different days (at least two days in between) due to it is not recommended anyone exercise more than once in two days [32]. According with each execution an average percentage of completeness by set were recorded (See Table 1 and 2).

Low Level			
attempt	first set	second set	third set
1	82%	73%	66%
2	96%	86%	89%

Table 1. Completeness' averages for Low Level

Average-to-High Level			
attempt	first set	second set	third set
1	95%	87%	71%
2	97%	96%	94%

Table 2. Completeness' averages for Average-to-High Level

For Low Level, the percentage of completeness increased significantly as increasing the execution allowing to achieve an incremental learning. For Average-to-High Level, the percentage of completeness gradually increased with each execution, however, the difference among third set and the others was not overwhelming.

6 Conclusions and Future Works

This paper involved the development and evaluation of a training system based in an ECA. We described the design and learning aspects, along with the standard physical activities considered for the implemented system. We introduced the general and most important issues of the training system like, *CAVE-like*, *Conversational Character*, and *Motion-NUI SubSystem*.

The obtained results show that people doing regular physical exercises had no difficulties with virtual exercising but they had an appreciable period of adjustment during first execution. On the other hand, people in Low Level had considerable difficulties and a major period of adjustment on all testing. As regards period of adjustment, we considered that it is necessary to make usability tests and include interviews with children.

Future works will be oriented to:

- improve user's experience, adding new challenges related to physical activities,
- develop additional evaluations of the existing virtual training system,
- extend the pilot test.

References

1. W. Sherman and A. Craig. *Understanding virtual reality: Interface, application, and design*. Elsevier, 2002.
2. J. Blascovich and J. Bailenson. *Infinite reality: Avatars, eternal life, new worlds, and the dawn of the virtual revolution*. William Morrow & Co, 2011.
3. M. Mancini and C. Pelachaud. Generating distinctive behavior for embodied conversational agents. *Journal on Multimodal User Interfaces*, 3(4):249–261, 2009.
4. J. Cassell, T. Bickmore, L. Campbell, H. Vilhjálmsón, and H. Yan. Embodied conversational agents. chapter Human Conversation As a System Framework: Designing Embodied Conversational Agents, pages 29–63. MIT Press, Cambridge, MA, USA, 2000.

5. M. Mancini. *Multimodal Distinctive Behavior for Expressive Embodied Conversational Agents*. Universal Publishers, 2008.
6. Q. Chen, P. Torroni, S. Villata, J. Hsu, and A. Omicini. *PRIMA 2015: Principles and Practice of Multi-Agent Systems: 18th International Conference, Bertinoro, Italy, October 26-30, 2015, Proceedings*. Lecture Notes in Computer Science. Springer International Publishing, 2015.
7. N. Jofré, G. Rodríguez, Y. Alvarado, J. Fernández, and R. Guerrero. Virtual humans conversational character for a cave-like environment. In *XX Congreso Argentino de Ciencias de la Computación*, volume 20, pages 937–946, 2014.
8. J. Pérez, E. Cerezo, and F. Serón. E-vox: A socially enhanced semantic eca. In *Proceedings of the International Workshop on Social Learning and Multimodal Interaction for Designing Artificial Agents*, DAA '16, pages 2:1–2:6, New York, NY, USA, 2016. ACM.
9. S. Kopp and I. Wachsmuth. Model-based animation of co-verbal gesture. In *Computer animation, 2002. proceedings of*, pages 252–257. IEEE, 2002.
10. G. Burdea and P. Coiffet. Virtual reality technology. *International Journal of e-Collaboration*, 2(1):61–64, 2006.
11. J. Blake. *Natural User Interfaces in . Net*. Manning Pubs Co Series. Manning Publications Company, 2012.
12. J. Koenemann, F. Burget, and M. Bennewitz. Real-time imitation of human whole-body motions by humanoids. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2806–2812, May 2014.
13. D. Sheff. *Game over: How Nintendo conquered the world*. Vintage, 2011.
14. J. Andrews and N. Baker. Xbox 360 system architecture. *IEEE micro*, 26(2):25–37, 2006.
15. K. Miller, B. Adair, A. Pearce, C. Said, E. Ozanne, and M. Morris. Effectiveness and feasibility of virtual reality and gaming system use at home by older adults for enabling physical activity to improve health-related domains: a systematic review. *Age and ageing*, 43(2):188–195, 2013.
16. R. Mellecker and A. McManus. Active video games and physical activity recommendations: A comparison of the gamercize stepper, xbox kinect and xavix j-mat. *Journal of science and medicine in sport*, 17(3):288–292, 2014.
17. D. Nathan, D. Huynh, J. Rubenson, and M. Rosenberg. Estimating physical activity energy expenditure with the kinect sensor in an exergaming environment. *PLoS one*, 10(5):e0127113, 2015.
18. C. Campos and H. del Castillo. The benefits of active video games for educational and physical activity approaches: A systematic review. *Journal of New Approaches in Educational Research*, 5(2):115, 2016.
19. N. Jofré, G. Rodríguez, Y. Alvarado, J. Fernández, and R. Guerrero. El uso de la realidad virtual inmersiva en terapias motrices. In *XIX Workshop de Investigadores en Ciencias de la Computación*, 2017.
20. B. Li, M. Maxwell, D. Leightley, A. Lindsay, W. Johnson, and A. Ruck. Development of exergame-based virtual trainer for physical therapy using kinect. In *Games for Health 2014*, pages 79–88. Springer, 2014.
21. M. Pedraza-Hueso, S. Martín-Calzón, F. Díaz-Pernas, and M. Martínez-Zarzuela. Rehabilitation using kinect-based games and virtual reality. *Procedia Computer Science*, 75:161–168, 2015.
22. L. Shaw, R. Turrel, B. Wunsche, C. Lutteroth, S. Marks, and J. Buckley. Design of a virtual trainer for exergaming. In *Proceedings of the Australasian Computer Science Week Multiconference*, page 63. ACM, 2016.

23. E. Chi, G. Borriello, G. Hunt, and N. Davies. Guest editors' introduction: Pervasive computing in sports technologies. *IEEE Pervasive Computing*, 4(3):22–25, 2005.
24. R. Cózar-Gutiérrez and J. Sáez-López. Game-based learning and gamification in initial teacher training in the social sciences: an experiment with minecrafteredu. *International Journal of Educational Technology in Higher Education*, 13(1):2, 2016.
25. P. Tan, S. Ling, and C. Ting. Adaptive digital game-based learning framework. In *Proceedings of the 2nd international conference on Digital interactive media in entertainment and arts*, pages 142–146. ACM, 2007.
26. A. Derryberry. Serious games: Online games for learning. 2013.
27. P. Simons and M. Ruijters. Varieties of work related learning. *International Journal of Educational Research*, 47(4):241–251, 2008.
28. Zhengyou Zhang. Microsoft kinect sensor and its effect. *IEEE multimedia*, 19(2):4–10, 2012.
29. Y. Alvarado, N. Moyano, D. Quiroga, J. Fernández, and R. Guerrero. *Augmented Virtual Realities for Social Developments. Experiences between Europe and Latin America*, chapter A Virtual Reality Computing Platform for Real Time 3D Visualization, pages 214–231. Universidad de Belgrano, 2014.
30. Y. Alvarado, G. Claudia, V. Gil Costa, and R. Guerrero. *Computer Science & Technology Series XXII Argentine Congress of Computer Science-Selected Papers*,, chapter ARENA Simulation Model of a Conversational Characters Speech System, pages 119–129. Universidad de La Plata, 2017.
31. F. Serón, C. Bobed, and P. Latorre. *Augmented Virtual Realities for Social Developments. Experiences between Europe and Latin America*, chapter The Vox System, pages 130–159. Universidad de Belgrano, 2014.
32. World Health Organization et al. Recomendaciones mundiales sobre actividad física para la salud. 2010.

XIV Workshop Bases de Datos y Minería de Datos (WBDDM)

Coordinadores

Alfonso, Hugo – Bertone, Rodolfo – Reyes, Nora

Comité Científico

Bender, Cristina (Argentina)	Leguizamón, Guillermo (Argentina)
Brisaboa, Nieves R. (España)	Luaces, Miguel R. (España)
Chávez, Edgar (México)	Motz, Regina (Uruguay)
Corbalán, Leonardo (Argentina)	Olivas Varela, José Ángel (España)
Cuevas, Alfredo Simón (Cuba)	Paramá Gabía, José Ramón (España)
Deco, Claudia Deco (Argentina)	Paredes, Rodrigo (Chile)
Errecalde, Marcelo (Argentina)	Pedreira, Óscar (España)
Figueroa, Karina (México)	Pollo Cattaneo, María Florencia (Argentina)
Gagliardi, Edilma Olinda (Argentina)	Romero, Francisco Pascual (España)
Hasperué, Waldo (Argentina)	Seco, Diego (Chile)
Herrera, Norma (Argentina)	Serrano, Jesús (España)
Lanzarini, Laura (Argentina)	Sposito, Osvaldo (Argentina)

Automatización de la extracción de características en tareas de análisis de sentimiento

Juan M. Rodríguez^{1,2}, Hernán D. Merlino², Patricia Pesado¹, Ramón García-Martínez²

¹ Programa de Doctorado en Ciencias Informáticas. Facultad de Informática.
Universidad Nacional de La Plata. Argentina.

² Grupo de Investigación en Sistemas de Información. Departamento de Desarrollo Productivo
y Tecnológico. Universidad Nacional de Lanús. Argentina.
jmrodriguez1982@gmail.com, hmerlino@gmail.com,
ppesado@lidi.info.unlp.edu.ar, rgm1960@yahoo.com

Resumen. El siguiente artículo propone la utilización de un método de extracción de conocimiento para la Web (OIE), en particular ClausIE, para la obtención de características de películas de forma automática. En el marco de métodos de generación de resúmenes automáticos y de análisis de sentimiento, se compara este enfoque con otros dos en los cuales se utilizan pasos manuales para obtener las características de un servicio o producto. Los resultados obtenidos indican que ClausIE puede utilizarse para la extracción de características de forma semiautomática, requiere una intervención manual mínima que se explica en los resultados.

Palabras Clave. Análisis de sentimientos, extracción de características, extracción de conocimiento, extracción de relaciones semánticas, *open information extraction*, procesamiento de lenguaje natural.

1 Introducción

La tarea de realizar minería sobre críticas de cine y obtener un resumen de forma automática consiste en obtener primeramente el par: característica-opinión analizando una por una las oraciones constituyentes de la crítica para luego identificar la polaridad (positiva o negativa) de cada opinión. Y finalmente en construir una lista estructurada basada en las características y opiniones encontradas, promediando las polaridades de cada opinión para cada una de las características [Zhuang et al., 2006]. El presente trabajo se enfoca principalmente en mejorar la primera de las tareas mencionadas, es decir la identificación de características y palabras que expresan opiniones, pero principalmente en las características.

Las características, también llamadas aspectos, son elementos individuales que forman parte de una entidad mayor, siendo cada uno de ellos susceptibles de ser evaluado de forma independiente. Por ejemplo un restaurante tiene las siguientes características: comida, ambiente, servicio, precio. Incluso si se sabe que se está hablando de un restaurante en particular, que ofrece un plato particular como podría

ser: milanesas napolitana con papas fritas, este platillo puede constituir una característica.

La diferencia principal entre el análisis de sentimiento sobre críticas y el resumen automático de críticas con análisis de sentimientos radica en que en el primer caso solo se busca la polaridad global de un texto dado (la crítica) mientras que en el segundo se extraen de dicho texto la características principales y se evalúa la polaridad de cada una individualmente.

Los aspectos juegan un papel importante en el análisis de sentimiento ya que si bien es muy valioso contar con la opinión general sobre una cuestión, la revisión de aspectos o rasgos individuales juega un rol fundamental en la toma de decisiones, un ejemplo clásico es la revisión de un producto, en donde muchas veces un solo aspecto es decisivo para que el usuario decida, por ejemplo, comprarlo (típicamente el precio y/o la calidad).

En este trabajo se puso el foco en la extracción de características. Se buscó una solución automática basada en el uso de un método de extracción de conocimiento creado para la Web, o como se los llama en inglés: *Open Information Extraction* (OIE). En particular se planteó una solución basada en el método ClausIE [Del Corro&Gemulla, 2013].

1.1 Introducción a los métodos de extracción de conocimiento creados para la Web (OIE)

Extracción de conocimiento es cualquier técnica mediante la cual un proceso automatizable es capaz de analizar fuentes de información no estructurada, como por ejemplo textos escritos en lenguaje natural y extraer el conocimiento allí embebido para representarlo de una manera estructurada, manipulable en procesos de razonamiento automático, como por ejemplo: una regla de producción o un subgrafo en una red semántica. A la información obtenida como salida de este tipo de procesos se la llama: pieza de conocimiento [García-Martínez & Britos, 2004; Gómez et al., 1997]

En el año 2007 Michele Banko introduce un nuevo concepto en materia de extracción de conocimiento, al que llama en inglés: *Open Information Extraction* (OIE). Se trata de un paradigma de extracción de conocimiento en donde un sistema informático realiza una sola pasada sobre el total de las fuentes de información no estructurada en formato de lenguaje natural (llamado *corpus* de documentos), dadas como entrada y extrae un gran conjunto de tuplas relacionales sin requerir ningún tipo de participación humana. En el mismo trabajo Banko presenta un método llamado TEXT RUNNER, el cual es el primer método que trabaja dentro de este nuevo paradigma [Banko et al., 2007].

A partir de este trabajo se propusieron otros métodos de extracción de conocimiento bajo el paradigma que Banko llamó *Open Information Extraction* y que podríamos identificar de forma más concreta como métodos de extracción de conocimiento para la Web.

Los métodos de extracción de relaciones semánticas que trabajan de acuerdo con el paradigma anterior (OIE) devuelven una tupla para cada relación semántica

descubierta. La tupla tiene la forma (Entidad 1, Relación, Entidad 2), donde las entidades suelen ser objetos bien identificados, personas, lugares, empresas, fechas, etc., y la relación es la relación semántica entre las dos entidades, por lo general información fáctica del tipo: "Quién hizo qué a quién". Para ilustrar esto, considérese la siguiente oración en idioma inglés:

"Albert Einstein, que nació en Ulm, ha ganado el Premio Nobel".

Extrayendo las relaciones semánticas presentes en la oración y expresándolas como una tupla en la forma: "(Entidad 1, Relación, Entidad 2)" obtenemos lo siguiente:

- (Albert Einstein, ha ganado, el Premio Nobel)
- (Albert Einstein, nació en, Ulm)

1.2 Método escogido: ClausIE

En [Rodríguez et. al., 2015] se realizó una investigación documental sobre distintos métodos de extracción de relaciones semánticas para la Web y se encontró que ClausIE era, según sus autores [Del Corro&Gemulla, 2013] el método que lograba una mejor precisión. Esta aseveración fue puesta a prueba en [Rodríguez et. al., 2016] en donde se hizo una publicación parcial del resultado de una evaluación comparativa entre ClausIE y otros métodos de extracción de información similares (ReVerb y OLLIE). Una versión definitiva de los resultados se encuentra en proceso de publicación. Pero estos serían favorables a ClausIE, razón por la cual se escogió dicho método para este trabajo.

2 Trabajos relacionados

En [Blair-Goldensohn et al., 2008] utilizaron un método híbrido para extraer las características de las críticas, consistente en un método dinámico y uno estático de extracción. Buscaron sustantivos o sustantivos compuestos de hasta tres palabras y que aparecieran en ciertas frases que indicaban una carga de sentimientos (polaridad) y/o que respetaran ciertos patrones sintácticos que eran indicadores posibles de una opinión. Encontraron que los patrones eran más precisos que la ocurrencia de los sustantivos en frases con carga de sentimientos. El patrón más productivo que tuvieron buscaba secuencias de sustantivos que tuvieran inmediatamente antes un adjetivo, así encontraron por ejemplo frases como "...*great fish tacos*...", en críticas sobre restaurantes. Incluyeron "*fish tacos*" (tacos de pescado) como una característica, ya que este era un platillo característico de los restaurantes que habían sido evaluados en las críticas.

El segundo enfoque, el método estático para la extracción de características, consistió en lo siguiente: tomaron al azar 1500 oraciones de críticas sobre hoteles y restaurantes y las etiquetaron de forma manual indicando las características "de grano-grueso" que hallaron en ellas. Las llamaron características de grano-grueso

porque eran características generales que podían ser aplicadas a cualquier restaurante u hotel, no eran tan específicas como por ejemplo: “taco de pescado”. Las características fueron las siguientes: comida, ambiente, servicio y precio para restaurantes. Para los hoteles usaron las características: habitaciones, ubicación, comedor, servicio y precio. También incluyeron una categoría *otras*, para etiquetar sentencias que no incluyeran ninguna de las anteriores. Luego etiquetaron un clasificador y lo entrenaron con el conjunto de casos etiquetados, finalmente utilizaron el clasificador ya entrenado para detectar aspectos en cualquier otra oración.

En [Zhuang et al., 2006] se llevó a cabo un experimento similar al propuesto en este artículo, se realizó un resumen automático de críticas cinematográficas de IMDB, focalizado en encontrar opiniones sobre las características de una película dada. Los autores definieron a una característica de película como un elemento (puesta en escena, música, etc.) o bien como personas (director, actor, etc.) mencionadas en una opinión. Los autores definieron de forma manual la lista de características principales (de tipo elemento) que son relevantes en una película y para las características asociadas a personas usaron el elenco completo de participantes tal y como está publicado en IMDB para una película dada.

Las características de tipo elemento escogidas de forma manual fueron las siguientes seis:

- OA: general
- ST: guión
- CH: diseño de personajes
- VP: efectos visuales
- MS: efectos de sonido y música
- SE: efectos especiales

Cada característica fue asociada a múltiples palabras claves, por ejemplo la característica “guion”, fue asociada a las diferentes palabras claves en inglés: *story, plot, script, storyline, dialogue, screenplay, ending, line, scene, tale*. Para obtener las palabras claves, se trabajó con un conjunto de datos conformado por 1100 críticas cinematográficas de IMDB etiquetadas de forma manual. Luego las palabras claves asociadas a una característica las obtuvieron al quedarse solo con las más frecuentes.

3 Problemas encontrados

Los mismos autores de [Blair-Goldensohn et al., 2008] encontraron un problema fundamental con el primer enfoque, el del método dinámico y fue que los aspectos encontrados son de grano fino. No es trivial deducir que “sopa de pescado” y “sopa de langosta” forman parte de un aspecto mayor que podría ser: “sopas”, “entradas” o “comida”.

Respecto al segundo enfoque, el clasificador logró una precisión bastante alta, por ejemplo obtuvo un 86.9 % para la “servicio” y 90.3 % para “precio” en el caso de restaurantes. En el caso de hoteles logró un 83.9% de precisión para “servicio” y 83.3

% para “precio”. La exhaustividad (*recall* en inglés) fue un poco más baja, estuvo entre un 54.5 % y un 69.7 % para los casos mencionados. Sin embargo este método tiene la desventaja de necesitar un conjunto de casos etiquetados de forma manual.

El principal problema asociado al trabajo de [Zhuang et al., 2006] es la necesidad de conocer el conjunto de características relevantes de ante mano para poder generar el etiquetamiento manual.

4 Solución propuesta

Para la elaboración de las pruebas experimentales se utilizó un conjunto de datos de 2000 críticas cinematografías extraídas del sitio IMDB y etiquetadas a mano en dos conjuntos: un grupo de 1000 críticas positivas y otro de 1000 críticas negativas. El conjunto de datos fue creado originalmente por Pang y Lee en [Pang et al., 2002] para entrenar un clasificador de textos con el objetivo de realizar una tarea de análisis de sentimientos. Desde entonces el conjunto de datos ha estado disponible en la web y ha sido utilizado en otras publicaciones.

4.1 Obtención de características

Sobre el conjunto de datos se ejecutó el método de extracción de relaciones semánticas ClausIE. ClausIE devuelve por cada relación semántica una tupla de la forma: (Entidad 1, Relación, Entidad 2) en donde “Entidad” es cualquier elemento sintáctico que haga referencia a algo concreto: una persona, un lugar, una marca, etc. Aunque también puede ser una fecha u otro tipo de entidad más bien abstracta. ClausIE utiliza un algoritmo de detección de nombre de entidades para ello (NER por sus siglas en inglés). Se conjeturó que las características de una película tendrían que poder ser detectadas como entidades. Y en un corpus medianamente grande estas se repetirían con una frecuencia superior a otras entidades posibles. Por lo menos las características llamadas de grano-grueso según [Blair-Goldensohn et al., 2008].

Las extracciones semánticas obtenidas se ordenaron por la cantidad de veces que se repetía una “Entidad” inicial. Luego se filtraron los resultados para mostrar solo los que comienzan con el artículo en inglés “the”, de esta forma se evitó listar pronombres y otras palabras de uso frecuente. La lista obtenida se muestra en la tabla 1.

Tabla 1. Repeticiones de la primera entidad que comienzan con “the”

Entidad 1	Repeticiones
the film	3538
the movie	1637
the story	683
the plot	501
the audience	396
the script	387

the characters	320
the director	258
the two	234
the filmmakers	197
the acting	192
the actors	184
the camera	147
the world	143
the dialogue	140
the cast	128
the man	123
the ending	114
the music	112
the scenes	101
the result	100
the performances	99
the special effects	99

La lista que se muestra en la tabla 1 se corresponde muy bien con una lista de características, o según la nomenclatura de [Zhuang et al., 2006] palabras claves que indican características. Sin embargo esta lista requirió de dos pasos manuales, por lo cual su generación no fue completamente automática. Estos pasos fueron los siguientes:

- Un corte de forma arbitraria en 99 repeticiones, no se tomaron más elementos que los que aprecian hasta 99 veces.
- La eliminación manual de algunas entidades que si bien se repetían no corresponden a características de una película: “the two”, “the camera”, “the world”, “the man” (marcadas en negrita)

Si se compara la lista encontrada con la lista de palabras claves de características que se presenta en [Zhuang et al., 2006], se observa que hay 12 palabras en común de un total de 38. Sin embargo en lista de la tabla 1 hay 8 palabras de uso frecuente que no fueron usadas en el trabajo de [Zhuang et al., 2006]. Por último hay que indicar que con las 12 palabras en común encontradas, se cubren todas las características de grano-grueso definidas en [Zhuang et al., 2006], aunque en algún conjunto solo quede una palabra clave. Esto se refleja en la tabla 2.

Tabla 2. Características de grano grueso y sus palabras clave asociadas en [Zhuang et al., 2006].

Características	Palabras clave
OA	film, movie
ST	story, plot, script, storyline, dialogue, screenplay, ending, line, scene, tale
CH	character, characterization, role

VP	scene , fight-scene, action-scene, action-sequence, set, battle-scene, picture, scenery, setting, visual-effects, color, background, image
MS	music , score, song, sound, soundtrack, theme
SE	special-effects , effect, CGI, SFX

En negrita se muestra las 12 palabras claves en común. Las otras palabras claves encontradas pertenecerían a las características de grano grueso OA y CH, según la siguiente lista:

- **CH:** acting, actors, cast, performances
- **OA:** director, audience, filmmakers, results

4.2 Análisis de sentimiento de cada característica

Para el siguiente análisis se tomó la lista de la tabla 1, sin contar las palabras filtradas, como una lista de características ya que el objetivo de este artículo es la obtención de características de forma automática a partir de un conjunto de críticas en lenguaje natural. Para cada característica se realizó una tarea de análisis de sentimiento utilizando el lexicon de sentimientos SentiWordnet 3.0 [Baccianella et al., 2010].

Se procedió del siguiente modo: se recuperaron todas las extracciones semánticas para una crítica dada, luego se unió cada extracción en una sola oración concatenando “Entidad 01” con “Relación” con “Entidad 02”. Si en la oración resultante aparecía alguna de las características de la lista se evaluaba la misma utilizando el diccionario SentiWordNet 3.0. Luego según el resultado de la polaridad obtenido, positivo o negativo, se marcó dicha característica con un 1 o un -1 en una tabla de resultados final.

Finalmente con aquellas críticas para las cuales se encontraron características, se sumaron los valores de las polaridades de cada una de ellas para obtener un resultado o polaridad global. Este último paso se realizó con el fin de comparar el análisis de las características, el cual en conjunto debería ser idéntico al análisis global. Si esto no hubiese sido así, las características no habrían sido representativas de la película o bien el cálculo de su polaridad tendría que haber sido erróneo.

5 Resultados y Conclusiones

La precisión global para el análisis de sentimientos (más específicamente la obtención de la polaridad), usando SentiWordNet 3.0 sobre las 2000 críticas cinematográficas es 0,662; son 1324 críticas categorizadas correctamente. Este es el piso, sobre el cual se cimentan los análisis de características, un piso bajo, sobre todo al comprar los resultados obtenidos con métodos de clasificación supervisada como los que se usaron en [Pang et al., 2002].

Solo en 1187 críticas se encontró al menos una característica para poder analizar, lo que equivale al 59% de las mismas.

La suma de las polaridades positivas y negativas de cada una de las características, para obtener la polaridad global de la crítica arrojó una precisión de 0.619, es decir 735 clasificadas correctamente de las 1187 que tenían al menos una característica. Si bien es un número bajo, es una precisión cercana a la precisión global de SentiWordNet 3.0. Sobre ese mismo segmento de críticas (las 1187 que tienen al menos una característica) SentiWordNet 3.0 obtuvo por su cuenta una precisión de 0.666, es decir un total de 790 correctamente clasificadas.

Sin embargo la precisión promedio obtenida fue mayor que la calculada en [Zhuang et al., 2006], en donde se calculó la precisión promedio de diferentes pares de características-opiniones para distintas películas siendo el promedio total de la precisión 0.483. Sin embargo, al ser diferente el conjunto de críticas utilizado (el de los autores no está disponible) y diferente la forma de analizar la polaridad, las precisiones no son directamente comparables. Se la cita solo a modo referencia.

Por último, el principal resultado positivo es la extracción de forma casi automática (con una intervención manual mínima) de las características de un producto o servicio (en este caso películas). Las características pueden no ser exhaustivas, al comparlas con las utilizadas en el trabajo de [Zhuang et al., 2006] pero son representativas y sin duda utilizadas con mayor frecuencia en el conjunto de datos analizado. El análisis de sentimiento sobre las características individuales, no mejora el rendimiento global del método utilizado (en este caso el lexicón de sentimientos SentiWordNet) pero se mantiene coherente con la precisión del mismo.

6 Futuras líneas de investigación

Queda como trabajo futuro la revisión y comparación de este enfoque con otros métodos de extracción de características automáticos como SABRE [Caputo et al., 2017].

Referencias

- Baccianella, S., Esuli, A., & Sebastiani, F. (2010, May). SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In LREC (Vol. 10, pp. 2200-2204).
- Banko, M., Cafarella, M. J., Soderland, S., Broadhead, M., & Etzioni, O. (2007, January). Open information extraction for the web. In IJCAI (Vol. 7, pp. 2670-2676).
- Blair-Goldensohn, S., Hannan, K., McDonald, R., Neylon, T., Reis, G. A., & Reynar, J. (2008, April). Building a sentiment summarizer for local service reviews. In WWW workshop on NLP in the information explosion era (Vol. 14, pp. 339-348).
- Caputo, A., Basile, P., de Gemmis, M., Lops, P., Semeraro, G., & Rossiello, G. (2017). SABRE: A Sentiment Aspect-Based Retrieval Engine. In Information Filtering and Retrieval (pp. 63-78). Springer International Publishing.
- Del Corro, L., & Gemulla, R. (2013, May). ClausIE: clause-based open information extraction. In Proceedings of the 22nd international conference on World Wide Web (pp. 355-366). International World Wide Web Conferences Steering Committee.

- García-Martínez, R. & Britos, P. V. (2004). Ingeniería de sistemas expertos. Nueva Librería. ISBN 987-1104-15
- Gómez, A., Juristo, N., Montes, C., & Pazos, J. (1997). Ingeniería del conocimiento. Editorial Centro de Estudios Ramón Areces. ISBN 84-8004-269-9.
- Pang, B., Lee, L., & Vaithyanathan, S. (2002, July). Thumbs up?: sentiment classification using machine learning techniques. In Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10 (pp. 79-86). Association for Computational Linguistics.
- Rodríguez, J. M., Merlino, H., García-Martínez, R. (2015). Revisión Sistemática Comparativa de Evolución de Métodos de Extracción de Conocimiento para la Web. XXI Congreso Argentino de Ciencias de la Computación (CACIC 2015). Buenos Aires, Argentina.
- Rodríguez, J. M., Merlino, H. D., Pesado, P., & García-Martínez, R. (2016, August). Performance Evaluation of Knowledge Extraction Methods. In International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems (pp. 16-22). Springer International Publishing.
- Zhuang, L., Jing, F., & Zhu, X. Y. (2006, November). Movie review mining and summarization. In Proceedings of the 15th ACM international conference on Information and knowledge management (pp. 43-50). ACM.

Cálculo del exponente de Hurst utilizando Spark Streaming: enfoque experimental sobre un flujo de transacciones de criptomonedas

María José Basgall^{1,3}, Waldo Hasperué^{1,2}, Marcelo Naiouf^{1,2}, Aurelio F. Bariviera⁴

¹Instituto de Investigación en Informática (III-LIDI), Facultad de Informática - Universidad Nacional de La Plata

²Comisión de Investigaciones Científicas (CIC)

³UNLP, CONICET, III-LIDI, La Plata, Argentina

⁴Department of Business, Universitat Rovira i Virgili, Av. Universitat 1, 43204 Reus, Spain

{mjbasgall, whasperue, mnaiouf}@lidi.info.unlp.edu.ar
aurelio.fernandez@urv.cat

Resumen. Actualmente es cada vez más común encontrarse con problemas de Big Data, donde las aplicaciones desarrolladas para resolver dichos problemas son implementadas en frameworks específicos. Uno de los que más se utiliza es Apache Spark, que posee el módulo Spark Streaming el cual permite el tratamiento de datos provenientes de un flujo de información potencialmente infinito. En este trabajo se presenta una aplicación implementada en Spark Streaming que realiza el cálculo del exponente de Hurst, un indicador muy utilizado en el análisis de mercado para la detección de memoria a largo plazo. Los ensayos realizados se hicieron sobre flujos simulados de transacciones de criptomonedas que demuestran la capacidad de Spark Streaming para el tratamiento de este tipo de flujos.

Palabras clave: Spark Streaming, Stream processing, Criptomonedas, Exponente de Hurst.

1 Introducción

En la actualidad, se ha incrementado la capacidad de recolectar y almacenar información por lo que trabajar con grandes cantidades de datos es cada vez más usual. Debido a que las técnicas tradicionales de análisis de datos que son ejecutadas en una computadora no son capaces de lidiar con estos volúmenes de datos por cuestiones de escalabilidad, surge el término Big Data, que se refiere a la aparición de una enorme cantidad de datos heterogéneos y las técnicas de procesamiento y análisis que se requieren para analizarlos en tiempos razonables de acuerdo al problema [1][2][3].

Actualmente existen diferentes herramientas que simplifican la tarea de extraer conocimiento frente a estos escenarios. Una de las herramientas más utilizadas es Hadoop MapReduce [4][5], construida sobre los principios del procesamiento

paralelo y distribuido. MapReduce trabaja sobre el sistema de archivos HDFS [6], el cual es distribuido, ofrece eficiencia, almacenamiento tolerante a fallos y es adecuado para aplicaciones que utilizan grandes volúmenes de datos ya que proporciona un alto rendimiento de acceso a los mismos. El framework MapReduce realiza las tareas de distribución y paralelización de manera automática y transparente al programador, convirtiéndola en una poderosa herramienta para implementar soluciones distribuidas.

A pesar de su popularidad, se han encontrado múltiples limitaciones para desarrollar programas escalables con MapReduce, ya que es ineficiente para aplicaciones que comparten datos a través de varias etapas de un algoritmo, como en tareas iterativas, debido al uso intensivo de disco generando un importante decremento de performance. Además MapReduce tiene su punto fuerte en el procesamiento off-line, lo cual lo convierte en una herramienta no muy útil en problemas donde se necesita una respuesta dinámica y en tiempo real.

Por lo anterior, han surgido diferentes plataformas que sobrellevan los problemas que presenta MapReduce, una de ellas es Apache Spark [7][8], la cual está siendo cada vez más usada debido a que es una de las más flexibles y poderosas para realizar cómputo distribuido de manera rápida en escenarios Big Data mediante el uso intensivo de memoria RAM, permitiendo leer los datos en ella y consultarlos repetidamente, siendo una característica deseable para algoritmos que usan los datos de manera iterativa.

Spark tiene dos modos de trabajo: batch y online. Este último permite analizar los datos provenientes de un flujo de datos con la posibilidad de ofrecer una respuesta online. Estos flujos de datos pueden ser de diversas fuentes: publicaciones de usuarios en redes sociales, logs generados por servidores, datos provenientes de sensores, navegación web o transacciones monetarias como lo pueden ser el reciente uso del comercio utilizando criptomonedas.

En relación a las criptomonedas, en los últimos años se ha comenzado a utilizar un nuevo tipo de moneda digital que se podría caracterizar como "sintéticas" en el sentido de que no surgió por decisión de un Estado sino que emergió de un acuerdo privado y facilitado por el anonimato que se puede obtener de internet. Las monedas digitales o criptomonedas permiten pagos instantáneos a cualquiera y en cualquier lugar del mundo.

La más importante de ellas es la llamada Bitcoin (BTC) [9] cuyo software fue creado por Satoshi Nakamoto bajo licencia MIT [11]. Y la idea consiste en usar criptografía para controlar la creación y transferencia de dinero de manera descentralizada y sin tener que confiar en las autoridades centrales.

Existen alrededor de 750 criptomonedas, además de Bitcoin. Las que siguen en popularidad son Ethereum (ETH), Ripple (XRP), Litecoin (LTC). Sin embargo, se puede observar en la figura 1 donde se muestra el porcentaje de capitalización bursátil que BTC representa el 47% del mercado, seguido por ETH con un 23%.

En ciencias económicas un indicador que resulta muy útil es el de determinar la presencia de memoria o correlación a largo plazo en un mercado para saber si se pueden construir estrategias de negociación que permitan extraer un rendimiento superior al mercado. Una forma de realizar esa tarea es mediante el cálculo del exponente de Hurst [12]. El poder asegurar que existe una correlación a largo plazo es

un desafío al modelo financiero establecido, ya que de acuerdo a la teoría económica estándar, en un mercado competitivo los precios se debería mover aleatoriamente, reflejando una serie temporal sin memoria. Lograr este objetivo es de mucho interés para las personas que invierten en estos tipos de mercado. En un trabajo reciente [13], se realizó un estudio del exponente de Hurst durante el período 2011-2017, detectando una variabilidad en dicho exponente a lo largo del tiempo. Por otro lado, se comprobó que dicho exponente tiene un comportamiento similar cuando es medido en distintas escalas temporales (rendimientos de 5 a 12 horas). Por ello, un avance natural en esta línea de investigación consiste en estudiar el comportamiento del exponente de Hurst en un entorno online.

En este artículo se presenta el cálculo del exponente de Hurst mediante el método DFA (Detrended Fluctuation Analysis) para un flujo de datos simulado compuesto por transacciones de criptomonedas utilizando el framework Spark Streaming. Este trabajo es a modo experimental ya que si bien es posible encontrar en internet bases de datos de transacciones de criptomonedas, aún no es posible obtener de manera online los datos de las transacciones que se necesitan para el cálculo del exponente de Hurst. Sería esperable, dado el avance de las tecnologías de la información y la comunicación, que en un futuro este tipo de flujo esté disponible en acceso abierto para diferentes análisis.

El resto del artículo se divide de la siguiente manera: en la sección 2 se describe el framework Spark streaming. En la sección 3 se describe el algoritmo del cálculo del exponente de Hurst y su implementación en Spark Streaming. En la sección 4 se describen los ensayos realizados sobre flujos simulados de transacciones usando criptomonedas y finalmente en la sección 5 se mencionan las conclusiones y los trabajos a futuro en los que se está trabajando.

2 Spark Streaming

El procesamiento de flujos de datos (*stream processing*) es un área muy estudiada en los últimos años. Stream processing permite llevar a cabo tareas sobre un flujo continuo y potencialmente infinito de datos [14][15][16][17][18].

El objetivo de este tipo de procesamiento es permitir que las tareas analicen los datos de un flujo de forma online, brindando respuestas en tiempos muy cercanos al tiempo real. La principal característica de esta forma de trabajar sobre un flujo de datos, es que los datos del flujo llegan a una velocidad tal que no es posible almacenarlos en su totalidad y, si se pueden almacenar, el volumen de datos es tan grande que presenta la dificultad de analizarlo en tiempos de respuesta cortos [19]. La velocidad es un concepto a tener en cuenta en los escenarios de Big Data, más aún en los escenarios de stream processing.

Existen dos tipos de modelos de procesamiento en stream processing, un dato a la vez y micro-batching. En el primer enfoque, los datos se procesan a medida que llegan y luego se descartan. Por otro lado, con el modelo de micro-batching el cómputo se lleva a cabo sobre pequeños lotes de datos capturados cada un

determinado intervalo de tiempo. Spark Streaming [20] es un framework que implementa el modelo de procesamiento de datos por micro-batching.

Spark Streaming permite el procesamiento de flujos de datos de manera escalable, de alto rendimiento y tolerante a fallos. La ingesta de los datos puede llevarse a cabo mediante diferentes fuentes de datos como puede ser Kafka, Flume, Twitter, ZeroMQ, Kinesis, o sockets TCP. Estos datos pueden ser procesados mediante algoritmos que utilicen las funciones de alto nivel provistas por la librería como son map, reduce, join, etc, para luego ser escritos en archivos, base de datos o mostrados por pantalla.

En Spark se presenta un modelo de abstracción llamado Resilient Distributed Dataset (RDD), el cual es una colección distribuida de datos tolerante a fallas que permite persistir resultados intermedios en memoria. Los RDD también pueden ser almacenados en memoria caché como en disco, según sea necesario. Spark streaming trabaja con los denominados discretized stream (DStream), abstracciones de alto nivel que provee Spark Streaming. Éstos representan un flujo continuo de datos e, internamente, están representados por una secuencia de RDD (el cual es una colección de elementos tolerantes a fallas que puede ser operada en paralelo mediante transformaciones y acciones).

La gran ventaja de utilizar Spark Streaming es que el framework puede ser ejecutado en un cluster, donde cada nodo del cluster se encarga de ejecutar las transformaciones y acciones sobre diferentes RDD, dando lugar al procesamiento de datos en paralelo y distribuido.

Además es posible configurar el tiempo de la ventana temporal al momento de recolectar los datos que serán utilizados en cada micro-batch.

3 Cálculo del exponente de Hurst en Spark Streaming

Las transacciones de criptomonedas se almacenan como tripletas $\langle TS, V, P \rangle$ guardando el timestamp (TS), el volumen de comercio de criptomonedas (V) y el precio (P) [21]. Para el cálculo del exponente de Hurst solo se utiliza el timestamp (segundos transcurridos desde las 0 horas del 1 de enero de 1970 GMT).

El flujo de datos compuesto por tripletas $\langle TS, V, P \rangle$ se divide en ventanas temporales de la misma longitud (LV) y no solapadas, llamadas batches o lotes. Cada batch B_i comienza en el tiempo ti_i y finaliza en el tiempo tf_i , siendo $(tf_i - ti_i) = LV$, además $ti_i = 1 + tf_{(i-1)}$. Dentro de cada batch es posible que se obtenga del flujo una o más transacciones. De cada batch se extrae la transacción T_i donde su respectivo timestamp (TS_i) es el más cercano a tf_i .

Se calcula la distancia entre la hora de arribo de la transacción y el tiempo del fin del batch como $D_i = tf_i - TS_i$.

Con los D_i se calcula el exponente de Hurst usando el método llamado Detrended Fluctuation Analysis (DFA) que es más apropiado cuando se trabaja con datos no estacionarios [12].

Se calculan los $x(i)$ que representan una serie autosimilar de la serie original de precios como lo muestra la ecuación 1.

$$x(i) = \sum_{t=1}^i [D_t - \overline{D}] \quad (1)$$

donde \overline{D} es la media aritmética de los D_i , $i=1..M$.

La serie de los $x(i)$ se divide en M/m submuestras y de cada submuestra se realiza un ajuste polinómico, que por lo general es un ajuste lineal, para obtener los $x_{fit}(i, m)$. m son las longitudes de intervalos dentro de la submuestra, para permitir el comportamiento fractal de la serie temporal.

La ecuación 2 muestra la forma de calcular los coeficientes $F(m)$ utilizados para el cálculo del exponente de Hurst.

$$F(m) = \sqrt{\frac{1}{M} \sum_{i=1}^M [x(i) - x_{fit}(i, m)]^2} \quad (2)$$

Con los pares $(\ln(m), \ln(F(m)))$ se realiza una regresión lineal y la pendiente de dicha regresión es el exponente de Hurst.

3.1 Simulador de flujos

Debido a las dificultades de tomar datos de transacciones de criptomonedas en el mercado real, para el desarrollo de los diferentes experimentos se utilizó un flujo simulado. Si bien es difícil tomar los datos online, éstos se pueden descargar los datos desde los sitios oficiales de las empresas manejadoras de criptomonedas. Por lo tanto la simulación del flujo consiste en leer las transacciones desde un archivo y enviarlas a un socket TCP, ya que Spark Streaming es capaz de procesar datos escuchando estos sockets.

Para esto se implementó un script bash que lee el archivo de las transacciones línea por línea y las envía a la utilidad Netcat [22] presente en la mayoría de los sistemas GNU/Linux para que reciba cada línea del archivo utilizando un puerto específico y, a su vez, en la aplicación Spark Streaming se establece en el contexto que escuche ese mismo puerto para poder recibir los datos que está escuchando el comando Netcat.

Este simulador también permite el envío continuo de transacciones ficticias, creadas aleatoriamente.

3.2 Implementación en Spark streaming

Primero se configura el tamaño de ventana de cada micro-batch al tamaño de ventana deseado para el cálculo del exponente de Hurst. La longitud de la ventana no debe ser inferior a 500 observaciones, a fin de evitar sesgos en la estimación del exponente.

Con esto nos garantizamos que Spark recolecte todas las transacciones del flujo que caen dentro de cada ventana asegurando que todas las ventanas sean del mismo tamaño, condición necesaria para el cálculo del exponente de Hurst.

La implementación realizada en este trabajo está hecha en Java 7, debido a la complejidad del código, se comenta e ilustra con pseudocódigo las diferentes transformaciones y acciones realizadas para el cálculo del exponente de Hurst.

Mediante funciones de mapeo y reducción se obtiene la transacción T_i que representa la transacción más cercana al tiempo de finalización del batch, utilizando para ello una clave única, ya que la siguiente acción a tomar es la reducción de todas las tuplas recolectadas en la ventana:

```
mapeo = batch.mapToPair(tupla -> (1,tupla))
ultima_transacción = mapeo.reduceByKey
                      (a, b -> (a.timestamp > b.timestamp)?a:b)
```

Luego se genera, con la última transacción obtenida, una tupla usando el timestamp de la transacción como clave y la distancia D_i como valor. Esta tupla se persiste ya que es necesaria tener almacenado el “historial” de las últimas transacciones de cada batch del flujo procesado. Cabe aclarar que las siguientes operaciones se realizan por cada una de las ventanas de datos analizadas, donde cada ventana agrega una nueva transacción al “historial” de transacciones.

Utilizando la clase StatCounter provista por Apache Spark se calcula la media de las distancias D_i y se obtiene la menor de ellas, esta última es usada para calcular el número de lote (valor i) al cual pertenece la transacción obtenida en la última ventana:

```
map_stats = historial.updateStateByKey
            (values, current -> (new StatCounter(values, current))
stats = map_stats.mapToPair(sc -> (1, sc.mean))
```

Para el cálculo de los $x(i)$ primero se mapea la diferencia $(D(i)-\bar{D})$ usando el valor de i como clave:

```
dif_D = historial.mapToPair(t -> (i, t - sc.mean))
```

Luego se hace el producto cartesiano de dif_D para crear una relación que permita calcular la sumatoria de las diferencias $(D(i)-\bar{D})$ y así obtener la serie $x(i)$.

```
prod_carte = dif_D.transformWith(dif_D)
              (rdd_a, rdd_b -> rdd_a.cartesian(rdd_b))
```

Cada tupla de prod_carte $(i_a, D_{i_a}, i_b, D_{i_b})$ se mapea a un par usando i_a como clave y D_{i_b} como valor si $i_b \leq i_a$ o 0 en caso contrario. La reducción es usada para sumar los valores y así obtener la serie $x(i)$.

```
x(i) = prod_carte.reduceByKey(a, b -> a+b)
```


Para el cálculo de la regresión la serie $x(i)$ se mapea usando un par (m,l) como clave y $x(i)$ como valor, donde l es el número de submuestra al que pertenece cada valor $x(i)$. La idea es poder aprovechar al máximo la capacidad de Spark Streaming distribuyendo, por cada valor de m y submuestra l , el correspondiente cálculo de la regresión polinómica (lineal en este trabajo) y el cálculo de la serie $x_{fit}(i,m)$.

```
map_regres = x(i).flatMap(a -> [(m, l), a.x_value])
```

Como en Spark la reducción se hace por pares de valores no pudiendo acceder a todos los mismos en una misma operación, nos vimos obligados a hacer un mapeo extra para juntar todos los $x(i)$ en un único string y así poder acceder a todos los $x(i)$ pertenecientes a un mismo submuestreo para realizar la regresión lineal.

```
map_aux = map_regres.mapToPair
              (a -> (a.m, a.l), a.x.toString())
```

```
submuestras = map_aux.reduceByKey(a, b -> a+b)
```

Luego se usa la operación flatMap para generar los $x_{fit}(i,m)$. También se aprovecha la operación para devolver una tupla con el valor i como clave y como valor la diferencia $(x(i) - x_{fit}(i, m))^2$.

```
xfit_i_m = submuestras.flatMap(a -> (a.m, xfit))
```

Luego se hace un mapeo por valor de m y la reducción correspondiente para el cálculo de la sumatoria.

```
map_sum = xfit_i_m.mapToPair(a -> (a.m, xfit))
```

```
suma = map_sum.reduceByKey(a, b -> a+b)
```

Así obtenemos una tupla por cada valor de m utilizado. El anteúltimo paso consiste en el mapeo a un par usando una clave única y los valores de m y los $F(m)$ como valor. Como también necesitamos todos estos valores al mismo tiempo para el cálculo de la regresión lineal se convierten a string para concatenarlos y poder usarlos en la etapa posterior.

```
map_F_m = suma.mapToPair
              (a -> (1, m.toString(), F(m).toString()))
```

```
pares_m_F_m = map_F_m.reduceByKey(a, b -> a+b)
```

El último paso es el que recibe la única tupla producto de la operación anterior y con cada par $(m, F(m))$ recibido como valor hace la regresión lineal para el cálculo del exponente de Hurst.

```
hurst = pares_m_F_m.map(a, hurst_value)
```

4 Resultados

Las pruebas realizadas se llevaron a cabo utilizando un cluster de máquinas virtuales, cada una de 2 cores y 7 GB de RAM. Los nodos tienen sistema operativo Ubuntu 16.04, Spark 1.6.3 y Hadoop 2.6.5. El script generador de flujo implementado y utilizado en los ensayos genera alrededor de 1400 transacciones por segundo.

Uno de los interrogantes que se planteó durante el desarrollo de este trabajo es cómo varía la performance de Spark cuando se varía el tamaño de la ventana. Como con cada batch recolectado se guarda una transacción, al cabo de n batches se tendrán n transacciones con las cuales operar para el cálculo del exponente de Hurst. Esto provoca que el cálculo del exponente tarde cada vez más tiempo, ya que con cada ventana analizada se obtiene una nueva transacción, la cual obliga a realizar todos los cálculos.

Como primer experimento se simuló un flujo de transacciones utilizando dos tamaños de ventana diferentes para medir la performance del procesamiento de Spark. Se ejecutó el proceso en dos clusters configurados con uno y dos nodos. Los tiempos de ejecución de cada batch se pueden observar en la figura 1. Se puede observar que el tiempo de procesamiento se mantiene dentro del tiempo de la ventana temporal y que si bien al comienzo hay un overhead, producto del lanzamiento del cluster y la distribución inicial de tareas, luego se estabiliza, demostrando la capacidad de Spark de tratar flujos de alta frecuencia de datos.

5 Conclusiones y trabajo a futuro

Se presentó un algoritmo implementado en el framework Spark Streaming que obtiene de manera online, el cálculo del exponente de Hurst en un flujo simulado de transacciones de criptomonedas.

Los resultados obtenidos muestran la eficiencia y la capacidad de Spark para tratar con flujos cuya frecuencia es de 1400 transacciones por segundo pudiendo dar respuestas online usando ventanas de tamaño de 5 y 10 segundos. En el mercado real estos flujos tienen una frecuencia mucho menor que la simulada en este trabajo, por lo que el algoritmo presentado en este trabajo al ser ejecutado en Spark Streaming podría tratar de manera online los flujos del mercado real sin ningún problema.

Una de las condiciones que debe cumplir el flujo de transacciones para el cálculo del exponente es que cada batch debe por lo menos tener una transacción. En este trabajo se simuló el flujo de manera de asegurarse que cada batch analizado tenga al menos una transacción. En flujos reales donde la distancia en el tiempo de llegada entre diferentes transacciones no se conoce a priori, es necesario calcular, la distancia máxima entre transacciones.

Como trabajo a futuro se está implementando una aplicación que es capaz de almacenar parte del flujo para el cálculo dinámico del tamaño de ventana LV y así

poder calcular de manera correcta el exponente de Hurst. Esto trae aparejado el desafío de poder almacenar la mayor parte del flujo en memoria, para lo cual el procesamiento paralelo y distribuido que proporciona Spark Streaming es una gran ayuda. Problema que los algoritmos secuenciales no podrían lidiar en tiempo real para flujos de alta frecuencia de llegada de datos.

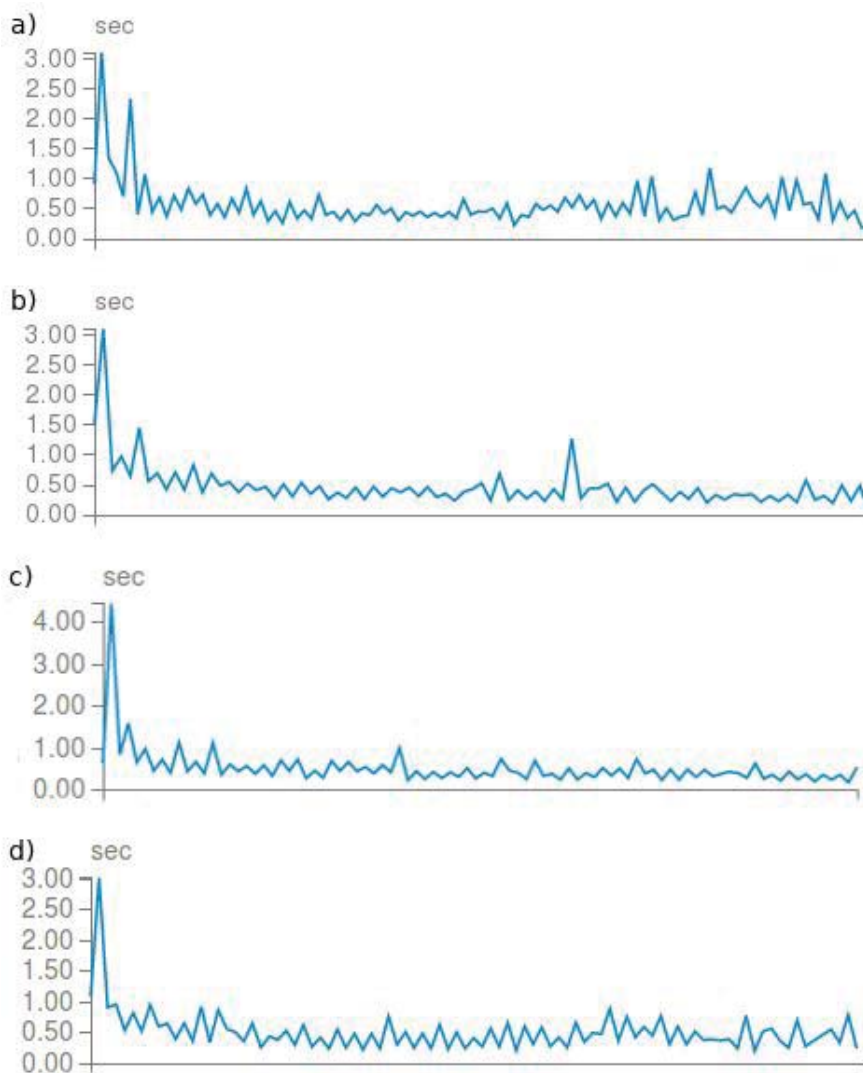


Fig. 1. Tiempo de ejecución de cada batch para:

- a)** 1 master y 1 slave con ventana de 5 segs.
- b)** 1 master y 1 slave con ventana de 10 segs.
- c)** 1 master y 2 slaves con ventana de 5 segs.
- d)** 1 master y 2 slaves con ventana de 10 segs.

Referencias

1. D. Laney, "3D data management: Controlling data volume, velocity, and variety," tech. rep., META Group, February 2001.
2. S. Kaisler, F. Armour, J. A. Espinosa, and W. Money, "Big data: Issues and challenges moving forward," 2013 46th Hawaii International Conference on System Sciences (HICSS 2013), vol. 00, no. undefined, pp. 995–1004, 2013.
3. B. Saha and D. Srivastava, "Data quality: The other face of big data" 2014 IEEE 30th International Conference on Data Engineering (ICDE), undefined, pp. 1294–1297, 2014.
4. J. Dean and S. Ghemawat, "MapReduce: Simplified Data Processing on Large Clusters," *Commun. ACM*, vol. 51, pp. 107–113, Jan. 2008.
5. Hadoop, White, T.: Hadoop: The Definitive Guide, 4th edn. O'Reilly Media Inc, 2015
6. HDFS.
<http://hadoop.apache.org/docs/current/hadoop-project-dist/hadoop-hdfs/HdfsDesign.html>.
Accedido en 07/2017.
7. Apache Spark. <http://spark.apache.org/> . Accedido en 07/2017.
8. Zaharia, M., Chowdhury, M., Das, T., Dave, A., Ma, J., McCauley, M., Franklin, M.J., Shenker, S., Stoica, I.: Resilient distributed datasets: a fault-tolerant abstraction for in-memory cluster computing. In: Proceedings of the 9th USENIX Conference on Networked Systems Design and Implementation, pp. 1–14. USENIX Association (2012)
9. Bitcoin https://en.bitcoin.it/wiki/Main_Page Accedido en 07/2017.
10. <https://coinmarketcap.com/currencies/bitcoin/> Accedido en 07/2017.
11. S. Nakamoto, Bitcoin: A peer-to-peer electronic cash system, 2009.
12. A. Carbone, G. Castelli, H.E. Stanley, Time-dependent Hurst exponent in financial time series, *Physica A: Statistical Mechanics and its Applications*, Vol 344, Issue 1, 2004, pp. 267-271.
13. A.F. Bariviera, M.J. Basgall, W. Hasperué, M. Naiouf, Some stylized facts of the Bitcoin market, *Physica A: Statistical Mechanics and its Applications*, Vol 484, 2017, pp. 82-90,
14. N. Takahashi et al., "A parallelized data stream processing system using dynamic time warping distance," in 2009 International Conference on Complex, Intelligent and Software Intensive Systems, Fukuoka, Japan, March 16-19, 2009, pp. 1100–1105.
15. Y. Noh et al., "Real-time data stream processing for ubiquitous home network systems," in 4th International Conference on Multimedia and Ubiquitous Engineering, MUE 2010, Cebu, Philippines, 11-13 August, 2010.
16. C. Kuka, "Processing the uncertainty: Quality-aware data stream processing for dynamic context models," in Pervasive Computing and Communications Workshops (PERCOM Workshops), 2012 IEEE International Conference on, pp. 560–561, March 2012.
17. D. Bonino and F. Corno, "spchains: A declarative framework for data stream processing in pervasive applications," *Procedia Computer Science*, vol. 10, 2012.
18. J. Stefanowski et al., "Processing and mining complex data streams," *Inf. Sci.*, vol. 285, pp. 63–65, 2014.
19. A. Rajaraman and J. D. Ullman, *Mining of Massive Datasets*. New York, NY, USA: Cambridge University Press, 2011.
20. <http://spark.apache.org/docs/latest/streaming-programming-guide.html> Accedido en 07/2017.
21. <http://api.bitcoincharts.com/v1/csv/> Accedido en 07/2017.
22. <http://netcat.sourceforge.net/> Accedido en 07/2017.

El Problema de la Dispersión Máxima en un entorno Multi-Objetivo

Mauricio Machuca Cabral, Benjamín Barán, and Fernando Sandoya

Facultad Politécnica
Universidad Nacional de Asunción,
Asunción, Paraguay
School of Natural Sciences and Mathematics
Escuela Superior Politécnica del Litoral
Guayaquil, Ecuador

mmachuca78@gmail.com, bbaran@pol.una.py, fsandoya@espol.edu.ec
<http://www.pol.una.py>, <http://www.espol.edu.ec>

Resumen El problema de la diversidad máxima o dispersión máxima (*MDP - Maximum Diversity Problem*) presenta un gran número de aplicaciones prácticas que surgen de la búsqueda de los elementos más disímiles de un conjunto de datos. Este tipo de problema utiliza modelos de diversidad y definiciones de distancia o disimilitud (observar que en el contexto de este trabajo distancia y disimilitud son considerados sinónimos) como forma de medir que tan diferentes son los elementos de un conjunto dado. Como diversos expertos pueden preferir diferentes definiciones de distancia para problemas específicos, se propone resolver el *MDP* con un enfoque multi-objetivo, considerando por primera vez, la utilización simultánea de múltiples definiciones de distancia. En este nuevo contexto multi-objetivo, este trabajo propone la utilización de un algoritmo evolutivo multi-objetivo (el reconocido NSGA-II), presentando varios casos de prueba que demuestran la eficiencia del algoritmo propuesto en comparación con la búsqueda exhaustiva.

Keywords: Maximum Dispersion Problem, Maximum Diversity Problem, MultiObjective Optimization, MultiObjective Evolutionary Algorithm.

1. Introducción

El problema de la diversidad máxima o dispersión máxima (observar que en el contexto de este trabajo diversidad y dispersión son sinónimos) consiste en seleccionar un subconjunto $M = \{i\}, i = 1, \dots, |M|$ de un conjunto $N = \{j\}, j = 1, \dots, |N|$ ($|M| < |N|$), de forma a maximizar la diversidad entre los elementos del subconjunto M seleccionado. En los últimos años se han propuesto diferentes modelos para la resolución de este tipo de problema, como por ejemplo:

- *Maximum Diversity Problem (MDP)* [6];
- *Max-Min Diversity Problem (MMDP)* [18];

- *Maximum Mean Dispersion Problem (Max-Mean DP)* [17];
- *Minimum Differential Dispersion Problem (Min-Diff DP)* [17].

Cada uno de estos modelos utiliza una definición diferente para interpretar el concepto de diversidad/dispersión, pero todas estas definiciones tienen la característica común de necesitar para su implementación de una *medida de diversidad* ($div(M)$) a ser aplicada sobre el conjunto de elementos estudiados. En la literatura podemos encontrar varios métodos que se han planteado para la resolución del *MDP* (mono-objetivo), como *GRASP and Path Relinking* [13], *Iterated Local Search* [4], *Variable Neighborhood Search (VNS)* [1], *Iterated Tabu Search (ITS)* [16], *Memetic self-adaptive evolution strategies* [5], *a tabu search based memetic algorithm* [23], *Opposition-based Memetic Search* [24], *Maximum-score Diversity Selection* [15], entre otros.

Las áreas de aplicación son las más diversas, y entre estas se pueden citar: selección eficiente de equipos de trabajo [20], preservación de la diversidad biológica [7], descubrimiento temprano de medicamentos [15], diseño de la estructura molecular, existencias de cría agrícola, composición de jurados [8], ubicación de instalaciones, diseño de productos, genética, sistemas ecológicos, tratamientos médicos y promoción de la diversidad étnica entre inmigrantes [12] por citar solo las aplicaciones más conocidas.

La medida de diversidad o medida de dispersión está relacionada a una definición de distancia, por lo tanto, para resolver el *MDP* necesitamos generar la matriz de distancias inter-elemento, para luego maximizar la diversidad (la definición de distancia utilizada depende en general del campo de aplicación del problema sobre el que se esté trabajando y del experto que lo escoje, es decir, diferentes expertos, utilizan diferentes definiciones de distancia). En este contexto, este trabajo propone resolver el *MDP* multi-objetivo, utilizando varias matrices de distancia (obtenidas cada una de ellas con una definición diferente de distancia) sobre la misma base de datos, convirtiéndose así cada una, en la matriz de entrada que permite calcular cada función objetivo del problema. De entre las meta-heurísticas disponibles para la resolución de problemas multi-objetivo, en este trabajo se utiliza el *Nondominated Sorting Genetic Algorithm II (NSGA-II)* por ser el algoritmo evolutivo más utilizado para resolver problemas multi-objetivo [11]. Este algoritmo es implementado y comparado con un algoritmo de búsqueda exhaustiva (BE) para resolver el *Maximum Diversity Problem (MDP)* en un contexto puramente multi-objetivo.

El presente trabajo está organizado de la siguiente manera: en la Sección 2 se discuten los trabajos relacionados al *Maximum Dispersion Problem*, en la Sección 3 se presentan definiciones de distancia, diversidad y el problema de la diversidad máxima, la Sección 4 presenta el planteamiento y la formulación matemática del problema; la Sección 5 presenta la propuesta de aplicación del *NSGA-II* para resolver el *MDP* multi-objetivo; en la Sección 6 se discuten los resultados experimentales obtenidos, y finalmente, la Sección 7 presenta las conclusiones y líneas de trabajo futuro.

2. Trabajos Relacionados

Diferentes heurísticas y metaheurísticas se han presentado en el pasado para el problema de dispersión máxima.

En el trabajo de Martí et al. [13] se resuelve el *Max-Mean Dispersion Problem*, aplicando un *GRASP (Greedy Randomized Adaptive Search Procedure)* con un *Path Relinking* en el cual la búsqueda local fue basada en la metodología *Variable Neighborhood*. Los valores de distancia que consideraron podían tomar tanto valores positivos como negativos, y no necesariamente satisfacer las propiedades usuales de una distancia, como la desigualdad triangular. Los resultados fueron comparados con trabajos previos también basados en *GRASP* demostrando las ventajas del método propuesto.

Sandoya et al. [21] describen modelos matemáticos para los modelos de diversidad y equidad; así como un compendio de heurísticas y meta-heurísticas para los mismos. La mayoría de estas heurísticas y meta-heurísticas, están basadas en métodos de construcción y búsqueda local, como lo son *GRASP* y *Tabu Search*. Dentro de las conclusiones finales se explica que existen implementaciones muy específicas para estos modelos, quedando todavía pendiente, la implementación de soluciones más genéricas.

Zhou et al. [24] presentan un *Opposition-based memetic algorithm (OBMA)* para la resolución del *MDP*. En él, se integraron el concepto de *opposition-based learning (OBL)* con el *memetic search framework*, explorando las soluciones candidatas y sus soluciones opuestas durante los procesos de inicialización y evolución. La diferenciación del *OBMA* con otros algoritmos meméticos radica en las siguientes tres características: un procedimiento de búsqueda de trayectoria doble que simultáneamente busca una solución candidata y su opuesta, la aplicación de *tabu search* para una optimización local efectiva y para mantener la diversidad de la población propusieron un *rank-based quality-and-distance pool updating strategy* el cual utiliza una fórmula de distancia para determinar si la nueva solución generada, debe o no ser agregada a la población.

3. Distancias, Diversidad y el Problema de la Diversidad Máxima

El problema de la Diversidad Máxima consiste en seleccionar un subconjunto de elementos de forma que la medida de diversidad del conjunto sea maximizada [10]. Este problema es ya conocido como un problema *NP – Hard* [6].

En Smyth y McClave [22] se define a la diversidad de un conjunto M como la disimilitud promedio entre todos los pares de elementos. A los efectos de medir dicha diversidad $div(M)$, se debe establecer una relación d_{ij} que represente la distancia o disimilitud entre cada pareja de elementos del conjunto M . La definición específica de distancia a ser utilizada depende del problema con el que se trabaje y de las consideraciones de cada experto.

Siendo d_{ij} la distancia entre los elementos i y j , y teniendo cada elemento K atributos, se definen las variables x_{ik} y x_{jk} como el k -ésimo atributo de los

elementos i y j respectivamente. Como ejemplo de medida de disimilitud, se presenta la definición de la distancia Euclidiana utilizando la notación arriba presentada.

$$d_{ij} = \sqrt{\sum_{k=1}^K (x_{ik} - x_{jk})^2} \quad (1)$$

Establecida la distancia d_{ij} entre elementos de un conjunto M , se puede pasar a definir la medida de diversidad en dicho conjunto. La medida de diversidad utilizada en este trabajo, es la dispersión de la suma, que se define como la suma de las distancias inter-elementos de todos los elementos del conjunto M . Esta medida es utilizada en el problema de optimización *Max-Sum* [21]:

$$div(M) = \sum_{i < j, i, j \in M} d_{ij} \quad (2)$$

El objetivo del *MDP* es obtener el subconjunto M más diverso, a través de la selección de los elementos más alejados unos de otros. Sin embargo, esta medida de diversidad tiene la característica de que elementos cercanos pueden quedar en el subconjunto óptimo, si estos se encuentran muy distantes de los demás elementos. En consecuencia, varios autores prefieren evaluar también otros modelos de diversidad como los citados en la Sección 1.

4. Planteamiento del Problema

4.1. Métricas de distancias consideradas

Cada área que trabaja con un problema específico considerando la diversidad máxima, puede utilizar una definición de distancia diferente. En consecuencia, surge la necesidad de resolver este problema como un problema de optimización multi-objetivo, dado que diferentes investigadores pueden preferir definiciones distintas de distancia (o disimilitud). En <http://www.ug.edu.ec/mdp/bases/>, se presentan 65 definiciones de distancia encontradas en la literatura [2], [14] y [9]. Inicialmente se ha calculado la correlación (ρ) entre todos los pares de definiciones relevadas de distancia, y luego, las que presentaron una correlación unitaria ($\rho = 1$) fueron re-clasificadas para este trabajo, quedando 50 definiciones diferentes de disimilitud con correlación de a par diferentes a la unidad, i.e. $|\rho| < 1$. En consecuencia, la novedad en este trabajo consiste en permitir la realización de una optimización del problema *MDP* utilizando todas las definiciones de distancia que necesiten los diversos expertos que analizan un mismo problema, por lo que el *MDP* es tratado en este trabajo como un problema con múltiples objetivos.

4.2. Formulación del Problema

El problema *MDP* multi-objetivo puede ser planteado como:

$$\text{Maximizar } \vec{div}(M) = [div^1(M) \dots div^p(M)]^T \in \mathbb{R}^p \quad (3)$$

donde la l -ésima función objetivo $div^l(M)$ se define como:

$$div^l(M) = \sum_{i < j} d_{ij}^l; \text{ donde } l \in \{1 \dots p\} \quad (4)$$

d_{ij}^l representa la l -ésima definición de distancia considerada.

5. Algoritmo Multiobjetivo para el Problema de la Máxima Diversidad

5.1. Búsqueda Exhaustiva

El problema de la diversidad máxima considerado en este trabajo es el *Multi-Objective Max-Sum Dispersion Problem*, definido en (3) y (4).

Por ser la primera vez que se plantea este problema en un contexto puramente multi-objetivo, no se cuenta todavía con un conjunto Pareto conocido. En consecuencia, se ha implementado un algoritmo de *Búsqueda Exhaustiva (BE)* para resolver el *MDP* multi-objetivo.

Si bien la BE puede encontrar la totalidad de los elementos más disímiles de una base de datos dada, queda claro que este algoritmo no representa una solución escalable para problemas *NP Hard*, pues si aumenta el tamaño de la base de datos o la cantidad de objetivos utilizados, el mismo ya no es capaz de realizar toda la búsqueda en tiempo útil, haciendo intratable el problema para la mayoría de los casos prácticos.

5.2. Algoritmo Evolutivo

El presente trabajo utiliza el *Nondominated Sorting Genetic Algorithm* versión 2, conocido como NSGA-II [3], un MOEA (*MultiObjective Evolutionary Algorithm*) reconocido como el más representativo y por ende utilizado en el estado del arte [19]. El Algoritmo 1 presenta el pseudo-código básico del NSGA-II para resolver el problema de la Máxima Diversidad, el cual recibe como datos de entrada: una base de datos de $|N|$ elementos, que conforman el conjunto N , un vector de definiciones de distancia dd (para este trabajo se tienen disponibles 50 definiciones distintas de distancia¹), la cantidad de elementos disímiles a seleccionar $|M|$, el tamaño de la población *PopSize*, probabilidad de cruzamiento p_c , probabilidad de mutación p_m y la cantidad de iteraciones *iter* a realizar por el algoritmo. Finalmente, este algoritmo retorna el mejor conjunto de soluciones no dominadas P y el frente Pareto correspondiente PF al final del proceso evolutivo.

¹ <http://www.ug.edu.ec/mdp/bases/>

Algoritmo 1 Algoritmo Genético NSGA-II para MDP

Entrada: N base de datos de elementos, dd vector de definiciones de distancia (2 a 50 métricas disponibles), M cantidad de elementos disímiles a seleccionar, $PopSize$ Tamaño de la población, p_c probabilidad de cruzamiento, p_m probabilidad de mutación, $iter$: número de iteraciones a realizar.

Salida: Solución P conjunto Pareto, PF frente Pareto $\max(div(M))$.

```

1:  $i \leftarrow 0$ 
2:  $P_i \leftarrow$  InicializarPoblación( $PopSize$ )
3: EvaluarPoblación( $P_i$ )
4:  $Q_i \leftarrow \emptyset$  [Población Auxiliar]
5: mientras CondiciónDeParada = falso hacer
6:    $Q_i \leftarrow$  CrearNuevaPoblación( $P_i, PopSize$ ) [usar selección, cruzamiento y mutación sobre  $J_i$  para crear una nueva población  $Q_i$  de  $PopSize$  individuos]
7:   EvaluarPoblación( $Q_i$ )
8:    $F_i \leftarrow Q_i \cup P_i$ 
9:    $\mathcal{F}_i \leftarrow$  OrdenamientoNoDominado( $F_i$ )
10:   $j \leftarrow 0$ 
11:  mientras  $|P_{i+1}| < PopSize$  hacer
12:    AsignarDistancia( $\mathcal{F}_{ij}$ )
13:     $P_{i+1} \leftarrow P_{i+1} \cup \mathcal{F}_{ij}$ 
14:     $j \leftarrow j + 1$ 
15:  fin mientras
16:  ORDENAR( $P_{i+1}, \geq_n$ )
17:   $P_{i+1} \leftarrow P_{i+1}[0 : PopSize]$ 
18:   $i \leftarrow i + 1$ 
19: fin mientras
20:  $PF \leftarrow \mathcal{F}_{i1}$ 
21: devolver P, PF

```

6. Resultados Experimentales

En esta sección se exponen los experimentos realizados. Como no se cuenta con un frente Pareto óptimo conocido, se toman los datos obtenidos por la BE como frente Pareto de referencia, cuando este conjunto puede ser computacionalmente calculado. Se ha trabajado con dos bases de datos a los efectos de buscar los $|M|$ elementos mas disímiles. Cada una de estas bases de datos consisten en un conjunto de datos numéricos generados aleatoriamente. Las bases de datos utilizadas en los experimentos están disponibles en <http://www.ug.edu.ec/mdp/bases/>. En los experimentos se han utilizado diferentes definiciones de distancia. Los datos que se presentan a continuación corresponden a $P = 2$ y $P = 5$ definiciones distintas de distancia, representando cada definición una matriz de entrada diferente que permite calcular una función objetivo del problema. En la Tabla 1, se detallan 4 de las 5 definiciones de distancia utilizadas en los experimentos, dado que la distancia Euclidiana fue definida en (1).

#	Nombre, Definición y Referencia	#	Nombre, Definición y Referencia
1	<i>Canberra Distance</i> $\sum_{k=1}^K \frac{ x_{ik} - x_{jk} }{ x_{ik} + x_{jk} }$ [2]	3	<i>Jaccard</i> $\frac{\sum_{j=1}^d (x_{ik} - x_{jk})^2}{\sum_{k=1}^K x_{ik}^2 + \sum_{k=1}^K x_{jk}^2 - \sum_{k=1}^K x_{ik}x_{jk}}$ [2]
2	<i>Sorensen</i> $\frac{\sum_{k=1}^K x_{ik} - x_{jk} }{\sum_{k=1}^K x_{ik} + x_{jk} }$ [2]	4	<i>Clark</i> $\sqrt{\sum_{k=1}^K \left(\frac{ x_{ik} - x_{jk} }{x_{ik} + x_{jk}} \right)^2}$ [2]

Tabla 1: Algunas definiciones de distancia utilizadas

A continuación se presentan tres casos diferentes de pruebas en los que se fue variando el tamaño de la base de datos de entrada, la cantidad de elementos disímiles a seleccionar y la cantidad de objetivos considerados.

Los algoritmos fueron implementados con el lenguaje de programación Java y ejecutados sobre un computador Intel i3 de 1,40 GHz, 12 GB de memoria RAM y sistema operativo Windows 10. El procedimiento básico consistió en tomar la base de datos de entrada y calcular todas las matrices de distancia inter-elemento (tantas como objetivos se necesiten considerar). Una vez obtenidas las matrices de distancia, se ha resuelto el problema de la Máxima Diversidad por Búsqueda Exhaustiva; luego el mismo procedimiento fue resuelto mediante la utilización del NSGA-II. A continuación se explican con más detalles los tres experimentos realizados. Para el experimento 1 se tomó una base de datos de 40 elementos y las definiciones de distancia *Euclidean Distance* y *Canberra Distance* como objetivos. Cuatro corridas fueron realizadas, buscando los 3, 6, 8 y 10 elementos más disímiles. Claramente se puede notar como aumenta significativamente el tiempo de procesamiento de la Búsqueda Exhaustiva en la medida que aumenta la cantidad de elementos disímiles a seleccionar. Para el caso de 10 elementos, se necesitaron casi 3 horas de cómputo para encontrar todas las soluciones. En la Tabla 2 se muestran los resultados obtenidos con la BE.

BE utilizando 2 objetivos				BE utilizando 5 objetivos				BE utilizando 2 objetivos			
Elem.	M	Cnt.Sol.	t (seg)	Elem.	M	Cnt.Sol.	t (seg)	Elem.	M	Cnt.Sol.	t (seg)
40	3	6	1,1	40	3	15	1,7	100	3	18	2,4
40	6	23	65,5	40	6	890	329	100	6	35	10.826
40	8	31	772	40	8	852	7.163	100	8	**	**
40	10	43	10.142	40	10	*	*	100	10	**	**

Tabla 2: Resultados de la Búsqueda Exhaustiva

En el experimento 2 se tomó la misma base de datos de 40 elementos, pero en vez de dos, se consideraron 5 definiciones de distancia como objetivos (*Euclidean Distance*, *Canberra Distance*, *Sorensen Distance*, *Jaccard Distance* y *Clark Distance*). En este caso, para la selección de los 10 elementos más disímiles ya no se pudieron encontrar todas las soluciones en un tiempo razonable por Búsqueda Exhaustiva (el tiempo de procesamiento fue de más de 7 horas).

Para el experimento 3, se tomó una base de datos de 100 elementos, las definiciones *Euclidean Distance* y *Canberra Distance* como objetivos y se repitieron las corridas para la selección de 3, 6, 8 y 10 elementos. Bajo éstas condiciones, la BE solo pudo encontrar soluciones para $|M| = 3$ y 6 elementos. Para la búsqueda de 8 y 10 elementos se tuvieron más de 8 horas de procesamiento sin poder encontrar todas las soluciones.

En la segunda parte de los experimentos, se utilizó el Algoritmo NSGA-II para resolver los mismos problemas arriba citados, utilizando $|M| = 3, 6, 8$ y 10 , y considerando 20, 50, 100 y 250 iteraciones. El NSGA-II fue parametrizado de la siguiente manera: popSize (Tamaño de la Población) = 100, p_c (probabilidad de cruzamiento) = 0,9 y p_m (probabilidad de mutación) = $1/|M|$

En la Tabla 3 se muestran los resultados obtenidos por el NSGA-II. En la columna t (seg) se muestra el tiempo de corrida del NSGA-II para las 250 iteraciones. Se puede notar que para estos casos, el NSGA-II es capaz de conseguir soluciones que se aproximan razonablemente al frente de referencia y en la medida que se aumentan la cantidad de iteraciones, lógicamente se encuentran soluciones más próximas al frente Pareto. En la Figura 1 se muestra la evolución para el experimento 3.

Instancias del Problema			BE		NSGA-II				t (seg)
Cnt.Obj.	Elementos	$ M $	Cnt.Sol.	t (seg)	Cnt. de Iteraciones				
					20	50	100	250	
2	40	3	6	1,1	4	3	4	5	4,29
2	40	6	23	65,5	9	14	17	21	4,81
2	40	8	31	772	16	20	21	26	4,64
2	40	10	43	10.142	10	14	30	33	4,61
5	40	3	15	1,7	3	11	14	14	21,51
5	40	6	890	329	7	33	52	186	19,27
5	40	8	852	7.163	22	48	67	93	17,69
5	40	10	*	*	37	86	83	83	18,24
2	100	3	18	2,4	8	10	13	15	8,47
2	100	6	35	10.826	11	27	22	32	7,25
2	100	8	*	*	8	13	29	37	7,58
2	100	10	*	*	10	16	18	41	8,43

Tabla 3: Resultados experimentales del NSGA-II comparados con la Búsqueda Exhaustiva (BE). La columna “Cnt.Sol.” muestra el número de soluciones encontradas por búsqueda exhaustiva mientras que las columnas “Cnt. de Iteraciones” muestran la cantidad de soluciones utilizando el NSGA-II con un número diferente de iteraciones

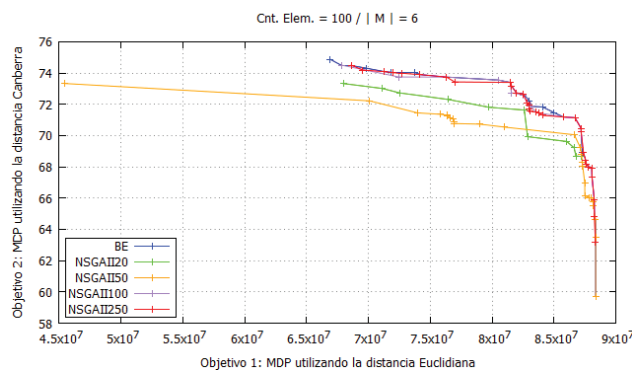


Figura 1: Experimento 3. BE vs NSGA-II con 20, 50, 100 y 250 Iteraciones

7. Conclusiones y trabajos futuros

Como se ha mencionado, el problema de la Dispersión Máxima *MDP* resulta relevante para áreas como biología, genética, medicina, para la justicia en la conformación de jurados, en las empresas para la conformación de equipos de trabajo entre otras aplicaciones. Existen varios modelos de diversidad que pueden ser aplicados, y a su vez en cada modelo se pueden utilizar diferentes definiciones de distancia. Dadas las características mencionadas, y como no existe un consenso entre los expertos sobre que definiciones de distancia utilizar, este trabajo propone por primera vez maximizar la diversidad de una base de datos determinada, con tantas definiciones de distancia como lo requieran los diferentes especialistas interesados en un problema. Se han presentado experimentos con bases de datos de 40 y 100 elementos respectivamente, y seleccionado los 3, 6, 8 y 10 elementos más disímiles de cada base de datos, para lo cual se utilizó un algoritmo evolutivo capaz de resolver el problema con tantas definiciones diferentes de distancias como sean necesarias. Como lo demuestran los resultados experimentales presentados, el algoritmo propuesto basado en el NSGA-II resuelve el problema de una manera razonable en tiempos admisibles. Como trabajo futuro se podrán considerar no solamente utilizar diferentes definiciones de distancia, sino también definiciones de diversidad combinadas, así como otros algoritmos evolutivos multi-objetivo, según sea el problema específico y la necesidad de los especialistas, ampliando también así, la visión que puedan tener los mismos para problemas específicos en los que estén trabajando.

Referencias

1. Brimberg, J., Mladenović, N., Urošević, D., Ngai, E.: Variable neighborhood search for the heaviest k -subgraph. *Computers & Op. Research* 36(11), 2885–2891 (2009)
2. Cha, S.H.: Comprehensive survey on distance/similarity measures between probability density functions. *City* 1(2), 1 (2007)

3. Deb, K., Agrawal, S., Pratap, A., Meyarivan, T.: A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: Nsga-ii. In: International Conference on Parallel Problem Solving From Nature. pp. 849–858. Springer (2000)
4. Della Croce, F., Grosso, A., Locatelli, M.: A heuristic approach for the max–min diversity problem based on max-clique. *Comp. & Op. Research* 36(8), 2429–2433 (2009)
5. de Freitas, A.R.R., Guimarães, F.G., Silva, R.C.P., Souza, M.J.F.: Memetic self-adaptive evolution strategies applied to the maximum diversity problem. *Optimization Letters* 8(2), 705–714 (2014)
6. Ghosh, J.B.: Computational aspects of the maximum diversity problem. *Operations research letters* 19(4), 175–181 (1996)
7. Glover, F., Ching-Chung, K., Dhir, K.S.: A discrete optimization model for preserving biological diversity. *Applied mathematical modelling* 19(11), 696–701 (1995)
8. Glover, F., Kuo, C.C., Dhir, K.S.: Heuristic algorithms for the maximum diversity problem. *Journal of information and Optimization Sciences* 19(1), 109–132 (1998)
9. Huang, A.: Similarity measures for text document clustering. In: Proceedings of the sixth new zealand computer science research student conference (NZCSRSC2008), Christchurch, New Zealand. pp. 49–56 (2008)
10. Kuo, C.C., Glover, F., Dhir, K.S.: Analyzing and modeling the maximum diversity problem by zero-one programming. *Decision Sciences* 24(6), 1171–1185 (1993)
11. von Lüken, C., Barán, B., Brizuela, C.: A survey on multi-objective evolutionary algorithms for many-objective problems. *Computational Optimization and Applications* 58(3), 707–756 (2014)
12. Martí, R., Gallego, M., Duarte, A., Pardo, E.G.: Heuristics and metaheuristics for the maximum diversity problem. *Journal of Heuristics* 19(4), 591–615 (2013)
13. Martí, R., Sandoya, F.: Grasp and path relinking for the equitable dispersion problem. *Computers & Operations Research* 40(12), 3091–3099 (2013)
14. McCune, B., Grace, J.B., Urban, D.L.: Analysis of ecological communities, vol. 28. MjM software design Glenden Beach (2002)
15. Meinel, T., Ostermann, C., Berthold, M.: Maximum-score diversity selection for early drug discovery. *J. of chemical information and modeling* 51(2), 237–247 (2011)
16. Palubeckis, G.: Iterated tabu search for the maximum diversity problem. *Applied Mathematics and Computation* 189(1), 371–383 (2007)
17. Prokopyev, O.A., Kong, N., Martinez-Torres, D.L.: The equitable dispersion problem. *European Journal of Operational Research* 197(1), 59–67 (2009)
18. Resende, M.G., Martí, R., Gallego, M., Duarte, A.: Grasp and path relinking for the max–min diversity problem. *Computers & Op. Research* 37(3), 498–508 (2010)
19. Riquelme, N., Von Lüken, C., Baran, B.: Performance metrics in multi-objective optimization. In: Computing Conference (CLEI), 2015. pp. 1–11. IEEE (2015)
20. Sandoya, F., Aceves, R.: Grasp and path relinking to solve the problem of selecting efficient work teams. INTECH Open Access Publisher (2013)
21. Sandoya, F., Martínez-Gavara, A., Aceves, R., Duarte, A., Martí, R.: Diversity and equity models. In: Handbook of Heuristics, pp. 1–20. Springer (2015)
22. Smyth, B., McClave, P.: Similarity vs. diversity. In: International Conference on Case-Based Reasoning. pp. 347–361. Springer (2001)
23. Wang, Y., Hao, J.K., Glover, F., Lü, Z.: A tabu search based memetic algorithm for the maximum diversity problem. *Eng. App. of Artificial Intell.* 27, 103–114 (2014)
24. Zhou, Y., Hao, J.K., Duval, B.: Opposition-based memetic search for the maximum diversity problem. *IEEE Transactions on Evolutionary Computation* (2017)

Extensión de una base de datos Objeto-Relacional para el soporte de datos de iris

Carlos Alvez, Graciela Etchart, Silvia Ruiz, Ernesto Miranda, Juan Aguirre

Facultad de Ciencias de la Administración, Universidad Nacional de Entre Ríos Concordia,
3200, Argentina
{caralv, getchart, sruiz, emiranda, juaagu}@fcad.uner.edu.ar

Resumen. Actualmente los sistemas biométricos han cobrado relevancia. En este trabajo se presenta una implementación de una extensión de bases de datos Objeto-Relacional para el dominio de biometría de iris. Si bien existen muchas extensiones de bases de datos para diferentes dominios, en ningún caso para aplicaciones biométricas. La extensión propuesta permite la gestión de datos biométricos del iris, tanto de las imágenes en sí mismas, como de sus metadatos asociados. La extensión está basada en estándares reconocidos internacionalmente, tanto para la gestión e intercambio de datos, como para la representación y compresión de las imágenes. Además, la extensión permite la gestión simple y eficiente de los datos del dominio, sobre todo, las consultas SQL se pueden realizar utilizando los métodos de búsqueda definidos en el sistema de tipos específicos de la extensión, aún por usuarios que no sean expertos en el dominio.

Palabras clave: Bases de datos, biometría, extensiones, Objeto-Relacional.

1 Introducción

Las bases de datos relacionales permiten la gestión eficiente de las aplicaciones tradicionales, que gestionan datos simples alfanuméricos. Sin embargo, para aplicaciones transaccionales más complejas, se ha extendido el uso de bases de datos Objeto-Relacionales (BDOR). Existen en la actualidad muchas extensiones de Sistemas de Gestión de BDOR (SGBDOR) para dominios específicos en el mercado, suministradas por las mismas empresas proveedoras de bases de datos o proyectos específicos, como por ejemplo Oracle Multimedia, que permite la gestión de videos, audio e imágenes en Oracle [1], Oracle Semantic Technology, que permite la gestión de modelos semánticos en Oracle [2], PostGIS, que permite el soporte de objetos espaciales y geográficos en PostgreSQL [3], etc., que intentan dar soporte a las aplicaciones que utilicen datos de estos dominios. Con soporte del dominio se hace referencia a tipos de datos y su comportamiento, métodos de accesos, operadores específicos, entre otros aspectos.

Sin embargo, más allá de las extensiones previstas, lo más importante que permiten realizar los SGBDOR, es crear nuevas extensiones para dominios específicos no

contemplados por éstos. Por ejemplo, los Data Cartridge provistos por Oracle [4], las extensiones de PostgreSQL [5], entre otros.

En relación a lo antes expuesto, no existen en la actualidad extensiones de bases de datos que brinden soporte para el dominio de las aplicaciones que utilizan datos biométricos. Sin embargo, existen muchas librerías y frameworks, muchos de ellos open source, que pueden ser de utilidad para la creación de estas extensiones, entre ellos se tienen, la librería como OpenCV [6] que permite el tratamiento de imágenes para soportar algunas de las funciones más genéricas; el sistema Masek and Kovesi [7] de la University of Western Australia para reconocimiento de iris open source y modular (en MATLAB); OSIRIS (Open Source para IRIS) [8] es un sistema de reconocimiento de iris desarrollado como framework del proyecto BioSecure (2007); el framework VASIR (Video-based Automatic System for Iris Recognition) [9] que cuenta con algoritmos de reconocimiento de iris implementado por el NIST para su utilización en la captura con video.

En este sentido, en el presente trabajo se pretende crear una extensión para el soporte de datos biométricos, en el caso particular de iris humano, que permita la gestión de imágenes de este rasgo y todos sus metadatos asociados; contemplando además, todos los métodos necesarios para la manipulación de los mismos, así como también, los métodos para búsqueda (verificación/identificación)¹.

Un punto importante para que la extensión propuesta sea lo más genérica posible y adaptable a las necesidades de los diferentes sistemas, es que la misma se base en estándares reconocidos internacionalmente, tanto para la gestión de los metadatos, como para la representación y compresión de las imágenes en sí mismas.

Lo que se pretende como principal objetivo de esta extensión, es que la gestión de los datos correspondientes a este dominio sea simple y eficiente, sobre todo, en la elaboración de consultas SQL. Es decir, utilizando los métodos de búsqueda definidos en el sistema de tipos implementados en la extensión, aún por usuarios que no sean expertos en el dominio.

En el resto del documento, los temas se organizarán de la siguiente manera: en la sección 2, se introduce a la problemática de iris y los estándares relacionados; en la sección 3, se presenta una breve descripción de la arquitectura de referencia para la implementación y las bases de datos objeto-relacionales; en la sección 4 se detalla la propuesta de implementación de la extensión para la gestión de iris y en la sección 5, se presentan las conclusiones y trabajos futuros.

2 Introducción a la problemática de iris y los estándares relacionados

La autenticación de personas basada en rasgos biométricos se ha vuelto muy popular en los últimos años como consecuencia de la baja en los costos de los sensores requeridos, su inclusión en dispositivos de consumo masivo y el surgimiento de

¹ Identificación, hace referencia a una búsqueda de 1 a n, donde se intenta determinar quién es la persona en cuestión. Mientras que verificación es una comparación 1 a 1, para determinar si la persona es quien dice ser.

vulnerabilidades debido al uso de múltiples claves de acceso a diferentes sitios que requieren cierto nivel de seguridad como ser cuentas de correo, sitios de banca electrónica, sistemas corporativos, etc.

De todos los rasgos utilizados en biometría, el iris, propuesto por Daugman en 1993 [10], es uno de los más estables a lo largo de la vida de una persona y también uno de los más difíciles de falsificar. Desde entonces ha habido varias propuestas en implementaciones para la gestión del iris, así como también estándares para permitir entre otras cosas, la interoperabilidad entre diferentes sistemas biométricos.

Uno de los estándares más importantes es el ANSI/NIST ITL 1-2011 [11], norma biométrica publicada en noviembre de 2011 y actualizada en 2015, 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.

Para el intercambio de datos de imágenes del iris se utiliza el registro Tipo-17 del estándar ANSI/NIST-ITL 1-2011. Entre las exigencias de conformidad de este registro, tiene además los requisitos para la representación y compresión de imágenes de iris.

Es importante asegurar que la región del iris tenga buena calidad y presente información fundamental para el reconocimiento; de lo contrario el sistema de reconocimiento tendrá un mal desempeño. Además, el intercambio de imágenes requiere en muchos casos transacciones web, con lo cual el tamaño de la misma es importante. El estándar ANSI/NIST-ITL 1-2011, en relación a las imágenes de iris, admite el uso de JPEG 2000 (Estándar ISO/IEC 15444:2004) [12] y PNG (Estándar ISO/IEC 15948:2004) [13].

El estándar internacional ISO/IEC 15444:2004 JPEG 2000 (Joint Photographic Experts Group) es un sistema de codificación de imágenes, que permite la compresión, transmisión y almacenamiento eficiente de imágenes fijas y de secuencias de imágenes. Esta norma define un conjunto de métodos de compresión sin pérdida y con pérdida para la codificación de dos niveles, en escala de grises de tono continuo, de color indexado, o de tono continuo en color con imágenes fijas.

En cuanto al estándar internacional ISO/IEC 15948:2004 PNG (Gráficos de Red Portables), éste especifica un flujo de datos y un formato de archivo asociado para una imagen sin pérdidas, comprimida, portátil y que puede ser transmitida a través de Internet. Soporta imágenes de color indexado, escala de grises y de color verdadero (RGB), con transparencia opcional. Además, permite profundidades desde 1 a 16 bits, es robusto y provee chequeo de integridad de archivos y detección de errores comunes durante la transmisión.

En la extensión que se presenta en la sección 4, se tendrán en cuenta los estándares citados en esta sección. A continuación, se presenta la arquitectura de referencia para la extensión propuesta.

3 Arquitectura de referencia y bases de datos Objeto-Relacionales

La arquitectura de referencia utilizada [14], contempla la integración de tres grandes módulos. El primero se encarga de la captura y el procesamiento de la imagen. Este módulo incluye la detección de las zonas de interés para la extracción de las imágenes de iris y la compresión de las imágenes según lo establecido en los estándares (JPEG2000 y PNG). El segundo módulo se encarga de las tareas de generación y transmisión de registros necesarios para una transacción ANSI/NIST IITL 1-2011, así como también, las operaciones necesarias para la recepción de estos registros de organismos asociados. En este trabajo, se utiliza el registro Tipo-17 del estándar citado que corresponde al rasgo iris utilizado para el intercambio de información entre organismos. El tercer módulo comprende la generación y el almacenamiento de metadatos necesarios para el reconocimiento, o sea, para los procesos de identificación y/o verificación de personas mediante el iris, lo que comprende el subsistema de reconocimiento de iris. Este módulo incluye todos los procesamientos de imágenes necesarios para la generación del código del iris (*iriscode*) [10] [15].

Las imágenes que gestionan los módulos mencionados, pueden obtenerse tanto de transacciones con otros organismos, como de capturas propias del sistema. Para esto último, se utiliza el módulo de captura y compresión de imágenes de iris.

Tanto la estructura del registro Tipo-17, como los metadatos para el reconocimiento del iris son estructuras complejas. Esto puede 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) [16]. Una de las mayores ventajas del modelo OR es la posibilidad de crear extensiones que permitan, entre otras cosas, 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, por ejemplo, el método de verificación que compara dos *iriscode*.
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.

En lo que concierne a este trabajo, esta extensión debe contemplar:

- La creación de tipos específicos para los datos biométricos: esto incluye tipos para los datos crudos (imágenes de iris en nuestro caso), tipos para la codificación de los mismos (*iriscode*), y todos los metadatos necesarios para generar los registros de transacciones ANSI/NIST-IITL1-2011.
- 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*) y métodos de generación e importación de registros ANSI/NIST-IITL1-2011.

Si bien no se encuentra en el alcance de este trabajo, se pretende en el futuro que esta extensión también contemple índices de dominio que permitan mejorar los tiempos de respuestas en los procesos de identificación.

En la siguiente sección se propone la implementación de la extensión basada en la arquitectura de referencia presentada en esta sección.

4 Implementación de la extensión propuesta

Las extensiones de los sistemas de gestión de bases de datos permiten definir para un dominio específico, entre otras cosas:

- Nuevos tipos de datos y su comportamiento
- Procedimientos y funciones
- Métodos de acceso y operadores

En este trabajo, se presenta una implementación de una extensión para un SGBDOR que provee nuevos tipos de datos y su comportamiento, y los procedimientos/funciones para la gestión de datos biométricos del iris.

Una de las principales decisiones para la implementación es qué SGBDOR utilizar. Se optó por PostgreSQL, debido a que es open source, y brinda muchas alternativas para generar nuevos métodos de acceso y operadores². Si bien, la indexación no es parte de este trabajo, se debe contemplar para la continuidad del desarrollo.

La implementación se basa en los tres módulos de la arquitectura de referencia tratada en la sección anterior. Para cada uno de estos módulos se analizaron diferentes librerías y frameworks existentes para ser utilizados en la implementación de los diferentes métodos, procedimientos y funciones.

4.1 Módulo de captura y procesamiento

En este módulo, se utilizó en primer lugar, la librería OpenCV para la obtención de las imágenes crudas. Esto incluye la detección en la imagen del rostro, ojo izquierdo y ojo derecho como se presenta en la Fig. 1. De aquí se obtienen las imágenes de ambos ojos. Luego para el procesamiento de imagen, se utilizaron las librerías OpenJPEG³ y LIBPNG⁴. Estas librerías permiten obtener formatos JPEG2000 y PNG respectivamente, que son los soportados por el estándar ANSI/NIST-ITL1-2011. Las imágenes capturadas se almacenan en instancias de tipos de objeto⁵ IMGJPG2000 y IMGPNG definidos en la extensión propuesta (Fig. 2 [14]). Estos tipos heredan del tipo de objeto ProcessedImage relacionado con las librerías OpenJPG y LIBPNG.

² Interfacing Extensions To Indexes. <https://www.postgresql.org/docs/9.5/static/xindex.html>

³ OpenJPEG. <http://www.openjpeg.org/>

⁴ LIBPNG. www.libpng.org/pub/png/libpng.html

⁵ Tipo de objeto (*object type*). En la terminología objeto-relacional, se refiere a un concepto similar al de clases en orientación a objetos. Las instancias de éstos son objetos.

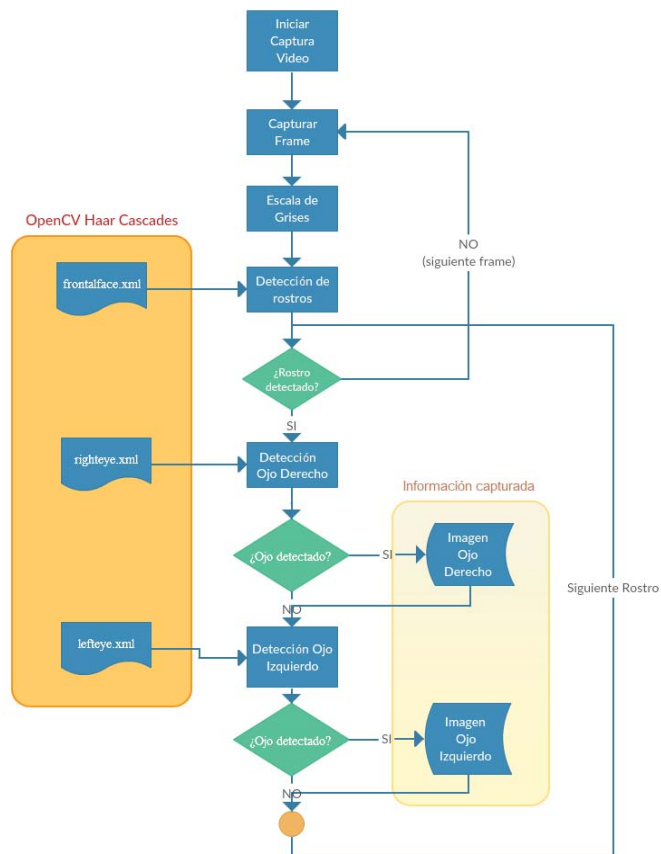


Fig. 1. Proceso de captura de los iris.

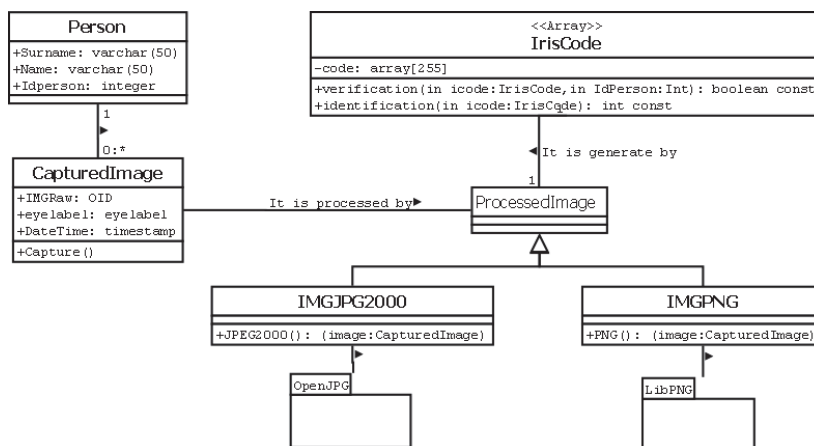


Fig. 2. Diagrama de clases con tipos de objetos requeridos para la captura y procesamiento de imagen y el reconocimiento de iris.

4.2 Módulo de reconocimiento de iris

Para este módulo se utilizó código del framework VASIR. La elección de este framework respecto de las otras opciones open source citadas en la sección 1, está motivada principalmente en dos puntos. Por un lado, su capacidad para analizar videos además de imágenes estáticas en el reconocimiento del iris. Por otro, cuenta con algoritmos efectivos para la verificación e identificación de sujetos bajo una amplia gama de imágenes y de diferentes condiciones ambientales (algoritmos implementados por el NIST).

Este módulo permite el pre-procesamiento y generación del código del iris (Fig. 3). Se crearon diferentes tipos de objeto para el procesamiento y gestión tanto de las imágenes como para el código del iris. Estos tipos contienen métodos para su comportamiento, por ejemplo, una instancia del tipo objeto IrisCode (Fig. 2), cuenta con los métodos miembros *verification* e *identification*. El método *verification* recibe como entrada una instancia de IrisCode y la identificación de la persona que dice ser; el método retorna verdadero si la distancia de Hamming de esta instancia respecto a la instancia almacenada en la base de datos para esta persona, se encuentra dentro del umbral establecido. El método *identification* recibe como entrada una instancia de IrisCode y retorna la identificación de la persona, cuya distancia de Hamming respecto a su instancia IrisCode almacenada en la base de datos, sea la menor y se encuentre dentro del umbral establecido. En caso que no se encuentre nadie por debajo del umbral, retorna cero.

Contando con estos métodos, las comparaciones (ya sea verificación o identificación), se pueden realizar con simples consultas SQL, utilizando los métodos de las instancias almacenadas en una tabla.

4.3 Módulo de intercambio de imágenes de iris

Este módulo permite las operaciones de importación y exportación del registro Tipo-17 del estándar ANSI/NIST-ITL1-2011. El formato utilizado es NIEM [17] (*National Information Exchange Model*) que define un espacio de nombres para el intercambio de datos biométricos mediante documentos de formato XML. Allí se define un documento de esquema XML (XSD) el cual permite validar los documentos XML generados en el proceso de exportación o antes del proceso de importación para asegurar la consistencia del documento (Fig. 4).

El proceso de importación consiste en tomar el documento XML generado en otro organismo. Luego el Parser XML valida que el documento esté bien formado y verifica su validez contra el esquema de NIEM. Verificada la validez, se procesan los campos y se genera una conexión con la base de datos Objeto-Relacional con el fin de almacenar los datos en la misma.

El proceso de exportación consiste en tomar los datos de la base de datos objeto-relacional, se procesan los campos y se construye el documento XML para la exportación. Finalizado este proceso se puede verificar, con el uso del Parser, que el documento generado sea correcto antes de enviarlo a otro organismo para su importación.

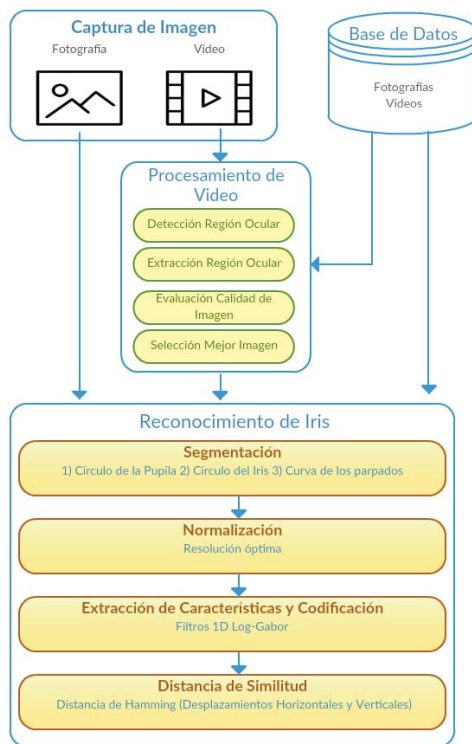


Fig. 3. Módulo de reconocimiento de iris.

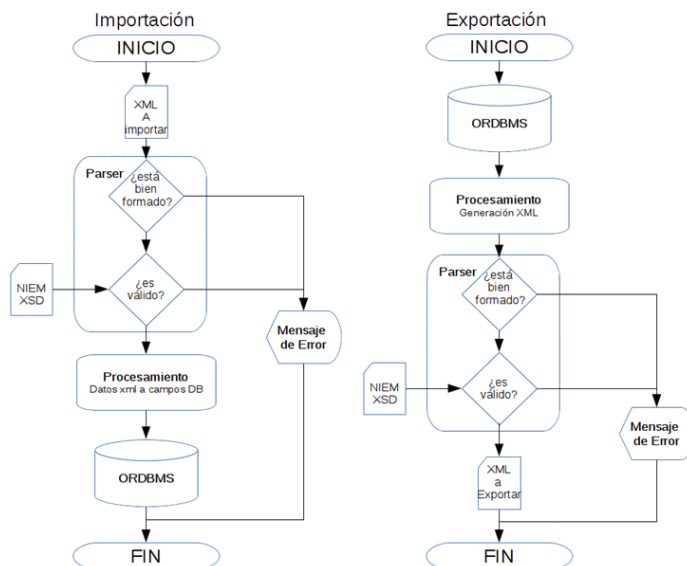


Fig. 4. Operaciones de importación y exportación de registros de Tipo-17 del ANSI/NIST-ITL1-2011.

5 Conclusiones y trabajos futuros

En el presente trabajo se muestra una implementación de una extensión de una base de datos Objeto-Relacional para el dominio de biometría, específicamente, para el iris humano. En la actualidad existen muchas extensiones de bases de datos para diferentes dominios, sin embargo, en ningún caso lo realiza para aplicaciones en biometría.

En el desarrollo de la misma se tomó una arquitectura de referencia [14] y para la implementación de los diferentes métodos y procedimientos se utilizaron librerías y frameworks open source disponibles. La extensión permite la gestión de datos biométricos del iris humano, tanto de las imágenes en sí mismas como de todos sus metadatos asociados; contemplando además, los métodos necesarios para la manipulación de los mismos, así como también, los métodos para búsqueda (verificación/identificación).

Un punto importante para la genericidad de la extensión propuesta, es que la misma está basada en estándares reconocidos internacionalmente, tanto para la gestión e intercambio de metadatos, como para la representación y compresión de las imágenes en sí mismas.

Por otro lado, la extensión permite que la gestión de los datos correspondientes a este dominio sea simple y eficiente, sobre todo, en la elaboración de consultas SQL, es decir, que se pueden realizar consultas SQL utilizando los métodos de búsqueda definidos en la extensión del sistema de tipos, aún por usuarios que no sean expertos en el dominio. Esto permite facilitar el desarrollo de aplicaciones biométricas basadas en el rasgo iris.

Como trabajos futuros, se pretende mejorar la performance en lo que refiere principalmente al método de identificación, buscando las alternativas de indexación que mejor se adecuen para este dominio.

Referencias

1. Pelski Sue. Oracle Multimedia Reference, 11g Release 2 (11.2). Part. E10776-03. (2010).
2. Murray Chuck. Oracle Database Semantic Technologies Developer's Guide, 11g Release 2 (11.2). Part. E25609-06. (2014).
3. PostGIS 2.3.4 Developer Manual - SVN Revision (15474). <http://postgis.net/stuff/postgis-2.3.pdf>.
4. Belden Eric, Timothy Chorma, Dinesh Das. Data Cartridge Developer's Guide - 11g Release 2 (11.2) - Part. E10765-02. (2010).
5. Extensiones en PostgreSQL. Chapter 35. Extending SQL. <https://www.postgresql.org/docs/9.5/static/extend.html>.
6. Laganieri Robert. OpenCV 3 - Transforming and Filtering Images [Video]. Packt (2017).
7. Masek Libor. Iris Recognition Thesis By Libor Masek. <https://www.mathworks.com/matlabcentral/linkexchange/links/1116-iris-recognition-thesis-by-libor-masek>.

8. Othman Nadia, Dorizzi Bernadette, Garcia-Salicetti Sonia. Osiris. Journal Pattern Recognition Letters archive. Volume 82 Issue P2, pp. 124--131. Elsevier Science Inc. New York, NY, USA. (2016).
9. Yooyoung Lee, Ross J. Micheals, P. J. Phillips, James J. Filliben, "VASIR: An Open-Source Research Platform for Advanced Iris Recognition Technologies", Journal of Research of NIST, Volume 118, pp. 218--259. (2013).
10. Daugman, J.: High condence visual recognition of persons by a test of statistical independence. IEEE Transactions on Pattern Analysis and Machine Intelligence 15(11), pp. 1148--1161. (1993).
11. Wing B. ANSI/NIST-ITL 1-2011. Update: 2015. Information Technology: American National Standard for Information Systems Data Format for the Interchange of Fingerprint, Facial & Other Biometric Information. (2015).
12. ISO/IEC 15444-1:2004 Information Tecnology-JPEG 2000 Image Coding System: Core Coding System, Second Edition. (2004).
13. ISO/IEC 15948:2004 Information Tecnology-Computer Graphics and Image Processing Portable Netware Graphics (PNG): Functional Specification, First Edition. (2004).
14. Ruíz, Silvia; Etchart, Graciela; Alvez, Carlos E.; Miranda, Ernesto; Benedetto, Marcelo Gabriel; Aguirre, Juan José. Iris Information Management in Object-Relational Databases. XXI Congreso Argentino de Ciencias de la Computación CACIC. (2015).
15. Daugman J. and Downing C., "Effect of severe image compression on iris recognition performance," IEEE Transactions on Information Forensics and Security, vol. 3, no. 1, pp. 52--61. (2008).
16. Melton Jim, "(ISO-ANSI Working Draft) Foundation (SQL/Foundation)", ISO/IEC 9075- 2:2003 (E), United States of America (ANSI). (2003).
17. NIEM National Information Exchange Model: "Biometric Schema Version 1.0". <http://release.niem.gov/niem/domains/biometrics/3.2/>.

Framework para el procesamiento lingüístico de artículos científicos.

Caso de estudio: Universidad Nacional de Chilecito

Jose Texier¹, Jusmeidy Zambrano¹, Emmanuel Frati¹

¹ Universidad Nacional de Chilecito, Chilecito-La Rioja, Argentina
{jtexier, jzambrano, fefrati}@undec.edu.ar

Resumen. El crecimiento de los datos producto del Acceso Abierto a las publicaciones académicas y científicas han generado estudios que posibilitan la interrelación de áreas como la lingüística y la computación en, por ejemplo, la extracción automática de datos para la creación de modelos teóricos, reconocimiento de estructuras para validación, etc. Este trabajo tiene como objetivo describir un framework elaborado a partir de la Rhetorical Structure Theory (RST) con el lenguaje Python a un corpus de 42 artículos científicos en español de la Universidad Nacional de Chilecito. El análisis se hizo sobre la base de un diseño modular informático y el trabajo manual de un lingüista experto, proceso que sirve para calibrar la propuesta y, luego, ser aplicada a un corpus mayor. Las variables que se consideraron fueron el título, resumen y palabras clave de los 42 artículos y la estructura canónica de un resumen científico (introducción, método, resultados y discusión -IMRD). Los resultados muestran que existen discrepancias en la frecuencia de ciertos elementos en los textos, pero a su vez, denotan coincidencias interesantes para este tipo de análisis textuales.

Palabras clave: reconocimiento de texto, retórica, Python, artículos científicos, Universidad Nacional de Chilecito, RST, IMRD.

1 Introducción

El auge de la tecnología ha transformado la investigación científica en el siglo XX [1], [2]. Con la llegada de internet y, sobre todo, con el movimiento de Acceso Abierto la producción académica y científica de la mayoría de las instituciones (por ejemplo: universidades, colegios, institutos, etc.) ha estado sujeta a controversias y acuerdos respecto de su visibilidad, preservación, difusión y uso. La tendencia actual es que la información académica y científica sea accesible sin restricciones legales y técnicas [3], [4], de esta manera se posibilita para algunas ciencias interdisciplinarias la elaboración de modelos computacionales que reproduzcan uno o más aspectos del lenguaje humano. Estos modelos en el campo de la lingüística computacional se pueden lograr gracias a teorías lingüísticas; análisis morfológicos y sintácticos de textos en un idioma y contexto determinado [5].

En relación con esta disposición de grandes de datos textuales, en este artículo se

centra en describir un *framework* para el análisis lingüístico de artículos científicos visibles en fuentes de Acceso Abierto, en función de la producción científica (artículos en español) de la Universidad Nacional de Chilecito (UndeC). De esta producción se obtienen los metadatos principales: título, palabras clave y resumen, autores y filiación. El trabajo se realizó bajo un enfoque lingüístico basado en Rhetorical Structure Theory (RST) y con el lenguaje de programación Python y el paquete NLTK [6]. Python (a pesar de no contar con rapidez de cómputo) ofrece diversas ventajas relacionadas con la flexibilidad, curva de aprendizaje rápida, funcionalidad dada por las librerías para el análisis de texto (extensible) y una sintaxis y semántica transparente [6]–[8].

Este trabajo aporta otro punto de análisis vinculado con la producción científica y académica de las instituciones y, está en consonancia con las prioridades de Argentina respecto de la preservación y difusión del conocimiento científico. Las funciones de la Universidad Pública [9] y la promulgación de la Ley 26.899 (*Creación de Repositorios Digitales Institucionales de Acceso Abierto, Propios o Compartidos*) sostienen las bases para que el conocimiento producido al interior de las instituciones públicas se difunda a la comunidad en general. Por tanto, los sistemas informáticos (Repositorios Institucionales y/o Bibliotecas Digitales) de muchas instituciones que permiten preservar y difundir la producción, aportan una gran cantidad de datos estructurados que sirven de base para desarrollar otras propuestas.

El artículo presenta en una segunda sección, un marco teórico que da cuenta, *grosso modo*, de algunos estudios previos que sustentan la propuesta realizada; la siguiente sección describe el *framework* desarrollado; luego, se analizan y discuten los resultados obtenidos y, en la última sección se plantean algunas conclusiones y sugerencias para trabajos futuros.

2 Marco de Referencia

La enorme cantidad de información digital disponible en áreas de conocimiento ha facilitado el desarrollo de propuestas vinculadas con: la extracción automática de textos [10], la creación de sistemas expertos que respalden la labor de especialistas en un área dada [11], los análisis de grandes corpus para generación de modelos teóricos, etc. [5].

Al tomar como punto de partida la extracción de textos para evaluar los conceptos y/o términos presentes en ellos, se pueden encontrar diversos enfoques que lo hacen de forma automática [10], [12]: i) lingüístico, ii) estadístico, iii) aprendizaje automático, y iv) métodos híbridos. De manera general, el enfoque lingüístico intenta filtrar mediante patrones de información. Los enfoques estadísticos, usan un número diferente de medidas y distribuciones estadísticas. Los sistemas de aprendizaje automático usan datos para aprender rasgos que sean útiles y relevantes para el reconocimiento [10].

En este trabajo se utilizó un enfoque mixto tomando como inicio el enfoque lingüístico según Acosta [10] al usar una teoría discursiva y luego herramientas de aprendizaje automático a través del paquete NLTK de Python [6]. El corpus se conformó con resúmenes en español de los artículos especializados en revistas arbitradas cuyos autores tuvieran como filiación institucional la UNdeC. Se consideró el resumen (*abstract*) porque es el género textual utilizado con asiduidad en la comunicación científica que da cuenta, en un número breve de palabras, los objetivos, marcos, metodologías, resultados y discusiones en un tema [13].

Para el estudio se vinculan las Ciencias de Computación y la Lingüística, puesto que ambas toman como objeto de estudio el lenguaje, y, en el caso específico, la lingüística computacional toma el Procesamiento de Lenguaje Natural (PLN) para habilitar a las computadoras en la tarea de procesar y entender el texto [14], [15]. Por ello, la propuesta se desarrolló sobre la base teórica del Procesamiento del Lenguaje Natural y la teoría discursiva Rhetorical Structure Theory (RST) de Mann y Thompson [16], que se ha empleado en diversas aplicaciones, como generación de texto, resumen automático, traducción automática, extracción de información, etc. [17]. La mayor parte de estas aplicaciones se han centrado en las lenguas: inglés, el japonés y el portugués [17]. El modelo usado para los resúmenes del corpus seleccionado fue el modelo de Swales [18], es decir, la estructura retórica: Introducción, Método, Resultados y Discusión (IMRD), elementos que se identificaron en el *framework* desarrollado.

Para el procesamiento del texto de los resúmenes, se analizaron cada una de las oraciones siguiendo la técnica de marcado del discurso (*Part-of-speech* - POS), que consiste en reconocer las entidades nombradas y extracción de información [19], [20]. El proceso se realizó en forma automatizada, a través de la asignación de una etiqueta (*tag* en inglés) de la categoría gramatical en cada palabra. También se usaron técnicas para identificar la raíz (o *stemmer*) de un conjunto de palabras similares, este proceso es conocido como Lematización, es decir, encontrar la raíz léxica de las palabras. Cuando un proceso de lematización no es posible de realizar, se recurre a un proceso de truncado (en inglés es *Chunker*), cuya finalidad es aproximar lo más posible las palabras a su raíz léxica. El etiquetado se basó en las propuestas por el grupo EAGLES (*Expert Advisory Group on Language Engineering Standards*) [21].

Al colocar en contexto teórico la propuesta desarrollada, se realizó un relevamiento de artículos que han trabajado de manera similar el *framework* propuesto, clasificado en tres grandes temas: análisis de los artículos científicos, generación de resúmenes automáticos/reconocimiento de términos, y análisis retóricos/contextos definitorios.

En cuanto al análisis de la estructura de los artículos científicos se encuentran tres trabajos, uno presenta un método para caracterizar el enfoque (contribución principal), dominio y técnicas/herramientas usadas en el artículo a partir de un matching con patrones semánticos usando el aprendizaje de bootstrapping [22]. El otro trabajo analiza los artículos científicos biomédicos a partir de tres esquemas: nombre de las secciones, zonas argumentativas y core de conceptos científicos, usando machine learning [23]. El tercer trabajo realizado en el 2008, se centra en la estructura retórica de los resúmenes científicos aplicando minería de texto gracias a

un corpus de Medline aplicando machine learning, identificando cuatro secciones objetivos, métodos, resultados y conclusiones [24].

El tema de generación de resúmenes automáticos es relevante porque representa el proceso inverso al *framework* propuesto. Teufel y Moens [25] realizan una investigación sobre el tema, sin embargo, se encuentran tres trabajos que extraen información en forma automatizada usando diferentes temas, pero utilizando principios de la generación de resúmenes automáticos [10], [26], [27]. El trabajo más importante de los tres es el laboratorio en línea presentado por Torres, ya que cuentan con un sistema de cuatro módulos: preprocesado, etiquetado POS, identificación de entidades nombradas (abreviatura en inglés NER) y *parse tree* [27].

En el tercer gran tema, se encuentran trabajos que se caracterizan por el uso retórico en los análisis de producciones científicas [24], [28], [29] y desde el punto de contextos definitorios [30], [31]. Hirohata et al. [24] identifica cuatro secciones en los abstract de las producciones científicas: objetivos, métodos, resultados y conclusiones. Prabhakaran et al. [28] hacen un análisis de 2.4 millones de abstracts de la Web of Science desde 1991 al 2010 y pueden determinar el crecimiento y/o declive de tópicos retóricos de ese corpus. En el tercer trabajo sobre análisis retóricos, los autores hacen un chequeo de las relaciones retóricas a partir de la RST (en Vasco) con el corpus de la revista médica de Bilbao [29]. En cuanto a la extracción de fragmentos textuales que contengan definiciones vinculadas por predicaciones verbales (contextos definitorios), se destaca el trabajo de Aguilar et. al. [30], porque definen un método enfocado a la biomedicina. El trabajo de Sierra [31], presenta un estado del arte en este contexto para entenderlo y estudiar la posibilidad de incorporar estas funcionalidades en una ampliación del *framework* desarrollado.

Actualmente, existen muchas herramientas que realizan PLN con virtudes y carencias según el objetivo deseado. A continuación se presenta un relevamiento de las herramientas de software encontradas:

- AnCora es un corpus del catalán (AnCora-CA) y del español (AnCora-ES) con diferentes niveles de anotación, <http://clic.ub.edu/corpus/es>.
- TreeBank, denominado también como Corpus parseado o más ampliamente Penn Treebank, <http://www.comp.leeds.ac.uk/amalgam/tagsets/upenn.html>.
- El RST Spanish Treebank, corpus en línea de textos especializados en español con relaciones discursivas de la RST de Mann y Thompson [16].
- Laboratorio en línea para el procesamiento automático de documentos [27].
- Stanford's Natural Language Processing, conjunto de herramientas de PLN, para documentos de texto los cuales pueden estar en diferentes idiomas [32]
- Natural Language Toolkit (NLTK), cuenta con un conjunto de bibliotecas de procesamiento de textos para la clasificación, simbolización, derivado, etiquetado, análisis sintáctico y semántico [6].
- TreeTagger, <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>.
- FreeLing, <http://nlp.lsi.upc.edu/freeling/>.
- Weka, <http://www.cs.waikato.ac.nz/ml/weka/>.
- Wagsoft Linguistic Software, <http://www.wagsoft.com/software.html>.

El *framework* desarrollado se ubica dentro del primer gran tema, análisis de la estructura de los artículos científicos, y se destaca que existe una implementación a

medida, a pesar de la existencia de una gran cantidad de herramientas, ya que la solución se posiciona en la teoría discursiva RST según la estructura retórica IMRD de Swales para resúmenes científicos.

3 Framework en Prueba

Las diferentes herramientas de software observadas y los trabajos relevados dan cuenta de que no existe un sistema específico que permita el análisis lingüístico de los resúmenes de artículos científicos, a partir del modelo de Swales [18]. De esta manera se puede verificar automáticamente que los resúmenes cumplan con la estructura textual canónica de “Introducción, Método, Resultados y Discusión - IMRD”. El *framework* desarrollado se analizó sobre la base de esa estructura canónica y con tecnologías de licencia abierta. En fases posteriores, el sistema evolucionará luego de un proceso de revisión, incorporando nuevas funcionalidades y analizando corpus más grandes. El sistema se realizó bajo un enfoque lingüístico basado en RST y con el lenguaje de programación Python. El *framework* está estructurado en seis etapas (ver Fig. 1):

1. Recolección de la producción científica de la UNdeC
2. Selección de los artículos científicos en revistas de investigación en idioma español.
3. Extracción de los metadatos necesarios para el análisis.
4. Análisis de palabras clave.
5. Análisis de los títulos.
6. Análisis de los resúmenes.

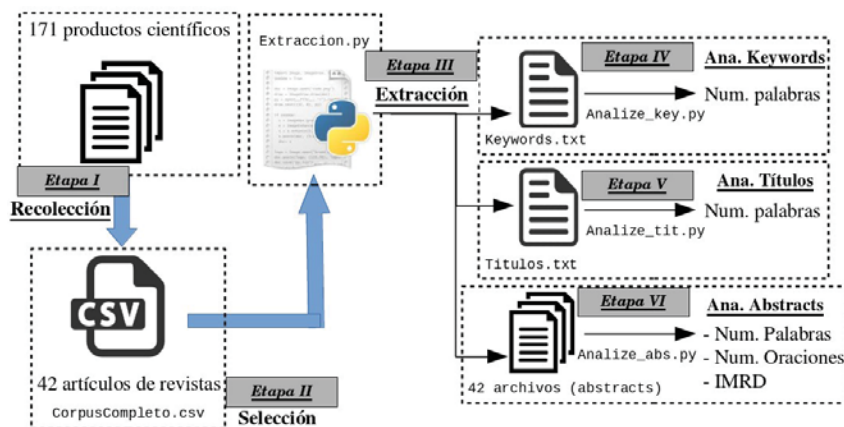


Fig. 1. Framework desarrollado.

3.1 Recolección

El corpus de la UNdeC, tomado el 30 de noviembre de 2016, se obtuvo de las siguientes fuentes: Google Scholar, RedALyC, SciELO y Scopus. Se realizó una

depuración de los resultados obtenidos con un total de 171 productos académicos y científicos con la afiliación de la Universidad Nacional de Chilecito (con al menos un autor) de esas cuatro fuentes. Los productos obtenidos tienen la siguiente tipología:

- Artículos en revistas: 100
- Conferencias: 53
- Libros: 3
- Reportes: 7
- Tesis: 8

Toda esta producción se incorporó en OpenRefine [33], herramienta que ofrece funcionalidades adicionales a los gestores de hojas de cálculo como LibreOffice Calc o Excel. Estos archivos y el proceso detallado de la investigación realizada se encuentran en un proyecto GitHub [34], el cual se puede consultar sin restricciones de acceso, legales ni técnicas.

3.2 Selección

De los 171 productos de la UNdeC, se seleccionaron los artículos de revistas en español. El total fue de 42 artículos extraídos en:

- Google Scholar, 24 artículos.
- RedALyC, 3 artículos.
- SciELO, 6 artículos.
- Scopus, 9 artículos.

Los metadatos de estos 42 artículos se almacenaron en un archivo CSV (*comma-separated values*) para adecuar la información a las siguientes etapas y poder adaptarlo a otros módulos en un futuro (Big Data por ejemplo).

3.3 Extracción

En esta etapa se realizó el proceso de extracción de metadatos del archivo CSV que proviene de la etapa II (*CorpusCompleto_UNdeC.csv*). En este proceso se usó el lenguaje de programación Python a través de un programa llamado *extraccion.py*. Esta etapa tiene como salida cuarenta y cuatro (44) archivos, que sirven de entrada para las etapas IV, V y VI. Los archivos son:

- Para la etapa IV se tiene el archivo *Keywords.txt*.
- En la etapa V se usó el archivo *Titulos.txt*.
- Para el análisis de los abstract (etapa VI) se usaron los 42 archivos que representan los 42 resúmenes del corpus seleccionado.

3.4 Análisis de palabras clave

De la etapa anterior se obtuvo el archivo “*Keywords.txt*”, el cual sirve de entrada

a un programa en Python llamado “`AnalyzeKeywords.py`”, donde se cuenta la cantidad de palabras, es decir, esta etapa consiste en totalizar las palabras clave que tienen los 42 artículos seleccionados. Al estar separada esta etapa en el *framework*, al igual que en las etapas restantes, permite incorporar a futuro cualquier tipo de proceso de acuerdo con los requerimientos deseados.

3.5 Análisis de títulos

En la etapa III se generó el archivo “`Titulos.txt`”, el cual sirve de entrada al programa de Python “`AnalyzeTitle.py`”, donde se totaliza la cantidad de palabras, es decir, el proceso es similar a la etapa IV.

3.6 Análisis de resúmenes

A diferencia de las etapas IV y V, en esta etapa se realizó el análisis lingüístico de acuerdo con los lineamientos establecidos, a partir de los 42 archivos generados en la etapa III que contienen los 42 resúmenes del corpus seleccionado, el proceso se realiza en el programa de Python llamado “`AnalyzeAbstract.py`”. En esta etapa se aprecia la modularidad diseñada en el *framework*, ya que se pueden realizar diversos procesos de forma independiente. Los resultados de esta fase se encuentran en la carpeta “`Analisis_Linguistico`” del proyecto GitHub [34]. En esta sección se realizan las siguientes actividades:

- Etiquetado de oraciones: se realizó a través de Python y el paquete Stanford CoreNLP [32], que tiene la posibilidad de hacer el etiquetado para el idioma español.
- Extracción de verbos: la nomenclatura para el etiquetado es la de EAGLE [21], que clasifica los verbos de tres maneras en forma general: principal (`vm`), auxiliar (`va`) y semiauxiliar (`vs`). Luego del etiquetado se extrajeron todas las palabras (o tokens) que tienen la clasificación de verbo.
- Identificación de verbos: luego de extraer los verbos, se busca el lema o lexema (su forma canónica) usando los procesos de truncamiento (*Stem*), que luego se comparan con un listado de verbos clasificados por los autores en las categorías IMRD, de acuerdo con la estructura retórica de Swales [18]. Todo esto gracias al paquete NLTK de Python [6].

4 Resultados

El proceso realizado consistió en hacer uso del *framework* y compararlo con la revisión manual realizada por el experto lingüista. Este análisis manual del experto se encuentra en el proyecto GitHub [34] con el nombre “`experto.pdf`”.

En cuanto a las diferentes etapas del sistema desarrollado produjeron los siguientes resultados:

- Para la sección de títulos, se encontraron 629 palabras, para un promedio de 14,98 palabras por título.
- En la sección de las palabras clave, se extrajeron 348 palabras, con un promedio de 8,49 por palabras clave.
- Para la sección de resúmenes, se obtuvieron 7.089 palabras, con un promedio de 172,90 por cada resumen.
- Un promedio de 6,42 oraciones por cada resumen de los artículos. El total de oraciones analizadas fue de 269, que pertenecen al corpus de los 42 artículos analizados.

En cuanto al análisis del experto, se resume con la Tabla 1:

- Las diferencias entre el análisis manual del experto y el *framework* desarrollado no exceden el 20%. Por ejemplo, en el corpus de resúmenes analizados el experto consideró que el 78,57% tiene una “introducción” y el *framework* detectó 85,71%, con una diferencia de 7,14%, lo que indica que hubo mucha coincidencia entre ambos análisis.
- En el caso de sección “discusión” la diferencia es la más alta (19,04%) lo que demuestra que los análisis fueron desiguales, para el experto, en los resúmenes solo el 38,10% tiene discusión y para el *framework* el 57,14%.
- En un total de 13 artículos (30,95%), el *framework* y el experto coinciden que esos artículos tienen las cuatro secciones en sus resúmenes de acuerdo con Swales. En cuanto a coincidencias en solo tres aspectos son 18 artículos (42,86%), 7 artículos (16,67%) en dos y 4 artículos (9,52%) en uno.

Tabla 1. Resultados del *framework* y del experto

	EXPERTO	FRAMEWORK	VARIACIÓN-DIFERENCIA
Introducción	78,57%	85,71%	7,14%
Método	78,57%	90,48%	11,91%
Resultados	59,52%	73,81%	14,29%
Discusión	38,10%	57,14%	19,04%

5 Trabajos Futuros

Esta investigación puede servir de base para realizar los siguientes trabajos:

- Un análisis de términos de acuerdo con los tesauros de la UNESCO y de la OECD para obtener una lista que posibilite un enfoque temático por áreas y/o disciplinas.
- Una implementación con todo el corpus de la UNdeC, ya que este trabajo se realizó como primera fase con un corpus de prueba de 42 artículos, verificados manualmente y obtenidos por el *framework*.

- El *framework* estará evolucionando para seguir con el análisis de todos los artículos de las conferencias del Latin American and Caribbean Consortium of Engineering Institutions - LACCEI - (aproximadamente 1.500) y con los artículos de la Universidad Nacional de La Plata (aproximadamente unos 25.000 artículos). En esta fase, se agregaran otras funcionalidades para que forme parte de un proyecto bajo los principios del Big Data [35]–[37], ya que implica el manejo de grandes cantidades de datos (volumen), la diversidad de temáticas (variedad) y la adaptación de un flujo continuo de datos en tiempo casi real (velocidad).
- Realizar un análisis semántico de la producción a partir del desarrollo de una ontología.
- Replicar este trabajo con un corpus pequeño en inglés, que permita hacer una verificación manual de los resultados del sistema y contrastarlos con el análisis de un experto.
- Generar trabajos interdisciplinarios (lingüistas e informáticos) al interior de las universidades para la enseñanza de los géneros académicos y/o científicos que generen instancias automatizadas de los procesos de escritura de los textos.

6 Conclusiones

Este trabajo se planteó como objetivo central, el describir un *framework* elaborado a partir de la Rhetorical Structure Theory (RST) con el lenguaje Python para un corpus de 42 artículos científicos en español de la Universidad Nacional de Chilecito. El análisis se hizo sobre la base de un diseño modular informático con tecnologías abiertas y el trabajo manual de un lingüista experto, lo que permitió hacer una implementación a pequeña escala para poder verificar de forma manual y evolucionar a un sistema con un corpus mayor. Las variables que se consideraron fueron el título, resumen y palabras clave de los 42 artículos y la estructura canónica de un resumen científico. El diseño se basó en módulos para que en un futuro, los módulos trabajen de forma independiente aceptando cualquier tipo de documentos y bajo los principios de Big Data, de manera tal que se puedan adaptar los análisis de otros datos como tuits o noticias diarias, por ejemplo.

Los resultados muestran que existen discrepancias en la frecuencia de ciertos elementos en los textos, pero a su vez, denotan coincidencias interesantes para este tipo de análisis textuales. Al comparar los dos resultados sobre la retórica IMRD (*framework* y experto) realizados, se puede concluir que existe una gran similitud pero no suficiente, ya que siempre se desea alcanzar un ciento por ciento de coincidencia. Un total de 13 artículos (30,95%) según el *framework* y el experto coinciden que esos artículos tienen las cuatro secciones (IMRD) en sus resúmenes de acuerdo con Swales, por tal razón, hay que estudiar más en profundidad las temáticas y exigencias de cada revista para poder tomar posición clara ante el bajo porcentaje de coincidencia. Además, serán necesarios otros niveles de análisis para redefinir la estructura textual y poder adaptar el *framework* a la realidad discursiva de algunas disciplinas.

En la actualidad la competitividad académica se refleja en la visibilidad web de la institución, por lo que se hace necesario que la institución logre estándares que permitan mayor impacto, por ello, los resultados obtenidos en esta investigación pueden ayudar a generar políticas institucionales sobre este aspecto.

7 Referencias

- [1] J. De Souza-Silva, J. Cheaz Peláez, and J. Calderón Romero, *La cuestión institucional, de la vulnerabilidad a la sostenibilidad institucional en el contexto del Cambio de Epoca*. Servicio Internacional para la Investigación Agrícola Nacional, 2001.
- [2] M. Castells, *The Rise of the Network Society: The Information Age: Economy, Society, and Culture Volume I*, 2nd Edition with a New Preface. Wiley-Blackwell, 2009.
- [3] J. Texier, “Los repositorios institucionales y las bibliotecas digitales: una somera revisión bibliográfica y su relación en la educación superior,” presented at the 11th LACCEI, 2013, Cancun, Mexico, 2013, p. 9.
- [4] P. Suber, “Ensuring open access for publicly funded research,” *BMJ*, vol. 345, 2012.
- [5] A. Domínguez Burgos, “Lingüística computacional: un esbozo,” *Boletín de lingüística*, no. 18, pp. 104–119, 2002.
- [6] S. Bird, E. Klein, and E. Loper, *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. O’Reilly Media, Inc., 2009.
- [7] R. Ali, *Python: Python For Beginners Crash Course Master Python Programming Fast and Easy Today*, 1st ed. USA: CreateSpace Independent Publishing Platform, 2015.
- [8] S. H. Edwards, D. S. Tilden, and A. Allevato, “Pythy: Improving the Introductory Python Programming Experience,” in *Proceedings of the 45th ACM Technical Symposium on Computer Science Education*, New York, NY, USA, 2014, pp. 641–646.
- [9] J.-I. Badell, C. Rovira, and M. Térmens, “Estudio de visibilidad web 2013 de los museos de Cataluña,” *Ibersid: revista de sistemas de información y documentación*, vol. 8, 2014.
- [10] O. L. Acosta, C. A. Aguilar, and T. Infante, “Reconocimiento de términos en español mediante la aplicación de un enfoque de comparación entre corpus” *Linguamática*. 2015.
- [11] G. Aquino and L. C. Lanzarini, “Keyword identification in spanish documents using neural networks,” *Journal of Computer Science & Technology*, vol. 15, no. 2, Nov. 2015.
- [12] J. A. Lossio-Ventura, C. Jonquet, M. Roche, and M. Teisseire, “Yet Another Ranking Function for Automatic Multiword Term Extraction,” in *Advances in Natural Language Processing*, 2014, pp. 52–64.
- [13] T. A. van Dijk, *La ciencia del texto: un enfoque interdisciplinario*. 1983.
- [14] “Asociación Mexicana para el Procesamiento del Lenguaje Natural Main/Home Page.” [Online]. Available: <http://www.ampln.org/>. [Accessed: 30-Apr-2017].
- [15] M. Vallez and R. Pedraza, “El Procesamiento del Lenguaje Natural en la Recuperación de Información Textual y áreas afines,” *Hipertext.net*, 2007.
- [16] W. Mann and S. Thompson, “Rhetorical Structure Theory: Toward a functional theory of text organization,” *Text-Interdisciplinary Journal for the Study of Discourse*, vol. 8, 2009.
- [17] I. D. Cunha, J.-M. Torres-Moreno, and G. Sierra, “Aplicaciones lingüísticas del análisis discursivo automático,” *Comunicación Social en el Siglo XXI*, vol. II, 2011.

- [18] J. Swales, *Genre Analysis: English in Academic and Research Settings*. Cambridge University Press, 1990.
- [19] A. Ritter, S. Clark, Mausam, and O. Etzioni, “Named Entity Recognition in Tweets: An Experimental Study,” in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Stroudsburg, PA, USA, 2011, pp. 1524–1534.
- [20] D. Ye, Z. Xing, J. Li, and N. Kapre, “Software-specific Part-of-speech Tagging: An Experimental Study on Stack Overflow,” in *Proceedings of the 31st Annual ACM Symposium on Applied Computing*, New York, NY, USA, 2016, pp. 1378–1385.
- [21] “Expert Advisory Group on Language Engineering Standards (EAGLES),” 25-Mar-2016. [Online]. <http://www.ilc.cnr.it/EAGLES96/home.html>. [Accessed: 30-Apr-2017].
- [22] S. Gupta and C. D. Manning, “Analyzing the Dynamics of Research by Extracting Key Aspects of Scientific Papers,” *IJCNLP*, 2011.
- [23] Y. Guo, A. Korhonen, M. Liakata, I. S. Karolinska, L. Sun, and U. Stenius, “Identifying the Information Structure of Scientific Abstracts: An Investigation of Three Different Schemes,” in *2010 Workshop on Biomedical Natural Language Processing*, USA, 2010.
- [24] K. Hirohata, N. Okazaki, S. Ananiadou, M. Ishizuka, and M. I. Biocentre, “Identifying Sections in Scientific Abstracts using Conditional Random Fields,” *IJCNLP*, 2008.
- [25] S. Teufel and M. Moens, “Summarizing Scientific Articles: Experiments with Relevance and Rhetorical Status,” *Computational Linguistics*, vol. 28, no. 4, pp. 409–445, Dec. 2002.
- [26] P. Saint-Dizier, “Processing natural language arguments with theplatform,” *Argument & Computation*, vol. 3, no. 1, pp. 49–82, Mar. 2012.
- [27] J. López, C. Sánchez-Sánchez, and E. Villatoro-Tello, “Laboratorio en línea para el procesamiento automático de documentos,” 2014.
- [28] V. Prabhakaran and O. Rambow, “Predicting Power Relations between Participants in Written Dialog from a Single Thread - P14-2056,” 2014. [Online]. Available: <http://www.aclweb.org/anthology/P14-2056>. [Accessed: 30-Apr-2017].
- [29] M. Iruskieta, “The RST Basque TreeBank: an online search interface to check rhetorical relations.”
- [30] C. Aguilar, O. Acosta, G. Sierra Martínez, S. Juárez, and T. Infante, “Extracción de contextos definitorios en el área biomedicina,” *Extraction of Definitional Contexts from Biomedical Corpora*, Sep. 2016.
- [31] G. Sierra, “Extracción de contextos definitorios en textos de especialidad a partir del reconocimiento de patrones lingüísticos,” *Linguamática*, vol. 1, no. 2, 2009.
- [32] “Stanford CoreNLP – Natural language software | Stanford CoreNLP.” [Online]. Available: <https://stanfordnlp.github.io/CoreNLP/index.html>. [Accessed: 27-Jun-2017].
- [33] OpenRefine, “OpenRefine,” 2017. [Online]. <http://openrefine.org/>. [Accessed: Apr-2017].
- [34] J. Texier and J. Zambrano, “Framework-PC - GitHub.” [Online]. Available: <https://github.com/dantexier/Framework-PC>. [Accessed: 30-Apr-2017].
- [35] V. Rajaraman, “Big data analytics,” *Reson*, vol. 21, no. 8, pp. 695–716, Aug. 2016.
- [36] A. Gandomi and M. Haider, “Beyond the hype: Big data concepts, methods, and analytics,” *International Journal of Information Management*, vol. 35, 2015.
- [37] M. Tascón, “Introducción: Big Data. Pasado, presente y futuro,” *Telos: Cuadernos de comunicación e innovación*, no. 95, pp. 47–50, 2013.

MetaCLAS: A Prototype Evolutionary Proposal to Automatically Suggest Clustering Methods and their Parameters

Macarena A. Latini, Rocío L. Cecchini, and Jessica A. Carballido

Instituto de Ciencias e Ingeniería de la Computación (UNS-CONICET),
Departamento de Ciencias e Ingeniería de la Computación
Universidad Nacional del Sur
San Andrés 800, Bahía Blanca, Argentina
`{macarena.latini,rlc,jac}@cs.uns.edu.ar`

Resumen Uno de los principales problemas al que nos enfrentamos al momento de realizar agrupamiento de datos consiste en elegir cuál es el mejor método de clustering para clasificarlos, y cuál es la cantidad ideal (k) de *grupos* en los que se deberían separar esos datos. En este trabajo presentamos una primera aproximación de un método que, a partir de un conjunto de datos estandarizados, sugiere el método de clustering y el valor de k que mejor los agrupa. Para esto considera cuatro índices de evaluación de la estructura final de clusters: *Dunn*, *Silüeta*, *Entropía* y *Widestgap*. El algoritmo está implementado como un algoritmo genético en el cual los individuos son posibles configuraciones de métodos de clustering y sus parámetros. En este primer prototipo, el algoritmo sugiere entre los métodos de partición K -means, PAM, CLARA y Fanny. Asimismo, además de sugerir el método que presentó mejor desempeño, también se obtiene como resultado el valor de los parámetros para ejecutarlo. El prototipo fue desarrollado en un entorno de R y se pudo corroborar que sus resultados son consistentes con una combinación de resultados provistos por otros métodos con objetivos similares. La idea de este trabajo es que sirva de base inicial para un desarrollo que incorpore opciones para reducción de la matriz de datos, evaluación de más métodos de agrupamiento y optimización de los operadores genéticos del algoritmo.

Keywords: algoritmos genéticos, clustering particional, computacion evolutiva

1. Introducción

Este artículo presenta una propuesta, basada en algoritmos evolutivos, para la resolución de uno de los principales problemas conocidos en el ámbito del *Clustering*, el cual consiste en la identificación de los métodos más adecuados para llevar a cabo el agrupamiento no supervisado de datos y la estimación de sus correspondientes parámetros. Además, el artículo pretende ser una guía básica de pasos a seguir al momento de realizar *Clustering* de datos, por lo que

se inicialmente mencionarán algunos aspectos ampliamente conocidos en el área.

¿Qué es el Clustering? El clustering o agrupamiento de datos consiste en la tarea de encontrar grupos de objetos de tal forma que quienes pertenecen a un grupo se parezcan más entre sí de lo que se parecen con objetos de otros grupos. La definición formal de *Clustering* (o agrupamiento), dependerá inevitablemente de la definición formal de *Cluster* y ambos términos parecen estar definidos de diversas maneras en la literatura. En [5] podemos encontrar una definición bastante simple para *Clustering*: “regiones contínuas del espacio que contienen una densidad relativamente alta de puntos, separadas de otras regiones del espacio por regiones cuya densidad de puntos es relativamente baja”. Si bien esta descripción parece bastante natural, cuantificar la cercanía de dichos puntos dependerá de los tipos de clusters presentes en el conjunto de datos sobre el que se trabaja. Ya que los datos pueden agruparse en formas compactas, alargadas o formando algún tipo de trazo dentro de cada cluster.

Interpretación y preparación de los datos. Como medida de interpretación general, en este artículo vamos a considerar que cada fila de la matriz de datos que se va a procesar es una *observación*, mientras que cada columna es una *variable*. Es decir, si la matriz de datos tiene dimensión $n \times m$, estaremos trabajando con n *observaciones* y m *variables*. Es importante tener en cuenta que los datos deben ser *estandarizados* previo a la aplicación del método de *clustering*, para lograr que las variables sean comparables entre sí.

2. Antecedentes

2.1. Conceptos básicos

Existen varias cuestiones a tener en cuenta al momento de elegir la estrategia con la cual se llevará a cabo el agrupamiento. Por ejemplo, si se tratará de un agrupamiento estricto, en cuyo caso cada elemento pertenecerá a un único grupo, o no. Si se podrá aplicar un método supervisado, para el cual se contará con casos etiquetados, o si sólo se podrá aplicar un método no supervisado.

Uno de los factores más importantes a considerar es cuál será la métrica que se usará para medir la distancia. El agrupamiento de las observaciones requerirá de algún método que permita evaluar la distancia (o similitud/dissimilitud) entre los elementos que se están intentando agrupar (en nuestro caso observaciones). Por medio de estas distancias se podrá conformar una matriz de distancias que es requerida por algunos de los métodos de clustering. Dos de las métricas más comunes para medir la distancia entre dos observaciones x e y son:

- *Distancia Euclídea*, definida como: $D_e = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$
- *Distancia Manhattan*, definida como: $D_m = \sum_{i=1}^n |(x_i - y_i)|$

En ambos casos x_i e y_i corresponden respectivamente a la i -ésima componente de los vectores x e y de longitud n . Existen otras métricas para medir distancia, como *Pearson*, *Kendall* o *Spearman* [17,18,10,11,21], las cuales están basadas en *correlación*.

Matriz de Distancia. Una vez definida la métrica de distancia a utilizar se puede calcular la *matriz de distancia*, M_d , de $n \times n$, tal que la componente $M_d[i, j]$ indicará la distancia (o similitud/dissimilitud) entre el vector (fila) i y el j de nuestra matriz original. Como es de suponer, esta matriz será simétrica y con *ceros* (o *unos*) en la diagonal. No todos los métodos utilizan esta matriz.

2.2. Estrategias clásicas para realización de Clustering

Como mencionamos anteriormente, las estrategias clásicas para intentar identificar las características de los grupos de datos subyacentes dentro de un conjunto de datos particular se dividen en dos grupos principales: *Métodos jerárquicos* y *Métodos no jerárquicos*.

Métodos jerárquicos. Estos métodos permiten trabajar con distintos tipos de variables y son útiles cuando no se conoce previamente el número de clusters, siempre y cuando el conjunto de datos no sea muy grande. Estos algoritmos pueden ser, a su vez, *Aglomerativos*, en los que en cada etapa se van agrupando los clusters calculados hasta el momento de tal forma de obtener un número menor de clusters, o *Divisivos*, que trabajan de manera inversa. En ambos casos se trabaja minimizando alguna función de distancia (o maximizando alguna función de similitud), para lo cual se utiliza la matriz de distancias (o similitudes).

Métodos no jerárquicos (o de particionamiento). Estos métodos, a diferencia de los *jerárquicos*, trabajan procurando alcanzar la mejor partición posible de los datos para un número dado de clusters (k). Este número debe ser determinado de manera previa a la ejecución del algoritmo. En general, estos métodos no trabajan sobre una matriz de distancia sino sobre los datos originales.

Existe una gran variedad de algoritmos no jerárquicos, entre los más comunes podemos mencionar *K-means* [14], *K-medoids* (PAM) [9], *CLARA*, *DBSCAN* [4]. A continuación se explicarán de forma resumida en qué consisten los métodos que se utilizaron en este trabajo.

K-means [14]. Este método, propuesto en 1967, es uno de los algoritmos no supervisados más utilizados para realizar el particionamiento de un conjunto de datos en k subconjuntos. La idea básica consiste en lograr determinar grupos de objetos de tal forma que se minimice la variación total dentro de los clusters (*within-cluster variation*) medida por el error cuadrado total (E_d) de la siguiente manera: $E_d = \sum_{i=1}^k \sum_{j=1}^{n_i} d(x_{ij}, c_i)^2$, donde, d es la métrica de distancia utilizada, k es el número de clusters, x_{ij} es la j -ésima observación del cluster i , y n_i y c_i son la cardinalidad y el centroide del cluster i .

La forma y el momento en que se recalcula dicho centroide ha dado origen a distintas variantes del algoritmo, como *MacQueen* [14], *Forgy/Lloyd* [6,13] y *Hartigan* [7].

PAM [9,12]. PAM proviene de Partitioning Around Medoids. Este método, propuesto en 1987, es una de las implementaciones más conocidas del algoritmo *K-medoids*. El objetivo del método es encontrar una secuencia de objetos (o

medoides) tales que se encuentren ubicados de manera central en los grupos o *clusters*. Es más robusto ante el ruido y *outliers* que *K-means* porque minimiza una suma de disimilaridades (entre pares de puntos) en vez de una suma de distancias euclídeas cuadradas. Algunos resultados experimentales han mostrado que PAM funciona relativamente bien para conjuntos de datos pequeños [12], pero se vuelve ineficiente con conjuntos de datos grandes debido a que el tiempo de ejecución de cada iteración aumenta en forma cuadrática con N [21].

CLARA [12,21]. En el método clara (Clustering LARge Applications) fue desarrollado con el objetivo de resolver los problemas de PAM frente a conjuntos de datos grandes. La idea subyacente de CLARA es procesar aleatoriamente una muestra X_i de tamaño N_i del conjunto de datos entero, X , y determinar el conjunto de *medoides*, θ_i que representa mejor a X_i utilizando PAM. CLARA utiliza PAM sobre un número de subconjuntos de X , denotados como X_1, \dots, X_m , de tal forma que en cada ejecución de PAM se obtiene un conjunto de *medoides*, $\theta_1, \dots, \theta_m$. Luego, se evalúa la calidad del agrupamiento asociado a cada uno de los conjuntos θ_i teniendo en cuenta el conjunto de datos entero. Trabajando de esta manera, el algoritmo logra requisitos de tiempo y almacenamiento lineales.

Fanny [12]. Fanny (Fuzzy Analysis clustering) es un algoritmo de particionamiento difuso (“fuzzy”). Es decir, a diferencia del resto de algoritmos vistos hasta ahora, las observaciones no se clasifican en un único cluster sino que se les asigna las probabilidades de pertenecer a cada grupo. Estas probabilidades no pueden ser negativas y deben sumar 1. Si u_{ij} es la probabilidad del elemento i de pertenecer al cluster j , el objetivo del algoritmo Fanny es minimizar la función $\sum_{v=1}^k \frac{\sum_{i=1}^n \sum_{j=1}^n u_{iv}^r u_{jv}^r d(i,j)}{2 \sum_{j=1}^n u_{jv}^r}$, donde n es el número de observaciones, $d(i, j)$ es la disimilitud entre las observaciones i y j y k es el número de clusters. El valor r se denomina *exponente de membresía o pertenencia* y si es cercano a 1 produce como resultado grupos menos difusos, mientras que si tiende a infinito se logra una difusión completa.

2.3. Validación de los Clusters

Si bien es claro que existen algunas características básicas que se desea alcanzar con el agrupamiento encontrado, como por ejemplo que la densidad entre los elementos del grupo sea superior a la densidad entre esos elementos y otros elementos externos, no es trivial evaluar qué tan bueno es dicho agrupamiento de la manera más adecuada posible. Además, se debe tener en cuenta que la bondad de cada grupo será relativa a la aplicación o problema que se está intentado resolver. Existe un gran número de índices que permiten evaluar diferentes aspectos del resultado de un algoritmo de clustering y, si bien se han clasificado de varias formas [1,21,2], la categorización más ampliamente utilizada es la que los agrupa en *internos* y *externos* [1,21]. La principal diferencia está en si la medida usa o no información externa para la validación, es decir, información que no es producto de la técnica de clustering. En general, de una u otra manera todas las medidas internas buscan analizar dos características principales de la

estructura: *cohesión* y *separación*. La primera busca que el miembro de cada cluster sea lo más cercano posible a los otros miembros del mismo cluster y la segunda apunta a tener clusters ampliamente separados. Los índices de validación interna más utilizados son *Dunn* y *Siluetas*. En este trabajo, se eligieron dos índices adicionales, *Entropía* y *Widestgap*, que permiten un análisis más global del resultado final del algoritmo.

Dunn [3,15]. El objetivo principal de este índice es dar un valor sobre la cohesión de los clusters analizando la varianza entre los miembros del cluster y la separación entre los clusters. La distancia entre los miembros de un cluster debe ser lo más pequeña posible y entre los clusters lo más grande posible. Su valor va desde cero hasta infinito y se calcula como:

$$D = \min_{1 \leq i \leq n} \left\{ \min_{1 \leq j \leq n, i \neq j} \left\{ \frac{d(i, j)}{\max_{1 \leq k \leq n} d'(k)} \right\} \right\} \quad (1)$$

donde $d(i, j)$ representa la distancia entre los clusters i y j , y $d'(k)$ mide la distancia dentro del cluster k .

Siluetas [15,19]. Este índice se utiliza para evaluar la compactitud (inter-cluster) y separabilidad (intra-cluster) de una estructura de clusters. El coeficiente de *Siluetas* para un agrupamiento se calcula a partir del valor $s(i)$ de cada elemento de la siguiente manera:

$$S = \frac{1}{N} \sum_{i=1}^N s(i) \quad (2)$$

donde $s(i)$ se calcula a partir de $a(i)$ (la distancia media entre el objeto y todos los otros objetos del mismo cluster) y $b(i)$ (la distancia media entre el objeto y todos los otros objetos del cluster más próximo) como: $s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$. Un valor de $s(i)$ cercano a cero indica que el objeto i está en la frontera de dos clusters. Un valor de $s(i)$ cercano a uno indica que el elemento se ha agrupado adecuadamente y, por el contrario, si el valor de $s(i)$ es negativo, es probable que el elemento haya sido incorrectamente agrupado.

Entropía. El tercer índice analizado mide la entropía de la distribución de los elementos de un cluster. En todos los ámbitos en que se analice, la entropía se concibe como una *medida del desorden*. Por esta razón, cuánto más chico sea este valor, mejor será la calidad de la estructura del clustering. Si bien esta medida suele ser utilizada como una herramienta para comparar dos *agrupamientos* diferentes, por medio de la medición de la *información mutua* existente entre los mismos, en este artículo sólo se utiliza como una forma de evaluar el desorden de un agrupamiento particular, funcionando como una métrica de validación interna. Según la definición de *entropía asociada a un agrupamiento C* dada en [16] se calcula como:

$$H(C) = - \sum_{k=1}^K P(k) \log P(k) \quad (3)$$

donde K es el número de clases del agrupamiento C y $P(K)$ es la probabilidad de que un elemento pertenezca al grupo k , dada por $P(k) = \frac{n_k}{n}$, donde n_k es

el número de elementos en el grupo k y n es el número de observaciones en el conjunto de datos.

Widestgap [22,8]. Finalmente, la cuarta medida considerada calcula la brecha más grande dentro de cada cluster, y retorna la mayor de todas. El valor del índice de *widestgap* para un cluster C se calcula como:

$$Wg = \max_{c \in C} wg(c), \text{ donde } wg(c) = \max_{\substack{c \in C \\ \text{para } D, E: \\ D \cup E = C \\ D \cap E \neq \emptyset}} \left\{ \min_{\substack{x \in D \\ y \in E}} d(x, y) \right\} \quad (4)$$

Como puede verse en esta ecuación, una vez calculados los valores de los espacios (o brechas) más grandes encontrados dentro de cada cluster (wg_c), Wg se define como el máximo de dichos espacios. Este espacio debería ser minimizado pues implica clusters más compactos.

3. Algoritmo propuesto: MetaCLAS

MetaCLAS es un algoritmo genético implementado en R que, a partir de una matriz de datos, propone un método de clustering y sus correspondientes parámetros. Hasta ahora, los posibles métodos son K -means, PAM, CLARA y Fanny, cuya elección depende de la calidad de los resultados obtenidos para la matriz de entrada. Los índices utilizados para evaluar la calidad del resultado del método de clustering son: Dunn, Silueta, Entropía y Widestgap.

Representación de los individuos. Cada individuo se compone por los campos “**MC**, **K**, **Algoritmo**”. **MC** es un entero que representa el método de clustering (K -means, PAM, CLARA o Fanny). **K** es la cantidad de clusters a obtener (fanny, clara) o una lista de centroides (K -means) o una lista de medoides (PAM) según corresponda para el método de clustering elegido en **MC**. Por último el campo **Algoritmo** guarda, en el caso de K -means, una referencia al nombre del algoritmo (*Hartigan-Wong*, *Lloyd*, *Forgy* o *MacQueen*), y en los otros tres casos (PAM, CLARA o Fanny) guarda el método utilizado para calcular la distancia entre dos observaciones.

Al momento de crear el individuo es importante mantener consistencia entre el método de clustering y sus parámetros. Por esto se realizó la validación del k según las restricciones del método siguiendo estas reglas: los métodos K -means, PAM y CLARA requieren que $0 < k < n$. El método Fanny requiere que $0 < k < (n/2) - 1$. Esto es considerado a lo largo de todo el algoritmo para conservar la factibilidad de los individuos.

La población inicial se crea de manera aleatoria respetando en cada caso las restricciones de los individuos, acorde a los parámetros requeridos por cada método. El método de selección utilizado es el de *torneo binario*.

Cruzamiento. Para el cruzamiento de dos individuos, se seleccionan los padres con una probabilidad $PC = 0,7$, y se cruzan aleatoriamente usando una de las siguientes opciones:

- Opción 1: Los hijos heredan el valor de k y reciben intercambiado el método de clustering de los padres.

- Opción 2: Los hijos heredan el método de clustering del padre y reciben intercambiado el valor de k .

Este proceso requiere una corrección del valor de k para preservar la propiedad de individuos factibles mencionada anteriormente.

Mutación. Para mutar a un individuo, utilizamos una técnica de reemplazo total. El individuo es seleccionado con una probabilidad $PM = 0,2$ y en su posición se genera un individuo nuevo.

Evaluación de la aptitud de cada individuo. La evaluación del fitness de un individuo determinado se hace en dos pasos. En el primer paso, se ejecuta el método correspondiente al individuo, utilizando los parámetros especificados para el mismo, para lo cual se utilizaron los paquetes *stats* y *cluster* de R. Luego, se utiliza la función *cluster.stats* {fpc} para validar el resultado del método. Dicha función calcula varias estadísticas de validez para una estructura de clustering y una matriz de disimilitud. En este caso, de todas las estadísticas que devuelve *cluster.stats*, se analizaron los valores de los índices Dunn, Silueta, Entropía y Widestgap. El objetivo final de la función de aptitud es maximizar los primeros dos índices y minimizar los últimos dos.

Dado que nos encontramos ante un problema de varios objetivos, en esta primera aproximación se decidió utilizar la forma más directa de abordarlo que es mediante una función de agregación. La misma consiste en combinar todas las funciones objetivo $f_i(x)$ en una única función $F(f_1(x), \dots, f_k(x))$. La función más utilizada es la combinación o agregación lineal de los objetivos en base a la siguiente ecuación:

$$F = \sum_{i=1}^k w_i f_i(x) \quad (5)$$

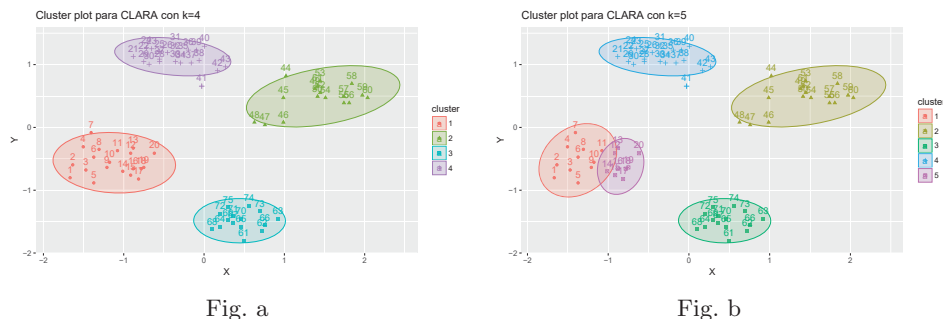
donde w_i son los pesos de cada función objetivo, siendo común que sean normalizadas, tal que la suma de todos los pesos sea igual a 1. En este trabajo todos los objetivos tienen igual peso. Los objetivos a maximizar se suman y lo demás se restan. Más específicamente, las funciones objetivo fueron determinadas como: $f_1 = D$, $f_2 = S$, $f_3 = -H$ y $f_4 = -W_g$ en base a las ecuaciones dadas en (1), (2), (3) y (4).

4. Evaluación

Para comprobar el desempeño del algoritmo se utilizó un ejemplo tradicional en el análisis de conglomerados. Se trata del conjunto de datos introducido por [20] que está compuesto por 75 observaciones sobre dos variables, x e y .

En la Figura 1 se pueden ver dos posibles soluciones de clustering encontradas con diferentes números de clusters. Se puede ver que en el grupo de cuatro clusters, la separación entre clusters es visualmente reconocible, mientras que a medida que nos movemos al caso de cinco grupos, la interpretación e incluso la definición de los grupos es menos clara.

Figura 1: Posibles agrupamientos encontrados para los datos *Ruspini*, para 4 y 5 grupos respectivamente.



4.1. Diseño de experimentos

La experimentación se organizó en 100 corridas independientes del MetaCLAS. Para cada corrida se registra la configuración de clustering sugerida por el mejor individuo utilizando los índices de evaluación presentados en la sección anterior. Una vez finalizadas las corridas, se coteja el resultado utilizando la función *NbClust* del paquete con el mismo nombre. Esta función utiliza 30 índices para determinar el mejor número de clusters. Sin embargo, a diferencia de nuestro método, no propone el algoritmo que arroja los mejores resultados. Sólo realiza el análisis usando un algoritmo no jerárquico, el *kmeans*, y uno jerárquico, el HAC (Hierarchical Agglomerative Clustering), sin dar la posibilidad de poder modificar estos métodos.

4.2. Análisis de resultados

En la tabla 1 mostramos un resumen de los valores obtenidos en las 100 corridas del MetaCLAS. Cabe destacar que solo una de las 100 veces, el algoritmo sugirió una configuración con 2 clusters. El resto de las corridas sugirieron entre 4 y 5 clusters. Es importante recordar que nuestro método decide cuándo una configuración es mejor que otra en términos de los índices Dunn, Silueta, Entropía y Widestgap. De acuerdo a lo que podemos ver en los resultados, nuestro algoritmo sugiere que utilizando el método CLARA con $k = 5$ obtenemos los mejores valores para los objetivos perseguidos en términos generales. No se ve una gran diferencia entre las métricas utilizadas para este método. Sí podemos ver que el método PAM con distancia Euclídea también demuestra una buena performance, para un $k = 4$. En este punto es cuando se pone en evidencia la importancia de incorporar medidas de evaluación externas que, en vista de estos resultados, ayuden a completar el análisis. Si sólo nos concentramos en el valor de k a sugerir, tendremos como resultado que nuestro algoritmo prefiere estructuras con 4 clusters. Cabe destacar que cada vez que se sugirió $k = 4$, la estructura fue idéntica para todos los casos, mientras que para $k = 5$ había distintas variantes.

Como dijimos anteriormente, validamos los resultados de este caso de estudio con la función *NbClust*. Al invocar dicha función con el dataset de *ruspini* y una

Tabla 1: Algoritmo más adecuado sugerido y número de *clusters* correspondiente para cada ejecución del MetaCLAS. En donde, H-W:Hartigan-Wong, L:Lloyd, F:Forgy, McQ:MacQueen, E:Euclídea, M:Manhattan.

	K-means				PAM		Fanny		CLARA		
	H-W	L	F	McQ	E	M	E	M	E	M	
k=4	5	1	3		14	8	9	8	8	12	68
k=5					5				14	12	31
	9				27		17		46		

variación de k desde 2 hasta 8, el resultado obtenido es que entre todos los índices:

- 1 propone que el mejor número de clusters es 2,
- 3 proponen que el mejor número de clusters es 3,
- 6 proponen que el mejor número de clusters es 4,
- 1 propone que el mejor número de clusters es 5,
- 2 proponen que el mejor número de clusters es 8,

y la conclusión a la que llega es que “*de acuerdo con la regla de mayoría, el mejor número de clusters es 4*”.

Este resultado nos da dos indicios: el primero es que en términos de la cantidad de clusters, estamos sugiriendo lo mismo que este método que es bien conocido y usado ampliamente en la literatura. El segundo es que elegimos correctamente cuatro índices que resumen las características deseables de una estructura de clusters. En cuanto al método que estamos sugiriendo (clara) no podemos constatar este resultado ya que no existe, hasta donde sabemos, un algoritmo cuyo objetivo sea también proponer el método más adecuado para una estructura determinada.

5. Conclusiones

En este artículo presentamos el algoritmo MetaCLAS, un algoritmo genético cuyos individuos representan distintas configuraciones de métodos de clustering, parámetros de los mismos y distintos valores de k . El algoritmo fue validado con el dataset *Ruspini*, el cual es ampliamente usado en la bibliografía de testeado de clustering. Los métodos entre los cuales nuestro algoritmo sugiere el que mejor se desempeña son los métodos de partición K -means, CLARA, PAM y Fanny. Para seleccionar la mejor configuración de método/parámetros/ k , usamos los índices de validación interna *Dunn*, *Siluetas*, *Entropía* y *Widestgap*. Luego de realizar 100 corridas independientes de nuestro algoritmo, pudimos comprobar usando el paquete *NbClust* de R que el método propuesto sugiere las mejores configuraciones. Por esto consideramos que el método presentado en este trabajo es un buen punto de partida para una futura implementación en la cual, tomando esta primera implementación como base, vamos a agregar la posibilidad de evaluar distintas reducciones de la matriz como parte del individuo para manejar grandes volúmenes de datos. Además planeamos incluir más métodos de clustering y como objetivo adicional, resta optimizar los operadores de cruzamiento y mutación del algoritmo genético.

Agradecimientos. Agradecemos a CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas), y a los subsidios PIP 112-2012-0100471, y UNS (Universidad Nacional del Sur) PGI 24/N042.

Referencias

1. Charu C. Aggarwal and Chandan K. Reddy. *Data Clustering: Algorithms and Applications*. Chapman & Hall/CRC, 1st edition, 2013.
2. Guy Brock, Vasyl Pihur, Susmita Datta, and Somnath Datta. clvalid: An r package for cluster validation. *Journal of Statistical Software, Articles*, 25(4):1–22, 2008.
3. J. C. Dunn. Well-separated clusters and optimal fuzzy partitions. *Journal of Cybernetics*, 4(1):95–104, 1974.
4. M. Ester, H. P. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on KDDM*, pages 226–231. AAAI Press, 1996.
5. Brian S. Everitt, Sabine Landau, and Morven Leese. *Cluster Analysis*. Wiley Publishing, 4th edition, 2009.
6. E. Forgy. Cluster analysis of multivariate data: efficiency versus interpretability of classifications. *Biometrics*, 21:768–780, 1965.
7. J. A. Hartigan and M. A. Wong. A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1):100–108, 1979.
8. Christian Hennig. Cluster validation by measurement of clustering characteristics relevant to the user, 2017. arXiv:1703.09282v1.
9. L. Kaufman and P. J. Rousseeuw. Clustering by means of medoids, 1987.
10. M. Kendall. A new measure of rank correlation. *Biometrika*, 30(1-2):81–93, 1938.
11. Maurice G. Kendall. *Rank Correlation Methods*. Griffin, London, England, 1970.
12. L. Kaufman and P. J. Rousseeuw *Finding groups in data: an introduction to cluster analysis*. Wiley-Interscience, 9th edition, 1990.
13. S. Lloyd. Least squares quantization in PCM. *IEEE Trans. Inf. Theor.*, 28(2):129–137, 1982.
14. J. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, pages 281–297, 1967.
15. U. Maulik *et al.* *Multiobjective Genetic Algorithms for Clustering - Applications in Data Mining and Bioinformatics*. Springer, 2011.
16. Marina Meil. Comparing clusterings an information based distance. *Journal of Multivariate Analysis*, 98(5):873–895, 2007.
17. Karl Pearson. Notes on the history of correlation. *Biometrika*, 13(1):25–45, 1920.
18. J.L. Rodgers and W.A. Nicewander. Thirteen ways to look at the correlation coefficient. *The American Statistician*, 42(1):59–66, 1988.
19. P. J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65, 1987.
20. Enrique H. Ruspini. Numerical methods for fuzzy clustering. *Inf. Sci.*, 2(3):319–350, July 1970.
21. Sergios Theodoridis and Konstantinos Koutroumbas. *Pattern Recognition, Fourth Edition*. Academic Press, 2009.
22. B. S. Villanueva, K. Gibert, and M. Sánchez-Marrè. Using CVI for understanding class topology in unsupervised scenarios. *CAEPIA*, volume 9868 of *Lecture Notes in Computer Science*, pages 135–149. Springer, 2016.

Prototipo de sistema de captura y monitoreo de datos OBD-II de vehículos

Claudio Aciti^{1,2}, Mauricio Urraco¹, and Elías Todorovich^{1,3}

¹ Universidad Nacional del Centro de la Pcia. de Buenos Aires, Tandil, Argentina, caciti@exa.unicen.edu.ar

² Universidad Nacional de Tres de Febrero, Departamento de Ciencia y Tecnología, Caseros, Argentina

³ Universidad FASTA, Facultad de Ingeniería, Mar del Plata, Argentina

Abstract. En este trabajo se diseñó y construyó un prototipo de un sistema de captura de parámetros de vehículos según tecnología OBD-II. El prototipo es capaz de leer distintas magnitudes generadas por un automóvil a través de una interfaz ELM-327 USB que permite la comunicación con la computadora de a bordo del vehículo en conjunto con los datos obtenidos mediante un GPS USB. Al prototipo desarrollado se le añadió un módem 3G y un adaptador Wifi USB. El sistema se implementó en una plataforma Raspberry Pi B con sistema operativo Raspbian. El módem 3G se utilizó para permitir el envío de la información recolectada a través de Internet a un equipo o central remota, y el adaptador WiFi USB permitió configurar el Raspberry Pi para poder ser utilizado como un Access Point de manera que fuese posible conectarse con un dispositivo móvil que soporte el sistema operativo Android para visualizar los datos leídos. Las pruebas fueron realizadas sobre un automóvil Chevrolet Corsa Wagon Life Gls 1.4 4p modelo 2009.

Keywords: OBD-II, Monitor, Raspberry Pi

1 Introducción

El sistema de diagnóstico a bordo (OBD) en vehículos es un sistema integrado que, mediante el uso de sensores, brinda al conductor y al técnico información relacionada al motor del vehículo y a su funcionamiento [8]. Con el avance en las tecnologías electrónicas incorporadas por los fabricantes en vehículos motorizados, han surgido una serie de beneficios asociados al mejor desempeño de los motores. Numerosos vehículos han utilizado sistemas de control electrónico para aumentar la eficiencia, tanto de los sistemas de alimentación e inyección de combustible como del encendido. A la par, se han desarrollado diferentes formas para diagnosticar los problemas asociados a estos nuevos dispositivos electrónicos y es así como en la actualidad una computadora a bordo controla sensores y actuadores que mantienen al motor funcionando bajo condiciones favorables. Al conjunto de actuadores, sensores y software de diagnóstico se le denomina Sistema OBD o Sistema de Diagnóstico a Bordo [5].

La estandarización de este sistema comenzó en los años 60, en EEUU, y se desarrolló entre fabricantes, gobierno y entidades norteamericanas preocupadas por el medio ambiente. A partir de 1970 se dió inicio a una serie de normas y requisitos de emisiones graduales para el mantenimiento de los vehículos por periodos prolongados de tiempo. Para cumplir con estos estándares, los fabricantes recurrieron a controles electrónicos de la alimentación de combustible y los sistemas de ignición. Los sensores miden el rendimiento del motor y ajustan los sistemas para proporcionar un funcionamiento óptimo. Además se puede acceder a estos sensores para obtener un diagnóstico temprano del vehículo [9]. En 1988, se establece un conector estándar y un conjunto de señales de prueba de diagnóstico. Esto dio lugar a la primer generación de requerimientos de sistema de diagnóstico a bordo, llamada OBD-I. La segunda versión del sistema de diagnóstico a bordo, llamada OBD-II, es una mejora tanto en capacidad como en normalización respecto a lo que fue la regulación inicial de OBD. El estándar OBD-II especifica el tipo de conector de diagnóstico y sus pines, los protocolos de señalización eléctrica disponible, y el formato de los mensajes [11,13].

Para poder llevar a cabo este monitoreo se utiliza un scanner, el cual se conecta a la interfaz de diagnóstico y permite el acceso a la información del vehículo. Existen scanners básicos y avanzados. Los básicos muestran desde códigos de error hasta información almacenada por el vehículo y los más avanzados poseen una interfaz que permite ver el estado de los sensores en tiempo real, gráficas y datos almacenados, los cuales facilitan enormemente el diagnóstico por parte del usuario. Estos últimos scanners tienen un alto costo, lo que dificulta su acceso por parte de mecánicos y empresas.

Hoy en día, muchas empresas disponen de flotas de vehículos que se utilizan para uso interno o para dar servicios de transporte. La utilización de un sistema recolector de información les permite controlar el uso de los vehículos para disminuir gastos en mantenimiento e inversión, prevenir fallas y accidentes, detectar malos hábitos de manejo de sus empleados, reducir gastos operativos de combustible, de neumáticos, etc. A nivel nacional existen algunas empresas que brindan servicios basados en métricas vehiculares. Por ejemplo, Sitrack, ofrece un sistema de control de flotas que permite el monitoreo del uso de los vehículos y el accionar de los conductores enviando la información de lo ocurrido en tiempo real a un equipo remoto [3]. Otra empresa argentina, Ful-Mar, provee una interfaz de comunicación con la computadora de a bordo que permite la lectura y registro en línea de la información capturada del vehículo en funcionamiento [1]. A nivel mundial, entre las más destacadas se encuentra Plug-N-Track, una empresa norteamericana que comercializa un dispositivo para servicios de seguimiento que provee geo-localización satelital y alerta sobre hábitos de manejo indeseados [2]. Si bien es un mercado que se encuentra en constante crecimiento, todavía hay mucho que avanzar con el uso de este tipo de tecnologías.

El objetivo principal de este trabajo es desarrollar un prototipo de un sistema recolector de datos para automóviles que permita registrar in situ lo que ocurre en el sistema de a bordo y enviar la información capturada a una central remota.

Además, desarrollar una aplicación en Android para que un usuario visualice los datos capturados a través de un dispositivo móvil.

Se definen entonces como objetivos desarrollar un prototipo para recolectar datos de automóviles capaz de:

1. Monitorear el sistema de a bordo en vehículos que cumplan con el estándar OBD-II para capturar parámetros del motor tales como velocidad, RPM, posición del acelerador, temperatura del refrigerante, torque, etc.
2. Obtener la geo-localización del vehículo mediante monitoreo satelital.
3. Conectarse a Internet haciendo uso de redes telefónicas y enviar los datos remotamente cuando la conexión esté disponible; en caso contrario almacenar los datos en la memoria interna hasta que se disponga de conexión.
4. Diseñar e implementar una aplicación para Android que permita al usuario, de manera opcional, conectarse al sistema de captura de datos para visualizar los datos que están siendo recolectados.

1.1 Limitaciones

El GPS presenta una limitación debido a que en algunas zonas puede no funcionar, o funcionar de manera incorrecta presentando mucha demora para obtener las coordenadas. De manera similar sucede con el módem 3G, debido a que en Argentina aún no se cuenta con buena señal de cobertura, es posible que en ciertas zonas no sea posible establecer la conexión a Internet. Por otra parte, existe el problema de que los dispositivos conectados al Raspberry Pi consumen mayor cantidad de corriente de la que puede entregar dicho dispositivo a través de sus puertos USB. Por ejemplo, el módem 3G cuando se encuentra intentando establecer la conexión puede llegar a consumir picos de hasta 800mA, pudiendo ocasionar problemas eléctricos que como consecuencia terminan provocando una falla permanente en el Raspberry Pi. Una solución a este problema es el uso de un HUB USB con alimentación independiente.

2 Trabajos previos

Existe una gran cantidad de referencias sobre aplicaciones de OBD-II. Entre las más relevantes, [4] presenta un sistema de adquisición de datos vehiculares para el manejo automático de flotas de automóviles empleando interfaces OBD, GPS, y tecnologías WiFi. Ellos se centran en el desarrollo de algoritmos para la toma de decisiones y sugieren interesantes aplicaciones para el control de tráfico. Además, los métodos donde explican el diagnóstico remoto del estado de los vehículos resultaron útiles para el desarrollo de este trabajo.

El método de posicionamiento presentado en [10] corrige y suplanta la velocidad inercial provista por un sistema de posicionamiento como GPS mediante el valor de velocidad del sistema OBD-II del vehículo en cuestión. Ese trabajo está enfocado en tener posicionamiento disponible en múltiples condiciones dentro de una ciudad. Si bien el trabajo que se presenta en este artículo también relaciona

medidas provenientes de un GPS y un sistema OBD-II, el foco aquí está en el monitoreo.

Los resultados de un sistema como el aquí propuesto, podrían servir de insumo en trabajos como [6]. Ahí se propone un método para calcular un recorrido con los objetivos de ahorrar combustible y reducir las emisiones de un vehículo. El método toma en cuenta el estilo de conducción del chofer, las características y componentes de las posibles rutas, e información de tráfico en tiempo real. Para evaluar el sistema se usaron conjuntos de datos que incluían datos OBD-II de taxis con posiciones GPS asociadas. [14] también se centra en el ahorro de combustible tomando en cuenta estilos de manejo de automóviles. El consumo e impacto ambiental se obtiene a partir de parámetros OBD-II. Los autores desarrollan un sistema que finalmente ofrece sugerencias a los choferes para ahorrar combustible. En la línea que apunta a reducir la emisión de gases, en [15] ya habían realizado un resumen de los factores que influyen en la emisión de contaminantes mediante estudios experimentales utilizando OBD.

En [12] el objetivo es el seguimiento y análisis de flotas de vehículos terrestres. Ese sistema mide de velocidad, distancia y consumo de combustible integrando datos OBD-II y GPS. Luego transmite la información mediante Wifi a un servidor remoto.

3 Desarrollo del sistema

El objetivo del presente trabajo es desarrollar un sistema ad-hoc capaz de comunicarse con la computadora a bordo OBD-II de un vehículo para poder obtener distintas métricas relacionadas con el funcionamiento del mismo. Dichos datos recolectados del motor del vehículo son complementados con la posición geográfica brindada por un GPS-USB y pre-procesados para luego poder ser enviados a un equipo remoto. La conectividad a Internet es lograda mediante el uso y configuración de un módem 3G. Los requerimientos funcionales para este prototipo de sistema de captura y monitoreo de datos OBD-II de vehículos, son:

- Establecer la conexión con el sistema OBD-II a bordo.
- Enviar mensajes y procesar las respuestas obtenidas.
- Leer una serie de PIDs y decodificar las respuestas de acuerdo a la fórmula correspondiente a cada magnitud.
- Obtener las coordenadas de la posición del vehículo a partir del GPS.
- Establecer y mantener la conexión a Internet mediante redes móviles.
- Almacenar la información recolectada para ser enviada por redes móviles a un equipo remoto cuando se disponga de Internet. En caso contrario, se debe almacenar la información hasta que pueda ser enviada.
- Proporcionar una interfaz gráfica a través de un dispositivo móvil (tablet o smarthphone) para visualizar los datos que están siendo recolectados.

Se decidió utilizar un Raspberry Pi ya que es una computadora reducida en algunas capacidades de hardware, lo que le permite principalmente producirse y comercializarse a precios bajos. Su tamaño y consumo de energía también

son reducidos. Sobre la plataforma corre una versión de Linux Debian, llamada Raspbian, aunque admite otros SO. Todas esas características son favorables para proyectos como el aquí presentado. En este caso se usó una Raspberry Pi B, que tiene dos puertos USB y 512 MB de RAM.

En la Fig. 1 hay una vista general del sistema que muestra como interactúa el Raspberry Pi con los dispositivos USB para capturar las magnitudes del motor del vehículo y los datos del GPS, establecer y mantener la conexión a Internet y utilizar la interfaz WiFi con el fin de convertirse en un Acces Point. Además se puede observar a grandes rasgos que la micro-computadora invoca a un servicio web para enviar los datos capturados a un equipo o central remota. Por otra parte un usuario puede mediante un dispositivo móvil con Android visualizar los datos a medida que éstos son leídos.

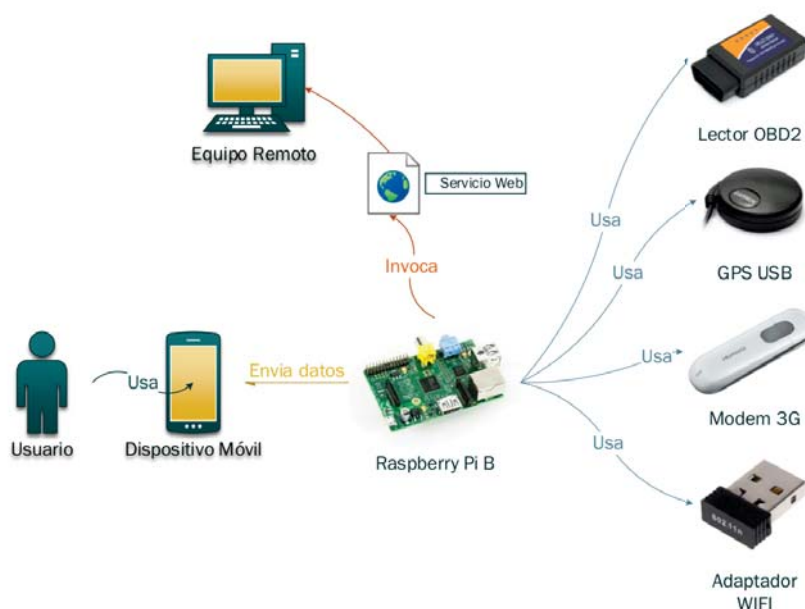


Fig. 1. Vista general del sistema.

Inicialmente se conectan todos los dispositivos a la Raspberry Pi: la interfaz ELM-327, el GPS USB, el adaptador WiFi y el módem 3G. Una vez que el vehículo se encuentra en marcha o encendido, se conecta el cable de alimentación desde el Raspberry Pi a la fuente conectada al vehículo. Cuando el micro-ordenador termina de iniciar el sistema operativo, el módulo recolector de datos, que compone el hilo principal del sistema, se inicia automáticamente. Este módulo intenta establecer comunicación con la interfaz ELM-327 mencionada previamente que sirve para leer la computadora de a bordo. Si ocurre un error

durante el establecimiento de la comunicación con la interfaz o si ésta fallara entonces todo el sistema fallaría, dando lugar a que no inicie. Por otra parte, si la conexión se establece con éxito, entonces el módulo recolector de datos inicia un bucle infinito donde se realizan las siguientes tareas:

1. Se invoca al módulo lector de GPS, implementado como un script en Python, y obtiene los datos leídos por este. Si no se reciben datos entonces el sistema continua su ejecución normalmente pero no reporta ningún valor para la latitud, la longitud y el tiempo.
2. Se leen las métricas de la computadora de a bordo OBD-II. En caso de que alguna de las lecturas falle entonces no se reportan valores para esa magnitud particular.
3. Los datos recogidos del vehículo y del GPS se escriben en paralelo a un archivo de datos y a un puerto TCP específico.
4. Cada 60 segundos se invoca al módulo de envío de datos, el cual es responsable de leer los datos escritos en el archivo, enviarlos a un equipo remoto, y vaciar el archivo leído.

Simultáneamente con este proceso recolector de datos, funciona el módulo disponibilizador de datos, el cual lee los datos recolectados a través del puerto TCP donde escribe el proceso anterior y se encarga de disponibilizarlos en forma de página web a través de otro puerto TCP. Una persona que disponga de un smartphone o tablet con sistema operativo Android y capacidad WiFi puede conectarse al Raspberry Pi como si éste fuera un Access Point y abrir la aplicación móvil desarrollada en este trabajo para visualizar los datos.

Entonces, el sistema se divide en módulos:

1. Módulo recolector de datos del vehículo
2. Módulo de envío de datos
3. Módulo disponibilizador de datos al dispositivo móvil
4. Módulo lector de GPS
5. Aplicación para el dispositivo móvil

Estos módulos se implementan de manera tal que cada uno es independiente del resto, para facilitar el testeado, y poder integrados de forma gradual.

3.1 Módulo recolector de datos

Este es el módulo central del sistema. Por un lado establece, la comunicación con la computadora de a bordo del vehículo (OBD-II) para solicitar los diferentes parámetros que se quieren monitorear. Para ello, inicializa el sistema embebido ELM-327 mediante comandos AT y luego solicita e interpreta las distintas respuestas mediante el Modo 1. Por otro lado, invoca a los módulos lector de GPS y de envío de datos a un equipo remoto.

El sistema embebido ELM-327 detecta e interpreta 9 protocolos distintos de de bajo nivel OBD. Traduce y devuelve la información en un formato único para

cualquiera de ellos, permitiendo la independencia con el protocolo interno que utiliza el vehículo. Para configurar el sistema ELM-327 se utilizan comandos AT [7]. Entre otras opciones, es posible desactivar el eco de los mensajes, desactivar la cabecera en la respuesta, o forzar el protocolo de comunicación a interpretar. A continuación se muestran algunos comandos AT a modo de ejemplo:

- AT Z → Reinicia el dispositivo y muestra su nombre.
Envío: atz
Respuesta: ELM327 v1.5
- AT DP → Muestra el protocolo actual que se está interpretando.
Envío: atdp
Respuesta: AUTO, ISO 14230-4 (KWP FAST)

Si bien el sistema OBD-II cuenta con varios modos de funcionamiento, en este trabajo se hace hincapié en el modo 1 o de flujo de datos, el cual permite acceder mediante un PID (Parameter Identification Data) a datos analógicos o digitales de salidas y entradas a la Engine Control Unit (ECU). Existen dos conjuntos de PIDs: los estándar y los propietarios. La mayoría de los PIDs OBD-II en uso no son estándar, pero para los vehículos más modernos, hay muchas más funciones compatibles con la interfaz OBD-II que están cubiertos por los PID estándar, y hay relativamente menor superposición entre fabricantes de vehículos para PIDs propietarios. Si bien existen más de treinta PIDs estándar, mediante los cuales es posible leer diferentes métricas del vehículo, en este trabajo se consideran solo cuatro para mostrar el funcionamiento del sistema y el método de lectura de parámetros.

Velocidad: El PID correspondiente a la solicitud de la velocidad es “0d” y la respuesta ocupa un byte. El rango válido va de 0 a 255 km/h y la fórmula para decodificar la velocidad es A , siendo A el valor de respuesta.

RPM (Revoluciones por minuto): El PID correspondiente a la solicitud de las revoluciones por minuto del motor es “0c” y la respuesta ocupa dos bytes. El rango válido va de 0 a 0 16.383,75 rpm y la fórmula para decodificar las RPM es $((A * 256) + B)/4$, siendo A y B el primer y segundo byte de la respuesta.

Posición del acelerador: El PID correspondiente a la solicitud de la posición del acelerador es “11” y la respuesta ocupa un byte. El rango válido va de 0 a 100% y la fórmula para decodificar esta posición es $A * 100/255$, siendo A el valor de respuesta.

Flujo de masa de aire: El PID correspondiente a la solicitud del flujo de masa de aire o MAF es “10” y la respuesta ocupa dos bytes. El rango válido va de 0 a 655,35 gramos/segundo y la fórmula para decodificar MAF es $((A * 256) + B)/100$, siendo A y B el primer y segundo byte de la respuesta.

3.2 Aplicación móvil

El módulo disponibilizador de datos se encarga de que los datos leídos de la computadora de a bordo OBD-II y del GPS estén disponibles para cualquier usuario en el vehículo con un dispositivo móvil Android con capacidad WiFi. La idea es

que el usuario se conecte a través de WiFi al Raspberry Pi como si este fuera un AP (Access Point) y se comunique a la dirección ip del Raspberry mediante un determinado puerto desde el browser del teléfono o desde una aplicación con un browser embebido.

La aplicación Android es básica. Tiene un browser embebido que se direcciona automáticamente a un puerto tcp establecido en la ip del Raspberry Pi conectado. Para el desarrollo de esta aplicación se utiliza un framework llamado Ionic-Framework el cual permite implementar aplicaciones para Android y iOS a partir de código HTML5 y código JavaScript, y que además soporta el uso de AngularJS, lo cual facilita el desarrollo. La aplicación solo cuenta con una pantalla en la que muestra los valores de los datos capturados: velocidad, rpm, posición del acelerador, flujo de masa de aire, latitud, longitud y tiempo, en ese orden.

4 Caso de prueba

El sistema desarrollado fue testeado en un Chevrolet Corsa Wagon Life Gls 1.4 4p 2009. Inicialmente se probaron los módulos de manera independiente y luego se procedió con las pruebas del sistema completo. Se debieron enfrentar diversos problemas de índole eléctrica. En primer lugar, debido a que el Raspberry Pi B sólo puede entregar una cantidad limitada de hasta 500mA a través de los puertos USB y que el módem 3G sólo ya produce picos de consumo de hasta 800mA, entonces es imposible alimentar el módem mediante el Raspberry Pi. Por lo tanto existen dos soluciones: la primera es utilizar un HUB USB alimentado el cual necesita una fuente adicional y la segunda es utilizar el nuevo modelo, Raspberry Pi 2B, el cual puede configurarse para entregar hasta 1.2A entre todos sus puertos USB. Por lo tanto, con una fuente de 2A o más es posible que el módem 3G y los demás dispositivos conectados funcionen correctamente sin necesidad de un HUB USB alimentado. En este trabajo se optó por utilizar una fuente adicional.

La captura que se ve en la Fig. 2 muestra los valores de los distintos parámetros mientras el vehículo está detenido y cuando se conduce a baja velocidad dentro de la ciudad. El motor se encuentra funcionando a casi 2200 RPM mientras que la velocidad es de alrededor de 30 km/h. La posición del acelerador reporta un valor mayor que cero, porque durante la conducción se presiona el acelerador.

5 Conclusiones y Trabajos futuros

Se logró desarrollar e implementar un prototipo de un sistema que recolecta datos y los envía a un equipo remoto a través de Internet. Este prototipo, tanto en hardware como en software puede ser extendido.

El prototipo permite monitorear el sistema de a bordo en vehículos que cumplan con el estándar OBD-II para capturar diferentes métricas. Además permite obtener la geo-localización del vehículo mediante monitoreo satelital utilizando un GPS.

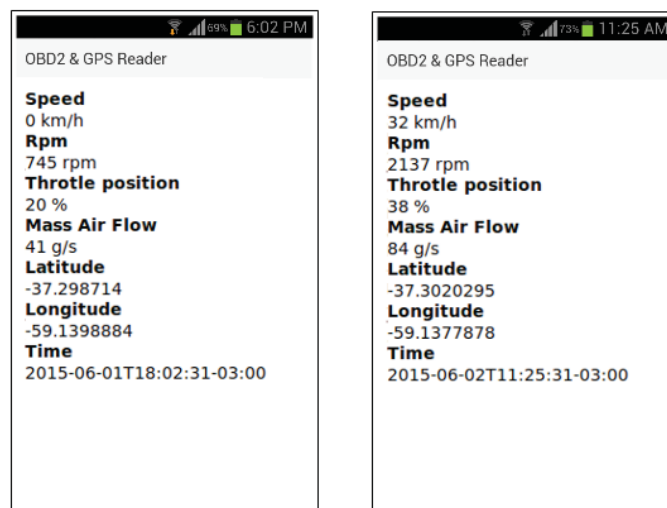


Fig. 2. Datos del vehículo encendido a velocidad cero y en marcha a baja velocidad.

Para enviar los datos, el sistema se conecta a Internet haciendo uso de redes telefónicas cuando la conexión este disponible; en caso de no tener conexión el sistema almacena los datos en la memoria interna hasta que pueda enviarlos.

Se desarrolló una aplicación móvil que permite al usuario conectare al sistema de captura de datos para visualizar los datos que están siendo recolectados.

Particularmente, en este trabajo, el prototipo debe ir instalado adentro de un vehículo en funcionamiento, por lo cual encontrar y adaptar una fuente de alimentación confiable y suficiente fue complicado. Como se explicó anteriormente, en este trabajo se optó por utilizar una fuente adicional.

Finalmente, y en base a lo apreciado durante las pruebas, se obtuvo una idea general del nivel de impacto que se puede esperar del uso de este prototipo.

Si bien este trabajo se centra en desarrollar un prototipo, a partir de esta base se prevee que es posible realizar extensiones como: Mejorar la presentación de los datos en la aplicación móvil; Capturar una mayor cantidad de métricas del motor del vehículo, de manera exhaustiva se pueden considerar todas las métricas estandarizadas para cualquier vehículo que soporte el protocolo OBD-II sin entrar en métricas específicas de cada marca; Se podría utilizar una batería independiente para alimentar el prototipo y así reemplazar la alimentación proveniente el automóvil. De esta forma se evita que el prototipo se apague si se detiene el vehículo o se corta el suministro eléctrico; Desarrollar un software que mediante diferentes algoritmos analice la información capturada e infiera datos más complejos tales como el desgaste del auto, el consumo promedio, reducción de las emisiones de gases, etc.

References

1. Ful-mar sa. <http://www.ful-mar.com.ar/>, accessed: 2017-07-21
2. Plugntrackgps. <https://www.plugntrackgps.com/>, accessed: 2017-07-21
3. Sitrack. <https://www.sitrack.com>, accessed: 2017-07-21
4. Aljaafreh, A., Khalel, M., Al-Fraheed, I., Almarahleh, K., Al-Shwaabkeh, R., Al-Etawi, S., Shaqareen, W.: Vehicular data acquisition system for fleet management automation. In: Vehicular Electronics and Safety (IC-VES), IEEE International Conference. pp. 130–133 (20011)
5. Corvalán, R., Osses, M., Urrutia, C., Barrientos, V.: Control de emisiones de fuentes móviles, informe final para la comisión nacional del medio ambiente. Tech. rep., Gobierno de Chile, Santiago de Chile (diciembre 2000)
6. Ding, Y., Chen, C., Zhang, S., Guo, B., Yu, Z., Wang, Y.: Greenplanner: Planning personalized fuel-efficient driving routes using multi-sourced urban data. In: 2017 IEEE International Conference on Pervasive Computing and Communications (PerCom). pp. 207–216 (March 2017)
7. Elm Electronics Inc.: Elm327 datasheet. Tech. rep., Elm Electronics Inc. (2008)
8. Henderson, B., Haynes, J.: OBD-II & Electronic Engine Management Systems. Haynes Repair Manuals, Haynes Manuals N. America, Inc. (2006)
9. Landin, C.A.M., Jimenez, U.Y.F.V.: Scanner Automotriz Interfaz PC. Master's thesis, Instituto Politécnico Nacional, Mexico D.F. (2010)
10. Lim, J., Choi, K.H., Kim, L., Lee, H.K.: Land vehicle positioning in urban area by integrated gps/beidou/obd-ii/mems imu. In: 2016 IEEE International Conference on Intelligent Transportation Engineering (ICITE). pp. 176–180 (Aug 2016)
11. Lopes, J.C.O.: Diseño de un escáner automotriz OBDII Multi-protocolo. Master's thesis, Universidad de San Carlos de Guatemala (2014)
12. Malekian, R., Moloisane, N.R., Nair, L., Maharaj, B.T., Chude-Okonkwo, U.A.K.: Design and implementation of a wireless obd ii fleet management system. IEEE Sensors Journal 17(4), 1154–1164 (Feb 2017)
13. McCord, K.: Automotive Diagnostic Systems: Understanding OBD-I & OBD-II. S-A Design Workbench Series, CarTech (2011)
14. Meseguer, J.E., Toh, C.K., Calafate, C.T., Cano, J.C., Manzoni, P.: Drivingstyles: a mobile platform for driving styles and fuel consumption characterization. Journal of Communications and Networks 19(2), 162–168 (April 2017)
15. Ortenzi, F., Campbell, F., Zuccari, F., Ragona, R.: Experimental measurement of the environmental impact of a euro iv vehicle in its urban use. In: SAE Technical Paper 2007-01-0966 (2007)

Sistema recomendador basado en tópicos latentes

María Emilia Charnelli^{1,3}, Laura Lanzarini², Javier Díaz¹

¹LINTI - Laboratorio de Investigación en Nuevas Tecnologías Informáticas

²III LIDI - Instituto de Investigación en Informática LIDI

Facultad de Informática, Universidad Nacional de La Plata

³ CONICET - Consejo Nacional de Investigaciones Científicas y Técnicas

mcharnelli@linti.unlp.edu.ar, laural@lidi.info.unlp.edu.ar,

jdiaz@unlp.edu.ar

Resumen. El filtrado colaborativo es una de las técnicas más utilizadas en los sistemas de recomendación. El objetivo del presente artículo es proponer un nuevo método que utiliza tópicos latentes para modelar los ítems a recomendar. De esta forma se incorpora la capacidad para establecer una semejanza entre estos elementos mejorando el rendimiento de la recomendación realizada. La performance del método propuesto ha sido medida en dos contextos muy diferentes arrojando resultados satisfactorios. Finalmente se incluyen las conclusiones y algunas líneas de trabajo futuras.

Palabras clave: Sistemas de Recomendación, Filtrado Colaborativo, Modelado de Tópicos Latentes.

1. Introducción

Los sistemas de recomendación analizan patrones de interés del usuario como artículos o productos, para proporcionar recomendaciones personalizadas que satisfagan sus preferencias [1]. Las sugerencias intervienen en varios procesos de toma de decisiones, tales como qué artículos comprar, qué películas mirar, o qué libros leer. El término ítem es utilizado para indicar lo que el sistema recomienda a los usuarios [2]. Para poder llevar esto adelante, es necesario modelar a los ítems que se quieren recomendar. La generación de un modelo a partir de la información textual y no estructurada de un conjunto de ítems representa un gran desafío. El análisis de tópicos latentes ha emergido como uno de los métodos más eficientes para clasificar, agrupar y recuperar datos textuales. Descubrir los tópicos en textos cortos es crucial para un amplio rango de tareas que analizan tópicos, como caracterizar contenido, modelar perfiles de intereses de usuarios, detectar tópicos latentes o emergentes. El modelo de tópicos bitérmino BTM [3] permite extraer de forma eficaz los tópicos que caracterizan a un conjunto de textos cortos. Con BTM se pueden obtener los temas subyacentes en un conjunto de documentos y una distribución global de cada tópico sobre cada uno de ellos, a través del análisis de la generación de bitérminos.

El enfoque más común para un sistema recomendador es la técnica de filtrado colaborativo basada en modelos de vecindad. Su forma original está basada en las similitudes entre usuarios [4]. Dichos métodos de usuario-usuario estiman puntajes desconocidos basados en puntajes registrados de usuarios con ideas afines. Posteriormente, se hizo popular el enfoque análogo pero ahora teniendo en cuenta las similitudes entre ítems [5] [6]. En estos métodos, se calcula un puntaje utilizando valoraciones realizadas por el mismo usuario en ítems similares. Una mejor escalabilidad y una precisión mejorada hacen que el enfoque por ítems sea más favorable en muchos casos [7] [8]. Además, los métodos de ítem-ítem son más susceptibles de explicar el razonamiento detrás de las predicciones. Esto se debe a que los usuarios están familiarizados con los elementos previamente preferidos por ellos, pero no conocen a los usuarios supuestamente parecidos. La mayoría de los enfoques de ítem-ítem utilizan una medida de similitud entre los ratings que tienen los mismos.

En este trabajo se propone un método basado en el enfoque ítem-ítem que utiliza un modelo de tópicos latentes para modelar a los ítems que se requieren recomendar y establece una semejanza entre estos elementos que mejoran el rendimiento de la recomendación. La evaluación del método propuesto se realiza mediante un conjunto de materiales educativos del repositorio digital Merlot[9] y un dataset de películas de MovieLens [10]. El presente artículo está organizado de la siguiente forma: la segunda sección describe el preprocesamiento efectuado sobre los datasets, la tercera sección muestra la extracción y modelado de tópicos latentes, la cuarta sección describe el método propuesto, en la quinta sección se muestran los resultados experimentales. Finalmente, en la sexta sección se presentan las conclusiones y las líneas de trabajo futuras.

2. Preparación de los datos

En este trabajo se utilizaron dos bases de datos, una de materiales educativos y otra de películas. La primera, brinda información de usuarios y materiales educativos del área de la Ciencias de la Computación publicados en el repositorio digital Merlot [9]. Los datos involucran más de 984 materiales y más de 260 usuarios que subieron, evaluaron o comentaron a cada una de las publicaciones. Así también, se dispone de información pública de las publicaciones y de los usuarios. De cada una de las publicaciones se obtuvo: título, tipo de material, fecha de creación, fecha de actualización, usuario que la realizó, valoración de revisores de 1 a 5, valoración de usuarios de 1 a 5, comentarios, y la descripción textual no estructurada. Mientras que el segundo dataset es sobre ratings de películas de MovieLens. Este dataset contiene 100.000 puntajes de 1 a 5 de 943 usuarios en 1682 películas, donde cada usuario evaluó al menos 20 películas; de las películas se conoce el título y la fecha; y además, se recolectaron los argumentos de cada una de ellas.

Cuando se trata de operar con información textual es preciso recurrir a técnicas de Minería de Texto (Text Mining) a fin de poder representar a cada descripción en un vector de términos. Esto fue llevado a cabo a través de un proceso

compuesto por varias etapas. En una primera etapa, se unificaron los contenidos en un único idioma. Luego se aplicó un filtro de stopwords, que se encarga de filtrar las palabras que coincidan con cualquier stopword indicado. Se filtraron stopwords del idioma inglés; palabras propias del contexto. También se eliminaron direcciones de páginas web, caracteres no textuales. Luego, cada palabra en el texto fue reducida a su raíz aplicando el algoritmo de stemming Snowball [11]. La importancia de este proceso radica en que elimina las variaciones sintácticas relacionadas con el género, número y tiempo verbal. Una vez que se obtienen las raíces de cada una de las palabras se calculó la frecuencia de aparición de cada una de ellas en las publicaciones y se escogieron las palabras que aparecen más de una vez.

3. Extracción de tópicos latentes

Para la extracción de los tópicos en las descripciones de los ítems se utilizó BTM (Biterm Topic Model) que es una técnica de aprendizaje no supervisado que descubre los tópicos que caracterizan a un conjunto de documentos breves.

Sea un conjunto de N_D documentos denominado corpus y sea W el conjunto de todas las palabras del corpus, un tópico se define como una distribución de probabilidad sobre W . Por lo tanto, un tópico puede ser caracterizado por sus T palabras más probables. Dado un número K de tópicos, el objetivo de BTM consiste en obtener las K distribuciones sobre cada una de las palabras. Un “bitérmino” denota a un par de palabras sin orden que co-ocurren en un documento corto. En este caso, dos palabras diferentes en un documento construyen un bitérmino. Dado un corpus con N_D documentos y un vocabulario W de palabras únicas, se supone que contiene N_B bitérminos $\mathbf{B} = \{b_i\}_{i=1}^{N_B}$ con $b_i = (w_{i,1} \in W, w_{i,2} \in W)$, y K tópicos expresados sobre W . Sea $z \in [1, K]$ una variable para indicar un tópico. La probabilidad $P(z)$ de que un documento en el corpus sea de un tópico z , se define como una distribución multinomial K -dimensional $\boldsymbol{\theta} = \{\theta_k\}_{k=1}^K$ con $\theta_k = P(z = k)$ y $\sum_{k=1}^K \theta_k = 1$. La distribución de palabras por tópico $P(w|z)$ puede ser representada como una matriz $\Phi \in R^{K \times W}$ donde la k -ésima fila ϕ_k es una distribución multinomial W -dimensional con entrada $\phi_{k,w} = P(w|z = k)$ y $\sum_{w=1}^W \phi_{k,w} = 1$. Dados los parámetros α y β , la suposición principal del modelo consiste en asumir que cada uno de los documentos del corpus fueron generados de la siguiente manera:

1. Se elige una distribución de tópicos $\boldsymbol{\theta} \sim \text{Dirichlet}(\alpha)$ para todo el corpus
2. Por cada tópico $k \in [1, K]$
 - Se extrae una distribución de palabras para el tópico $\phi_k \sim \text{Dirichlet}(\beta)$
3. Por cada bitérmino $b_i \in \mathbf{B}$
 - Se extrae una asignación de tópicos $z_i \sim \text{Multinomial}(\boldsymbol{\theta})$
 - Se extraen dos palabras $w_{i,1}, w_{i,2} \sim \text{Multinomial}(\phi_{z_i})$

Teniendo en cuenta el mecanismo de generación supuesto por BTM, se puede obtener la verosimilitud para todo el corpus dado los parámetros α y β a partir de la probabilidad de cada uno de los bitérminos:

$$P(\mathbf{B}|\alpha, \beta) = \prod_{i=1}^{N_B} \int \int \sum_{k=1}^K P(w_{i,1}, w_{i,2}, z_i = k | \boldsymbol{\theta}, \boldsymbol{\Phi}) d\boldsymbol{\theta} d\boldsymbol{\Phi} \quad (1)$$

$$= \prod_{i=1}^{N_B} \int \int \sum_{k=1}^K \theta_k \phi_{k, w_{i,1}} \phi_{k, w_{i,2}} d\boldsymbol{\theta} d\boldsymbol{\Phi} \quad (2)$$

Obtener exactamente los parámetros $\boldsymbol{\theta}$ y $\boldsymbol{\Phi}$ que maximizan la verosimilitud de la ecuación 2 es un problema intratable. Siguiendo lo propuesto en [12], los parámetros $\boldsymbol{\theta}$ y $\boldsymbol{\Phi}$ pueden ser aproximados utilizando muestreo de Gibbs [13].

Para inferir los temas de un documento, es decir, evaluar $P(z|d)$ para el documento d , se deriva la proporción de tópicos de un documento a través de los tópicos de los bitérminos. Si d contiene N_d bitérminos, $\{b_i^{(d)}\}_{i=1}^{N_d}$,

$$P(z|d) = \sum_{i=1}^{N_d} P(z|b_i^{(d)}) P(b_i^{(d)}|d) \quad (3)$$

3.1. Criterio de evaluación

Para evaluar la calidad de los tópicos obtenidos se utiliza la métrica de coherencia propuesta por Mimno et al. [14]. Dado un tópico z y sus T palabras más probables $V^{(z)} = (v_1^{(z)}, \dots, v_T^{(z)})$ donde $v_i^{(z)} \in W$ para $i = 1 \dots T$, el puntaje de coherencia es definido como:

$$C(z; V^{(z)}) = \sum_{t=2}^T \sum_{l=1}^t \log \frac{D(v_t^{(z)}, v_l^{(z)}) + 1}{D(v_t^{(z)})}$$

donde $D(v)$ es la frecuencia de la palabra v en todos los documentos, $D(v, v')$ es el número de documentos en donde las palabras v y v' co-ocurren. La métrica de coherencia está basada en la idea de que las palabras que pertenecen a un mismo concepto tenderán a co-ocurrir dentro de los mismos documentos. Esto es empíricamente demostrable porque el puntaje de coherencia está altamente correlacionado con el criterio humano. Para evaluar la calidad en general de un conjunto de tópicos, se calcula el promedio de la métrica de coherencia para cada uno de los tópicos obtenidos $\frac{1}{k} \sum_k C(z_k; V^{(z_k)})$. Estos resultados nos permiten determinar la cantidad de tópicos que mejor representan a todo el corpus.

4. Método propuesto

Sea K el número de tópicos que representan a un conjunto de ítems, se modelan cada uno de los mismos según la distribución de probabilidad mostrada en la ecuación 3.

Sea una lista de m usuarios $U = u_1, u_2, \dots, u_m$ y una lista de n ítems $I = i_1, i_2, \dots, i_n$. Cada usuario tiene una lista de ítems I_u , con un puntaje asociado a cada ítem r_{ui} . Cada ítem tiene asignado un puntaje de 1 a 5.

Con el objetivo de evaluar la semejanza entre dos ítems a partir de las distribuciones de probabilidad obtenidas con BTM, el método propuesto utiliza la divergencia de Kullback-Leibler [15]. Dadas dos distribuciones de probabilidad P y Q la función de divergencia se define como:

$$D_{KL}(P, Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}$$

A partir de esta divergencia, es posible definir la semejanza entre dos ítems p y q de la siguiente manera:

$$sim(p, q) = \exp(-D_{KL}(p, q))$$

Para estimar el rating de un nuevo ítem m dado un usuario u , se propone el siguiente método para predecir \hat{r}_{um} :

1. Se obtienen las distribuciones de probabilidad para cada uno de los ítems que el usuario evaluó I_u , según se mostró en la sección 3.
2. Se obtiene la distribución de probabilidad del material m .
3. Se calcula la semejanza sim de m con cada I_{u_j} .
4. Se ordenan las semejanzas, y se eligen las primeras t , donde t es un parámetro que define el tamaño de vecindad a considerar.
5. A partir de los t más semejantes, se calcula la predicción:

$$\hat{r}_{um} = \mu_m + \frac{\sum_{j=1}^t sim(m, j)(r_{uj} - \mu_j)}{\sum_{j=1}^t sim(m, j)}$$

donde r_{uj} es el puntaje del ítem j dado por el usuario u , y μ_j y μ_m son los puntajes promedios de j y m respectivamente.

5. Resultados experimentales

Se evaluó el modelo obtenido por BTM en el conjunto de los materiales educativos y de películas. Por cada número de tópicos entre 2 y 30 se promedió la coherencia obtenida, muestreando aleatoriamente el conjunto de prueba y entrenamiento en 1000 iteraciones. La figura 1 muestra el promedio de la coherencia del modelo con respecto a la cantidad de tópicos extraída en el dataset de materiales. Interesa el número de tópicos en el que se produce un quiebre en el crecimiento de la función de la coherencia promedio. En este caso, el valor óptimo se encuentra entre 5 y 7 tópicos latentes.

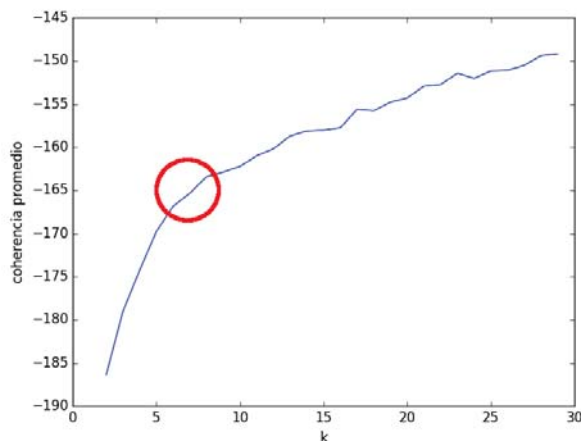


Figura 1. Dataset materiales. Coherencia promedio para distintos K

En la tabla 1 se muestran los tópicos obtenidos con $K=7$ para el dataset de materiales. Para cada uno de los tópicos se muestran las seis palabras más importantes, es decir, aquellas que tienen más probabilidad de pertenecer a dicho tópico.

Tópico	Palabras más importantes del tópico				
1	programming	software	data	algorithms	design
2	information	technology	computing	internet	systems
3	programming	java	language	tutorial	software
4	resources	design	systems	development	security
5	design	information	programming	interaction	human
6	binary	fractions	codes	numbers	tutorial
7	numbers	stars	interactivate	graph	simulation

Tabla 1. Dataset materiales. Modelo de tópicos obtenidos con BTM

La metodología de evaluación para el método propuesto aplica validación cruzada 10-fold. Este proceso de evaluación se repitió 50 veces para obtener una muestra significativa sobre la cual se promedian los resultados. Este proceso se aplicó para el método propuesto, identificado como KNN Topic Model, y sobre los métodos de filtrado colaborativo KNN, KNN Mean [7] basados en el enfoque ítem-ítem y usuario-usuario, SlopeOne [16] y sobre el método basado en modelos de factores latentes (SVD) [17].

El método propuesto y los métodos KNN reciben como parámetro la cantidad de vecinos a considerar. El tamaño del vecindario tiene un impacto significativo en la calidad de la predicción [4]. En la figura 2 se muestra el error RMSE (Root Mean Squared Error) para diferentes números de vecinos en los distintos algoritmos. El error decrece a medida que la cantidad de vecinos crece. El error para el método propuesto KNN Topic Model siempre está por debajo para distintos valores de vecindad. Además, se observa que luego de 40 vecinos el RMSE decrece lentamente para cada uno de los algoritmos.

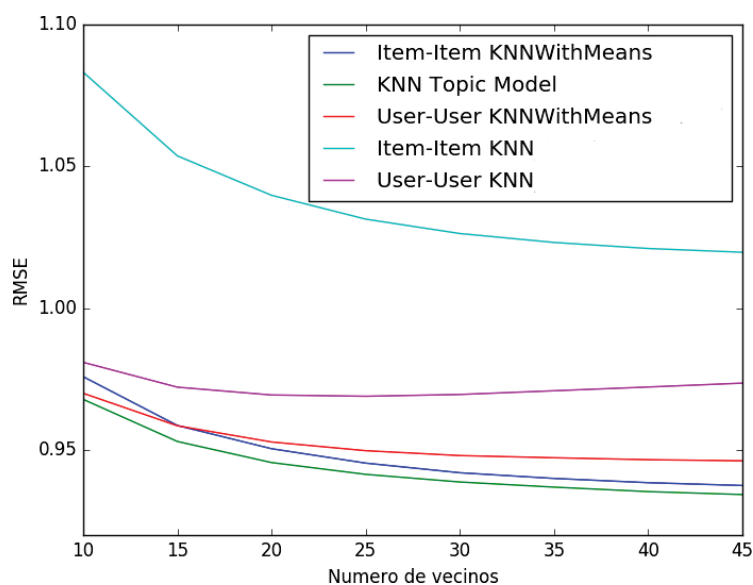


Figura 2. Dataset películas. Influencia del tamaño de vecindad

Los resultados de las 50 ejecuciones de la validación cruzada para cada algoritmo utilizando el dataset de materiales se muestran en la tabla 2 y los resultados al utilizar el dataset de películas se muestran en la tabla 3. Se estableció el número de vecinos $t = 40$ para todos los modelos basados en vecindad. Los ítems del dataset de materiales fueron representados como una distribución multinomial 7-dimensional y los ítems del dataset de películas como una distribución multinomial 10-dimensional. Para evaluar las predicciones del método propuesto frente a los resultados de los otros algoritmos, se calcularon las métricas de precisión RMSE (Root Mean Squared Error), MAE (Mean Absolute Error) y FCP (Fraction of Concordant Pairs), que mide la proporción de pares de ítems bien clasificados [18]. A diferencia de RMSE y MAE, el valor de FCP es mejor cuanto más alto es, porque mide una proporción.

Se observa que el método propuesto es competitivo frente a dos conjuntos de datasets diferentes. Para el dataset de materiales educativos el método KNN Topic Model obtiene un error RMSE más bajo y una proporción FCP más alta. Sin embargo, la métrica MAE es menor para SVD. Se destaca que con la poca información de los materiales que se disponen, a través de la utilización del modelado de tópicos es posible mejorar el FCP. En el dataset de películas se dispone de mayor información de los intereses de los usuarios, por lo que el método propuesto, si bien tiene un resultado competitivo, no supera el valor de FCP con respecto al enfoque KNN Mean usuario-usuario.

	KNN Model Topic	KNN item-item	KNN Mean item-item	KNN user-user	KNN Mean user-user	SlopeOne	SVD
Mean RMSE	0,6047	0,6848	0,8412	0,7757	0,6339	0,6575	0,6420
Mean MAE	0,4403	0,4566	0,5544	0,5126	0,4333	0,4336	0,3443
Mean FCP	0,6517	0,2075	0,4820	0,1400	0,3333	0,4133	0,4329

Tabla 2. Dataset materiales educativos. Resultados obtenidos

	KNN Model Topic	KNN item-item	KNN Mean item-item	KNN user-user	KNN Mean user-user	SlopeOne	SVD
Mean RMSE	0,9340	1,0203	0,9385	0,9732	0,9466	0,9426	0,9402
Mean MAE	0,7359	0,8044	0,7375	0,7680	0,7462	0,7409	0,7396
Mean FCP	0,6879	0,5990	0,6867	0,6948	0,6946	0,6865	0,6889

Tabla 3. Dataset películas. Resultados obtenidos

6. Conclusiones y líneas de trabajo futuras

En el presente artículo se logró modelar a un conjunto de ítems utilizando la detección de tópicos latentes a partir de las descripciones de los mismos. Esto permitió saber cuáles son los tópicos que describen a los ítems y cómo se relacionan entre sí. La metodología utilizada en el método propuesto y las métricas de validación aplicadas presentan resultados preliminares satisfactorios y competitivos frente a los métodos tradicionales. Como trabajo futuro se prevé la aplicación del método propuesto en otras bases de datos con información textual asociada. Resulta de interés, además, incorporar información acerca de las opiniones y gustos del usuario desde otros contextos. Los resultados de este trabajo se suman a lo previamente realizado en [19], donde se propuso un modelado

de usuarios a través de la información obtenida con BTM al identificar los tópicos de interés de los alumnos de la Facultad de Informática de la UNLP a través de sus publicaciones realizadas en grupos de Facebook. A su vez, este trabajo se relaciona con un proyecto más grande, cuyo objetivo es crear un sistema de recomendación de materiales digitales educativos.

Referencias

1. Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE transactions on knowledge and data engineering* **17** (2005) 734–749
2. Ricci, F., Rokach, L., Shapira, B.: Introduction to recommender systems handbook. In: *Recommender systems handbook*. Springer (2011) 1–35
3. Cheng, X., Yan, X., Lan, Y., Guo, J.: Btm: Topic modeling over short texts. *IEEE Transactions on Knowledge and Data Engineering* **26** (2014) 2928–2941
4. Herlocker, J.L., Konstan, J.A., Borchers, A., Riedl, J.: An algorithmic framework for performing collaborative filtering. In: *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, ACM (1999) 230–237
5. Linden, G., Smith, B., York, J.: Amazon. com recommendations: Item-to-item collaborative filtering. *IEEE Internet computing* **7** (2003) 76–80
6. Sarwar, B., Karypis, G., Konstan, J., Riedl, J.: Item-based collaborative filtering recommendation algorithms. In: *Proceedings of the 10th international conference on World Wide Web*, ACM (2001) 285–295
7. Bell, R.M., Koren, Y.: Scalable collaborative filtering with jointly derived neighborhood interpolation weights. In: *Data Mining, 2007. ICDM 2007. Seventh IEEE International Conference on*, IEEE (2007) 43–52
8. Takács, G., Pilászy, I., Németh, B., Tikk, D.: Major components of the gravity recommendation system. *ACM SIGKDD Explorations Newsletter* **9** (2007) 80–83
9. University, C.S.: Merlot - multimedia educational resource for learning and online teaching. <https://merlot.org> (2017 (accessed June 30, 2017))
10. Research, G.: MovieLens datasets. <https://grouplens.org/datasets/movielens/> (2017 (accessed June 30, 2017))
11. Gupta, V., Lehal, G.S.: A survey of common stemming techniques and existing stemmers for indian languages. *Journal of Emerging Technologies in Web Intelligence* **5** (2013) 157–161
12. Griffiths, T.L., Steyvers, M.: Finding scientific topics. *Proceedings of the National academy of Sciences* **101** (2004) 5228–5235
13. Geman, S., Geman, D.: Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence* (1984) 721–741
14. Mimno, D., Wallach, H.M., Talley, E., Leenders, M., McCallum, A.: Optimizing semantic coherence in topic models. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics (2011) 262–272
15. Kullback, S., Leibler, R.A.: On information and sufficiency. *The annals of mathematical statistics* **22** (1951) 79–86
16. Lemire, D., Maclachlan, A.: Slope one predictors for online rating-based collaborative filtering. In: *Proceedings of the 2005 SIAM International Conference on Data Mining*, SIAM (2005) 471–475

17. Mnih, A., Salakhutdinov, R.R.: Probabilistic matrix factorization. In: *Advances in neural information processing systems*. (2008) 1257–1264
18. Koren, Y., Sill, J.: Collaborative filtering on ordinal user feedback. In: *IJCAI*. (2013) 3022–3026
19. Charnelli, M.E., Lanzarini, L., Diaz, J.: Modeling students through analysis of social networks topics. *XXII Congreso Argentino de Ciencias de la Computacion CACIC 2016* (2016) 363–371

Uso de Bases de Datos Espacio-Temporales para la Atención de Eventos de Emergencia

Marcela Aballay¹, Anabella De Battista², Edilma Olinda Gagliardi³

¹ Maestría en Ing. en Sist. de Información
Facultad Regional Córdoba, Universidad Tecnológica Nacional,
Córdoba, Argentina
marcelaaballay@yahoo.com.ar

² Departamento Ingeniería en Sistemas de Información
Fac. Reg. Concepción del Uruguay, Universidad Tecnológica Nacional
Entre Ríos, Argentina

debattistaa@frcu.utn.edu.ar

³ Departamento de Informática
Univ. Nac. de San Luis
San Luis, Argentina
oli@unsl.edu.ar

Resumen Este artículo describe el desarrollo de un prototipo para la atención de eventos de emergencia y la administración de los móviles correspondientes del Municipio de la localidad de Córdoba, Argentina. Para la solución de esta problemática, se hizo uso de un modelo de base de datos espacio-temporal. En este artículo presentamos las características de su implementación, el uso de la base de datos espacio temporal, las consultas logradas, y las funcionalidades agregadas a la aplicación.

Palabras Claves: I+3 R-Tree, bases de datos espacio-temporales, gestión, emergencias, móviles

1. Introducción

Las bases de datos tradicionales, basadas en el modelo relacional, permiten la gestión eficiente de información que puede estructurarse en registros, donde cada uno posee campos totalmente comparables. Algunas aplicaciones requieren gestionar datos que no puede estructurarse, por lo que los modelos tradicionales de bases de datos no son aplicables, como por ejemplo objetos que se caracterizan por mantener dos atributos comunes: el espacio y el tiempo. Es decir que un objeto puede identificarse por la posición y el área que ocupa en cualquier instante de tiempo. Estos objetos dan origen al tipo de dato espacio-temporal y como soporte para el tratamiento de este tipo de datos surgen los Métodos de Acceso Espacio-Temporales (MAETs), cuyo principal objetivo es evitar que se deba realizar un examen exhaustivo de la base de datos al momento de responder consultas, además de brindar eficiencia ante consultas que se realizan con

alta frecuencia. En este artículo se presenta un prototipo de una aplicación que implementa el índice espacio-temporal I+3 R-Tree para objetos en movimiento, en particular para gestión de eventos de emergencias y móviles asociados para su atención. El trabajo se organiza de la siguiente manera: en la Sección 2 se expone una reseña de aspectos teóricos, definiendo los conceptos necesarios para la comprensión de este trabajo, en la Sección 3 se presenta la aplicación desarrollada y en la Sección 4 las Conclusiones.

2. Aspectos teóricos

En ocasiones resulta de interés poder modelar objetos que tienen una componente espacial asociada, por lo que se requiere el uso del modelo de bases de datos espaciales, que soporta tanto tipos de datos espaciales, como métodos de acceso para poder realizar consultas de manera eficiente. Es también frecuente que además de conocer el estado actual de un modelo de datos, resulte necesario consultar estados pasados. En este contexto cobra relevancia el modelo de bases de datos temporales, que soporta consultas de estados actuales y pasados. Para aquellos casos en los que es necesario gestionar objetos espaciales que pueden cambiar su posición y/o forma a lo largo del tiempo [1] se ha propuesto el Modelo de Bases de Datos Espacio-Temporales. Su objetivo es permitir modelar el dinamismo de los fenómenos que ocurren en el mundo real, administrando al mismo tiempo las componentes espacial y temporal de un objeto, mediante la implementación de tipos de datos apropiados, un lenguaje de consulta, métodos de acceso y modelos que permitan establecer planes óptimos de ejecución de las consultas.

Las principales consultas que se pueden realizar sobre una base de datos espacio-temporal, que incorporan predicados espaciales que involucran instantes o intervalos de tiempo son [2,3]:

- *TimeSlice*: devuelve los objetos que se encuentran en una determinada área en un instante de tiempo dado.
- *Intervalo*: obtiene los objetos que se encuentran en una determinada área en un intervalo de tiempo.
- *Eventos*: se obtienen todos los eventos que sucedieron en una región en un instante dado.
- *Trayectoria*: recupera las posiciones espaciales en las que un objeto ha permanecido en un intervalo de tiempo dado.
- *Consultas de los k vecinos más cercanos*: dado un objeto, se devuelven los k objetos más cercanos al objeto de entrada para un instante o intervalo de tiempo.

Por las características dinámicas de los objetos que se administran en estas aplicaciones, las bases de datos espacio-temporales deben ser capaces de almacenar grandes volúmenes de información a lo largo del tiempo, por lo que resulta imposible pensar en un procesamiento secuencial al realizar una consulta. Por

este motivo se propone la construcción de índices utilizando métodos de acceso espacio-temporales adecuados para la recuperación eficiente de datos.

Existe un método base de indexación espacial que implementan la mayoría de los Sistemas Gestores de Bases de Datos Espacio-Temporales (SGBDET) como Oracle, SQL y PostgreSQL que es el R-Tree, propuesto por Guttman en 1984 [4]. Dentro de las variantes del R-Tree se encuentran: métodos que consideran el tiempo como otra dimensión como 3D R-Tree [5], 2+3 R-Tree [6] y el I+3 R-Tree [7]; métodos en los que el tiempo se incorpora como información dentro de la estructura de los nodos, como el RT-Tree [8]; y métodos que usan múltiples versiones de la misma estructura para distintos instantes de tiempo, con reutilización de las partes invariantes, como el HR-Tree [6,9] y el MR-Tree [8].

Para este caso de estudio se propone la utilización del índice I+3 R-Tree que permite manipular puntos en movimiento, administra el tiempo como otra dimensión y responde a las consultas timeslice, intervalo, evento y trayectoria. Es una variante del 2+3 R-Tree en la que se reemplaza el 2D R-Tree de la estructura original, por un Índice.

La estructura del I+3 R-Tree (Figura 1) está formada por:

- Un R-Tree de tres dimensiones (3D R-Tree) que se utiliza para almacenar los cubos cerrados que representan las posiciones anteriores de los objetos y su tiempo de estadía en dicha posición. Cuando un objeto cambia su posición actual, se actualiza su posición en el índice, y se inserta en el 3D R-Tree un cubo que representa su posición anterior.
- Una estructura denominada I (índice) en la que se almacenan los cubos abiertos, es decir, aquellos para los cuales su instante final en una posición aún no está definida. También se guardan las referencias necesarias a los cubos anteriores que describen la trayectoria del objeto.

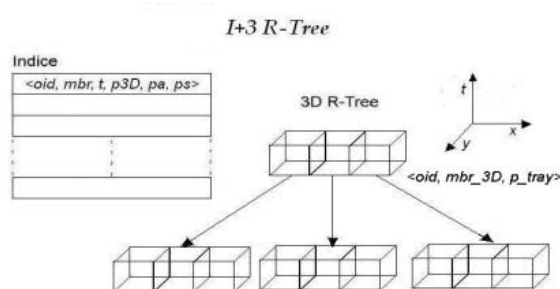


Fig. 1: I+3 R-Tree

Figura 1. Estructura del I+3 R-Tree

3. Herramienta para gestión de eventos de emergencia

En este trabajo se presenta el desarrollo de una aplicación para la atención de eventos de emergencia y la administración de los móviles correspondientes del Municipio de la localidad de Córdoba, Argentina.

La aplicación permite el registro de llamadas telefónicas realizadas a la línea 911. Cada llamada puede generar más de un evento si requiere la intervención de distintas áreas. Se han definido dos áreas: Emergencia Médica y Emergencia Urbana, con sus correspondientes despachos, que son las oficinas encargadas de la asignación de los móviles y del seguimiento del evento hasta su finalización. Al recibirse una llamada se generan el o los eventos asociados, o también puede asociarse a un evento ya existente, ya que pueden recibirse varias llamadas por el mismo hecho. (Figura 2).

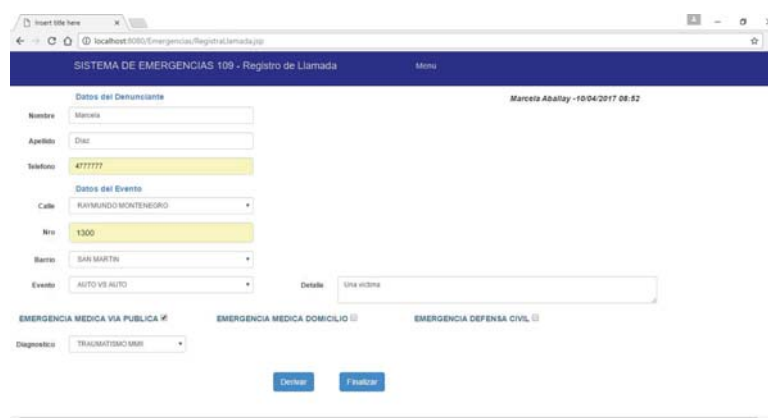


Figura 2. Pantalla para el Registro de Llamadas

El evento puede ser de uno de estos tipos:

1. **Emergencia Médica Domiciliaria:** se requiere una ambulancia para la atención de un paciente en su domicilio particular o en alguna institución pública dependiente de la Municipalidad (ej: escuelas, oficinas públicas). En este caso se registra el tipo de domicilio donde se requiere la atención.
2. **Emergencia Médica en la Vía Pública:** se solicita una ambulancia para la atención de personas en la vía pública.
3. **Emergencia de Defensa Civil:** eventos que requieren la intervención de una cuadrilla para algún trabajo de emergencia pero que no involucra a personas. Por ejemplo: árbol por caer, hundimiento de un pozo, panal de abejas, cable cortado, derrame de líquidos contaminantes, etc.

En el despacho se cuenta con una vista de los eventos pendientes y en curso con su correspondiente localización en el mapa (Figura 3).

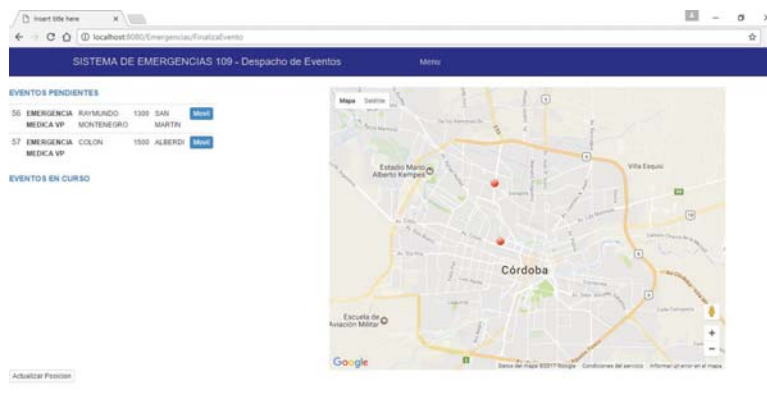


Figura 3. Vista del despacho con los eventos pendientes

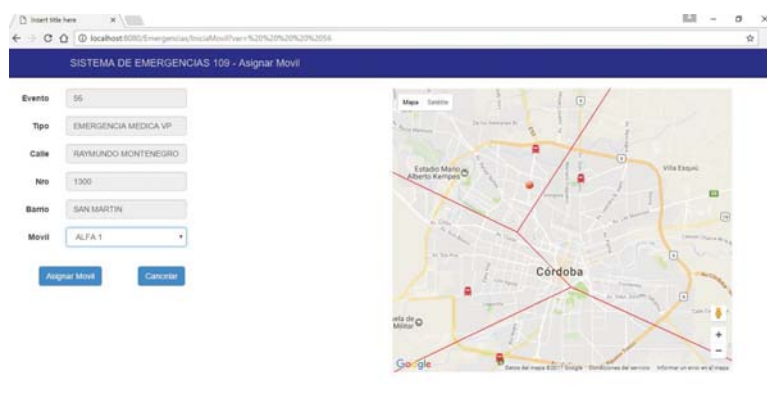


Figura 4. Vista del despacho con los móviles disponibles

Para la asignación de un móvil a un evento se obtiene de la tabla *Móviles* de la base de datos relacional aquellos que están en estado *Disponible* y se los ubica sobre el mapa, formando con ellos un diagrama de Voronoi (Figura 4). De esta forma el operador pueden ubicar rápidamente cual es el móvil más cercano al evento que se está gestionando.

Luego de asignado el móvil, éste se incorpora a la vista del despacho y se actualizan los eventos pendientes y en curso (Figura 5). Una vez finalizada la atención se cierra el evento y se libera el móvil, pasando a estar disponible para cubrir un nuevo evento. Periódicamente se registra la posición de los móviles, lo que permite posteriormente reconstruir su trayectoria en la atención de un determinado evento.

En este proyecto se utilizaron dos modelos de bases de datos para almacenar distintos tipos de datos: una base de datos relacional (Figura 6), en la que se

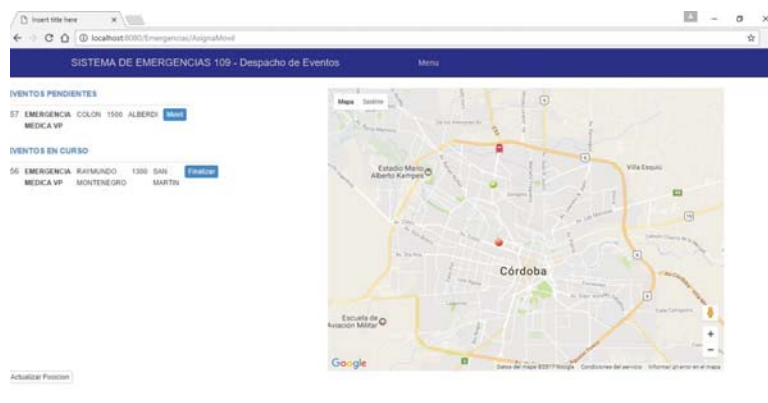


Figura 5. Vista del despacho con los eventos pendientes, en curso y móviles asignados

registra información estructurada, y una base de datos espacio-temporal (BDET) (con la implementación del índice I+3 R-Tree para realizar consultas de manera eficiente), en la que se registra la información relacionada con la ubicación y tiempos asociados al dinamismo de los móviles.

Respecto a la BDET, en esta aplicación se implementaron los siguientes tipos de consultas:

- Consulta de Trayectoria: se requiere conocer la trayectoria que realizó un móvil en la atención de un evento, desde que sale de su base, hasta el lugar del evento y luego al destino correspondiente (hospital para los casos en que se registran personas heridas).
- Consulta Instantánea: dado un evento, se puede determinar su área de influencia, y conocer todos los móviles disponibles que pueden ser asignados para atender dicho evento.
- Consulta por Intervalo: permite obtener por ejemplo todos los móviles asignados a la atención de un evento, conociendo su posición en un intervalo de tiempo.
- Consulta de Historial de Eventos: permite conocer los eventos que ocurrieron en una región en un momento determinado.
- Consulta de los k vecinos más cercanos: dado un móvil y conocido el lugar de un evento, se puede obtener el listado de los móviles más cercanos al mismo para un instante o intervalo de tiempo, para aquellos casos en los que se requiere presencia de más de un móvil para atender el evento.

El desarrollo de la aplicación fue hecho en Java, se utilizó como Sistema Gestor de Bases de Datos Relacional PostgreSQL y se implementó el método de acceso espaciotemporal I+3 R-Tree a través de sus librerías en Java.

El I+3 R-Tree está conformado por dos estructuras un 3D R-Tree y un Índice. Las tuplas en el 3D RTree están compuestas por $\langle oid, mbr_3D, p_tray \rangle$

- *oid*: código identificador del objeto.

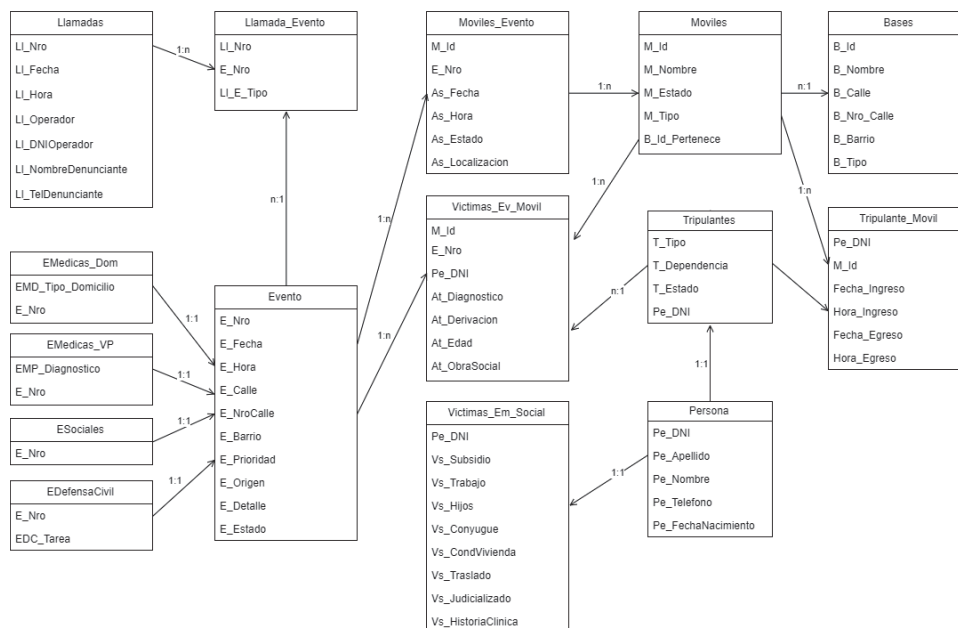


Figura 6. Base de Datos Relacional

- *mbr_3D*: región tridimensional cuya altura representa el intervalo temporal durante el cual el objeto se mantuvo en la posición espacial definida por su base.
- *p_tray*: puntero al cubo anterior correspondiente al mismo *oid*, utilizado para mantener un historial de trayectoria.

Las tuplas en el índice son de la forma $\langle oid, mbr, t, p3D, pa, ps \rangle$

- *mbr*: región aproximada que ocupa actualmente el objeto.
- *t*: tiempo de llegada del objeto a su ubicación actual.
- *p3D*: puntero al cubo anterior correspondiente al mismo *oid*, utilizado para mantener un historial de trayectoria.
- *pa*: puntero al objeto insertado en el instante de tiempo inmediatamente anterior.
- *ps*: puntero al objeto insertado en el instante de tiempo siguiente.

Inicialmente en el I+3 R-Tree se registran en el Índice los móviles con su posición inicial (el 3D R-Tree vacío) (Tabla 1).

Al recibir información de una nueva posición del móvil *om1* se actualiza el Índice y el 3D R-Tree (Tablas 2 y 3).

Tabla 1. Índice

Oid	Mbr	t	p3D	Pa	Ps
om1	om1_posicion0	om1_tiempo0	null	null	puntero a om2
om2	om2_posicion0	om2_tiempo0	null	puntero a om1	puntero a om3
om3

Tabla 2. Índice

Oid	Mbr	t	p3D	Pa	Ps
om1	om1_posicion1	om1_tiempo1	puntero al primer cubo del R-Tree	null	puntero a om2
om2	om2_posicion0 (la de la base)	om2_tiempo0	null	puntero a om1	puntero a om3
om3

Tabla 3. 3D R-Tree

Oid	mbr_3D	p_tray
om1	posicion0 - Desde tiempo0 hasta tiempo1	null

Para la visualización y georreferenciación de los móviles sobre el mapa de la ciudad de Córdoba se utilizaron las APIs de Google Maps [10]. De esta manera se pueden localizar en el mapa de la ciudad los puntos donde se encuentran los eventos en curso y pendientes de atención. En el caso del despacho, en la pantalla correspondiente a la asignación de un móvil para un evento, se implementó un diagrama de Voronoi. Esta estructura geométrica subdivide un área en regiones tomando como base una serie de puntos. Dichas regiones quedan formadas por los puntos más próximos a cada uno de los puntos de base [11]. Para este caso el área a dividir es el mapa de la ciudad de Córdoba y los puntos bases son los móviles disponibles. De esta forma se pueden identificar las distintas áreas de cobertura de cada móvil. Como en el prototipo no se dispone de una conexión real con los móviles para que transmitan cada determinado período su posición, se realizó una simulación, registrando en el I+3 R-Tree distintos tiempos y posiciones (latitud, longitud) para algunos móviles. Luego de finalizada la atención, en base a los datos registrados en la estructura espacio-temporal se puede resolver la consulta de trayectoria de un móvil, que es uno de los principales requerimientos de información para este sistema.

4. Conclusiones

A partir de los requerimientos funcionales planteados por el Municipio de Córdoba para la gestión de móviles que se asignan en la atención de emergencias, y debido a que ningún Sistema Gestor de Bases de Datos ofrece actualmente una gestión eficiente de datos tradicionales combinados con datos que cambian su forma y/o posición a través del tiempo, se desarrolló este prototipo de aplicación que combina una Base de Datos Relacional con una Base de Datos Espacio-Temporal. Teniendo en cuenta las características de los móviles que se pretende gestionar con esta aplicación se consideró el índice espacio-temporal I+3 R-Tree como el más conveniente de implementar ya que permite administrar objetos con las siguientes características:

- Los objetos se mueven a gran velocidad.
- El tamaño y la forma de los objetos no son importantes, el interés se centra en su posición en el tiempo.
- El espacio no cambia y la cantidad de objetos no tiene mucha variación.
- Los objetos se mueven en áreas previamente establecidas.
- Se necesita consultar la trayectoria realizada por los objetos.

La consulta de trayectoria resulta fundamental en esta aplicación y pudo resolverse eficientemente con la implementación del índice seleccionado, ya que la misma no está disponible en todos los índices de acceso espacio-temporal. La combinación del motor de Bases de Datos Relacional con el índice métrico I+3 R-tree para tratar objetos en movimiento ha resultado muy satisfactorio, y los resultados de las primeras pruebas son alentadores, dado que el prototipo ha respondido de manera satisfactoria a las consultas efectuadas, devolviendo adecuadamente los móviles disponibles en los radios más cercanos al lugar del evento.

Referencias

1. Y. Theodoridis, T. Sellis, A.N. Papadopoulos, and Y. Manolopoulos. Specifications for efficient indexing in spatiotemporal databases. In *Proceedings. Tenth International Conference on Scientific and Statistical Database Management (Cat. No.98TB100243)*, pages 123–132. IEEE Comput. Soc, 1998.
2. Pankaj K. Agarwal, Lars Arge, and Jeff Erickson. Indexing moving points (extended abstract). In *Proceedings of the Nineteenth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, PODS '00*, pages 175–186, New York, NY, USA, 2000. ACM.
3. Gilberto Antonio Gutierrez Retamal. *Metodos de Acceso y Procesamiento de Consultas Espacio-Temporales*. PhD thesis, 2007.
4. Antonin Guttman. R-trees: A dynamic index structure for spatial searching. *SIGMOD Rec.*, 14(2):47–57, June 1984.
5. Y. Theodoridis, M. Vazirgiannis, and T. Sellis. Spatio-Temporal indexing for large multimedia applications. *Multimedia Computing and Systems, International Conference on*, 0:0441+, 1996.

6. Mario A. Nascimento, Jefferson R. O. Silva, and Yannis Theodoridis. Evaluation of Access Structures for Discretely Moving Points. *Proceedings of the International Workshop on Spatio-Temporal Database Management*, pages 171–188, 1999.
7. Edilma Olinda Gagliardi, Fernando D. Carrasco, and Juan C. García Sosa. I+3 R-Tree: un método de acceso espacio-temporal. 2009.
8. X Xu, J Han, and W Lu. RT-tree: An improved R-tree index structure for spatio-temporal database. *4th International Symposium on Spatial Data Handling*, pages 1040—1049, 1990.
9. Mario A Nascimento, Jefferson R O Silva, Yannis Theodoridis, Christian S Jensen, Michael H Böhlen, Renato Busatto, Curtis E Dyreson, Heidi Gregersen, Dieter Pfoser, Simonaš Saltenis, Janne Skyt, Giedrius Slivinskas, Kristian Torp, Richard T Snodgrass, Sudha Ram, Michael D Soo, Andreas Steiner, and Switzerland Timeconsult. Access Structures for Moving Points. 1998.
10. Api google maps. <http://goo.gl/QH7y4K>.
11. Mark de Berg, Otfried Cheong, Marc van Kreveld, and Mark Overmars. *Computational Geometry: Algorithms and Applications*. Springer-Verlag TELOS, Santa Clara, CA, USA, 3rd ed. edition, 2008.

XIV Workshop Ingeniería de Software (WIS)

Coordinadores

Pesado, Patricia – Estévez, Elsa – Cechich, Alejandra – Kuna,
Horacio

Comité Científico

Astudillo, Hernán (Chile)	Olivas Varela, José Ángel (España)
Barbosa, Luis (Portugal)	Piatini, Mario (España)
Dujmovic, Jozo (EEUU)	Riesco, Daniel (Argentina)
Fillottrani, Pablo (Argentina)	Rossi, Gustavo (Argentina)
Genero, Marcela (España)	Ruiz, Francisco (España)
Janowski, Tomasz (Polonia)	Sousa Pinto, Jorge (Portugal)
Navarro Martin, Antonio (España)	

Análisis de estrategias para clasificar contenidos en foros de discusión

Valeria Zoratto, Nadina Martinez Carod, Facundo Otermin, Gabriela Aranda

{vzoratto|nadina.martinez}@fi.uncoma.edu.ar

Abstract. La información contenida en los foros de discusión de la Web es, en la mayoría de los casos, considerada muy valiosa por usuarios con problemas o necesidades similares. Por ese motivo, en el último tiempo se han multiplicado los esfuerzos para recuperar, analizar y reutilizar la información que se extrae de dichos hilos de discusión. En trabajos previos se ha presentado una estrategia de análisis de hilos de discusión basada en la herramienta de búsqueda Lucene. Siguiendo esa línea, en este artículo se presenta una extensión de la metodología anterior en la que se combina la funcionalidad de la base de datos léxica WordNet y el parser de lenguaje natural Stanford, con el objetivo de evaluar la inclusión de sinónimos considerando la estructura gramatical de los hilos bajo estudio.

1 Introducción

Con la evolución de la Web, han aparecido distintas plataformas de trabajo colaborativo que permiten que los usuarios se comuniquen y trabajen en forma conjunta sin importar que estén reunidos en un mismo lugar físico ni que lo hagan en el mismo instante. Estas plataformas tienen como objetivo general, además de compartir información en distintos formatos, permitir el intercambio de opiniones y conocimiento. Ejemplos de estas plataformas, ampliamente utilizadas en la actualidad, son las Wikis, los Weblogs y los foros de discusión. Estos últimos son especialmente interesantes dado que habitualmente son utilizados para, ante una dificultad, solicitar ayuda a usuarios más expertos en dominios específicos [1]. En particular, los foros de discusión sobre temáticas relacionadas al desarrollo y mantenimiento de software, tienen un gran volumen de contenido útil, producido por sus usuarios, que es deseable y valioso que pueda ser extraído y reutilizado [2]. Además, los foros tienen una estructura única [3], constituida por un conjunto de hilos, donde cada uno está compuesto por un título, una pregunta principal, y luego una serie de respuestas que representan el debate sobre dicha pregunta.

Usualmente, ante un problema técnico, los usuarios utilizan los motores de búsqueda multi-propósito para acceder a las discusiones de los foros, lo que los lleva a recorrer varias páginas hasta encontrar un problema similar al suyo entre diversos hilos de uno o de varios foros. Sin embargo, a veces la solución encontrada no es la adecuada para dicho problema y es necesario probar varias posibles soluciones hasta encontrar la correcta, transformándose en una tarea que insume gran cantidad de tiempo.

Con el objetivo de minimizar el trabajo manual, en [4] se presentaron las características de una herramienta para facilitar la búsqueda de soluciones a problemas comunes, estableciendo prioridades entre las soluciones disponibles en foros de discusión dispersos en la Web. Una componente de la herramienta propuesta, desarrollada en [5], evalúa la información obtenida de los foros a partir del análisis léxico de sus datos. A fin de mejorar los resultados obtenidos hasta el momento, se propone una extensión de dicha herramienta que permite evaluar la inclusión de sinónimos en la clasificación de los hilos de discusión, empleando WordNet (base de datos léxica ampliamente utilizada para el idioma inglés) [6], y Stanford Parser¹, para analizar la estructura gramatical del hilo.

El proceso que se busca mejorar está representado en la Figura 1, donde se muestra el desarrollo de la clasificación de hilos de acuerdo a un conjunto de entidades reconocibles, llamadas *documentos de referencia*. Como este ejemplo se basa en el lenguaje de programación Java, los documentos de referencia forman parte del repositorio de especificación de clases de Oracle (versión 5)², sin embargo, dicho conjunto de documentos puede variar al enfocar el estudio en otras temáticas.

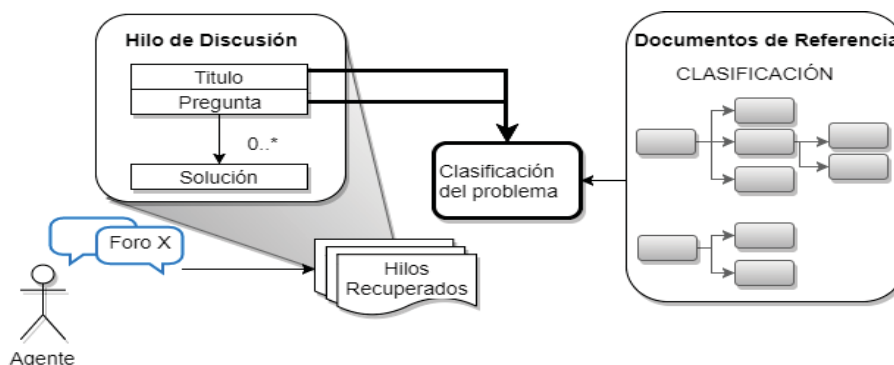


Fig. 1. Proceso de clasificación de hilos de acuerdo a documentos de referencia

Teniendo en cuenta este objetivo, en la Sección 2 se detallan los antecedentes que llevaron a la realización del primer caso de estudio. Posteriormente, en la Sección 3 se detalla la modificación de la herramienta propuesta en [5] y en la Sección 4 se evalúan los resultados alcanzados con dicha modificación. Luego, en la Sección 5 se comparan los resultados originales con los obtenidos en la versión modificada. Finalmente, en la Sección 6 se presentan las conclusiones y trabajo futuro.

¹

<https://nlp.stanford.edu/software/tagger.shtml>

² <http://docs.oracle.com/javase/1.5.0/docs/>

2 Antecedentes

En un foro de discusión, cada hilo está compuesto por un título y una pregunta principal, y luego se encadenan una secuencia de respuestas que representan el debate sobre dicha pregunta. Por el otro lado, la estructura de un documento Oracle tiene el nombre de la clase (*Class Name* en el documento Oracle), una lista de las interfaces que implementa (*Implemented Interfaces*), las subclases conocidas (*Known Subclasses*), y luego se describen los métodos de la clase con sus parámetros, comentarios y en algunos casos ejemplos de utilización. Estas estructuras se muestran esquematizadas en la Figura 2.

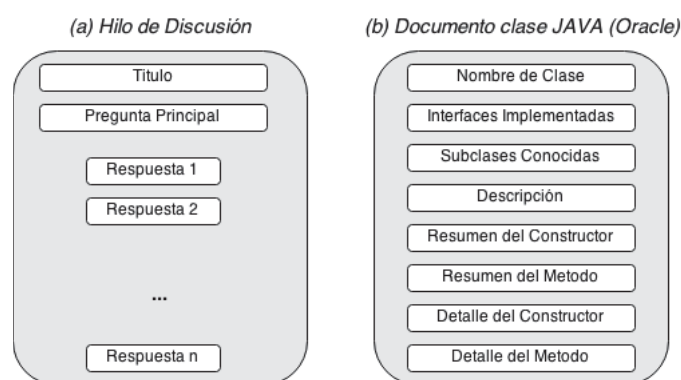


Fig. 2. Estructuras de los documentos estudiados

En base a la información contenida en las secciones presentes en cada tipo de documento (hilos de discusión y documentos Oracle), y a la tarea objeto de esta investigación de clasificar los hilos de foros de discusión de la manera más apropiada, se definieron las siguientes hipótesis:

- HIPÓTESIS A: Utilizar mayor cantidad de información sobre cada una de las clases Java documentadas en Oracle permite clasificar los hilos de discusión relacionados a ellas de forma más precisa.
- HIPÓTESIS B: Utilizar más información sobre el problema explicado en los hilos de discusión permite clasificarlos de forma más precisa respecto a los documentos Oracle de las clases Java.

A efectos de evaluar dichas hipótesis, en [5] se propuso una metodología de clasificación de hilos reales de discusión técnicos. Para evaluar las hipótesis propuestas, se establecieron las siguientes fases:

- Fase 1. Recuperación de documentos: Recuperar los hilos de discusión del foro de discusión y las especificaciones de clases Java del repositorio de Oracle (documentaos de referencia).
- Fase 2. Clasificación por expertos: Los hilos de discusión recuperados en la fase 1 son analizados por tres expertos para identificar las clases Java más relacionadas con cada uno.

- Fase 3. Pre-procesamiento de documentos: Preparar los documentos descargados para su posterior análisis, eliminando código html irrelevante e información innecesaria,
- Fase 4. Indexación de documentos de referencia: Utilizar la API Lucene [7] para indexar los documentos de referencia.
- Fase 5. Búsqueda de documentos relevantes: Utilizar la API Lucene para determinar la relación entre cada hilo de discusión y los documentos Oracle de las clases Java.
- Fase 6: Evaluación: Contrastar los resultados de los expertos con los retornados por la herramienta.

Dado que la herramienta Lucene permite redefinir el conjunto de palabras *stop-words*³, se modificó el conjunto predefinido para que no se consideren como tal las palabras reservadas del lenguaje Java (*for*, *then*, *if*, *this*) y se agregaron otras que han sido consideradas poco representativas en dicho contexto.

Esta metodología fue aplicada sobre un conjunto de hilos de discusión referidos a problemas sobre el uso del lenguaje Java. En particular, se recuperaron 50 hilos del foro de discusión Stack Overflow⁴, utilizando la funcionalidad de filtro por tags de dicho sitio, seleccionando el tag *java*. Uno de los problemas detectados a partir de la ejecución de un caso de estudio fue que algunos documentos Oracle aparecían relacionados a la mayoría de los hilos de discusión con un valor alto, y a partir del análisis correspondiente se detectó que se trataba de nombres de clases que representaban vocablos de uso común en el lenguaje natural dentro del ambiente de programación, como por ejemplo: *Class*, *Error*, *Type*, etc., por lo que se reprodujo la fase de indexación, pero en este caso eliminando dichos documentos del conjunto de referencia.

En [8] se presentó un análisis de dos casos de estudio siguiendo la metodología mencionada, en la que se observan los resultados obtenidos diferenciando, por un lado la cantidad y calidad de información obtenida de los hilos (casos F_1 , F_2 , F_3) y luego considerando diferentes porcentajes de información de la documentación extraída de Oracle (casos O_a , O_b , O_c). Los resultados obtenidos se grafican en la Figura 3 donde se puede observar que en el caso (a), al contar con mayor información de los documentos Oracle (caso O_c) la performance es mas baja que utilizando los documentos que contienen menos información (casos O_a y O_b). En el caso (b), la performance es mayor cuando se consideran los documentos que contienen el título del hilo y la pregunta principal (caso F_2), pero al utilizar el texto completo del hilo (caso F_3) mejora los resultados en comparación de las ocurrencias donde sólo se considera el título del mismo (caso F_1).

A partir de los resultados obtenidos en ese caso de estudio, surgió la inquietud de modificar el proceso incorporando sinónimos al texto recuperado de los hilos de discusión. Enfocado en la hipótesis B que sugiere que la precisión aumenta a medida que se cuenta con más información sobre el problema, se plantea la siguiente hipótesis, que será desarrollada a continuación:

³ Palabras que carecen de significado por sí solas y no brindan información acerca del contenido del texto

⁴ <http://stackoverflow.com/>

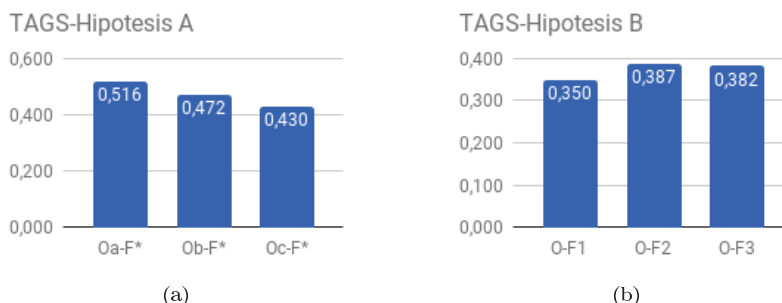


Fig. 3. Análisis de performance según las hipótesis planteadas

- HIPÓTESIS B.2: Agregar sinónimos al problema explicado en los hilos de discusión permite clasificarlos de forma más precisa respecto a los documentos Oracle de las clases Java.

3 Mejora en el procesamiento de hilos

La incorporación de sinónimos para mejorar los resultados obtenidos al clasificar los hilos de discusión, requiere que se extienda la etapa de procesamiento previa a la fase 5. Surge así una modificación al proceso propuesto en [8], tal como se presenta en la Figura 4. En dicha modificación, se agrega una nueva fase, entre las fases 3 y 5 (explicadas en la Sección 2), para implementar un segundo procesamiento de los hilos obtenidos en la fase 3, que los enriquezca incorporando sinónimos, mediante la utilización de WordNet y del parser Stanford.

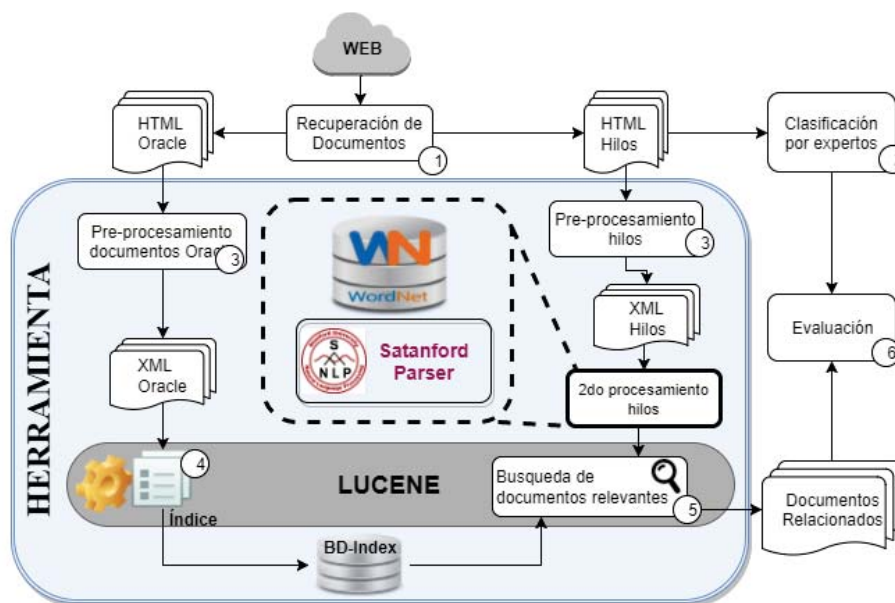


Fig. 4. Fases del desarrollo utilizando WordNet y Stanford parser

3.1 Procesamiento de los hilos para agregar sinónimos

En esta fase se utilizan como entrada los documentos pre-procesados durante la Fase 3 (*pre-procesamiento hilos*), donde los documentos *html* se traducen a formato *xml*, descartando código *html* irrelevante y agregando etiquetas para delimitar los bloques de interés. Además, en la Fase 3 se crean tres versiones de cada documento, tanto para los hilos como para los documentos Oracle, considerando un subconjunto de las partes que componen cada documento, como se presenta en la Tabla 1.

Table 1. Versiones de los diferentes hilos y documentos de referencia

Hilos		Documentos de referencia	
F_1	Sólo el título del hilo	O_a	Sólo el nombre de la clase
F_2	Título del hilo y la pregunta principal	O_b	Nombre de la clase y los nombres de todos sus métodos
F_3	Texto completo del hilo (título, pregunta principal, respuestas)	O_c	Todas las secciones del documento (excepto las secciones "Detalles de constructor" y "Detalle de métodos").

En base a las tres versiones de cada tipo de documento, se establecieron nueve combinaciones que formaron la base del análisis, tal como se muestra en la Tabla 2.

Table 2. Combinaciones de los tipos de documentos a analizar

	Sólo título (F_1)	Título + pregunta (F_2)	Texto completo (F_3)
Sólo el nombre de la clase (O_a)	$O_a F_1$	$O_a F_2$	$O_a F_3$
Nombre de la clase y métodos (O_b)	$O_b F_1$	$O_b F_2$	$O_b F_3$
Texto completo (O_c)	$O_c F_1$	$O_c F_2$	$O_c F_3$

Para considerar los sinónimos en la búsqueda de documentos de referencia relacionados, se utiliza el Stanford Parser para etiquetar las palabras según la gramática de la oración, teniendo en cuenta el contexto en el que se utiliza cada palabra. De esta manera, la palabra se clasifica como verbo, adjetivo, adverbio, sustantivo, u otro. El parser detecta qué grupos de palabras van juntas (como "frases") y qué palabras son el sujeto u objeto de un verbo. Por ejemplo, para la frase "*My dog also likes eating sausage.*", el parser etiqueta la cadena a "*My/PRP\$ dog/NN also/RB likes/VBZ eating/VBG sausage/NN ./.*". De la misma manera que [9], en esta instancia del estudio nos interesamos en el análisis de sustantivos y sus modificadores (adjetivos y adverbios).

Una vez que se cuenta con todas las palabras del hilo etiquetadas según la estructura gramatical, se utiliza WordNet para recuperar los sinónimos (según el tipo correspondiente) y se agregan a la cadena de búsqueda teniendo en cuenta que la palabra no sea un stopword. Por otro lado, se han tomado como caso

especial los valores numéricos que suelen agregar sinónimos innecesarios. Por ejemplo, para el valor 7, Wordnet considera sinónimos los términos *seven*, *VII*, *Sevener*, *heptad*, *septet* y *septenary*, que van más allá del significado del número en sí, por eso se decidió ignorar los sinónimos de los números en esta etapa del estudio.

Una vez que se han obtenido las nuevas versiones de los hilos, se procede a ejecutar la Fase 5 para obtener los documentos relevantes para cada consulta.

4 Evaluación y Análisis de resultados

Dado que el objetivo de los sistemas de recuperación de información es tratar de maximizar la cantidad de documentos recuperados que sean relevantes, se utilizará la medida F-Measure para evaluar la performance del método de recuperación, dado que brinda mayor información que las medidas de precisión y *recall*, y que es una de las medidas más utilizadas al evaluar técnicas de recuperación de información [10].

Un enfoque adicional es calcular estas medidas con valores de cortes (*cut-off*) sobre la cantidad de respuestas válidas obtenidas por el recuperador de información utilizado. Es decir, en el caso de corte N se analizan la cantidad de documentos relevantes, no relevantes, precisión, recall y F-Measure que se obtienen al considerar solamente los primeros N documentos adquiridos en el proceso de recuperación.

Es importante evaluar los resultados estimando valores de cortes sobre la cantidad de respuestas validas, es por ello que, teniendo en cuenta lo realizado en [5,8], se utilizaron las mismas medidas de corte, conservando como corte inicial $N=2$ y siguiendo una escala de $N=3, 4$ y 5 .

Conservando las combinaciones planteadas anteriormente (Sección 3.1) se obtuvieron 9 posibilidades sin utilizar las clases Oracle consideradas stopwords. Las mediciones para los valores de corte mencionados anteriormente, utilizando la medida F-Measure, se presentan en la Tabla 3.

En la Figura 5 se muestran las tendencias de la performance en este caso de estudio. Se puede observar que al considerar el título y la pregunta principal del hilo (F_2) mejora la performance de la clasificación respecto a considerar sólo el título (F_1) o el hilo completo (F_3). En los casos en que la búsqueda se orienta a problemas de Java en general, cuando sólo se utiliza el título en la clasificación, la baja performance puede deberse a que el nombre de las clases relacionadas no suele aparecer siempre en el título de los mismos; aunque sí suele aparecer al explicar el problema en la pregunta principal. En cuanto a la búsqueda considerando el hilo completo, la baja performance puede deberse a que las discusiones son largas y se pierde precisión al aumentar el contenido de las mismas (dan ejemplos o piden aclaraciones pero no aportan a la solución).

5 Análisis comparativo de los resultados

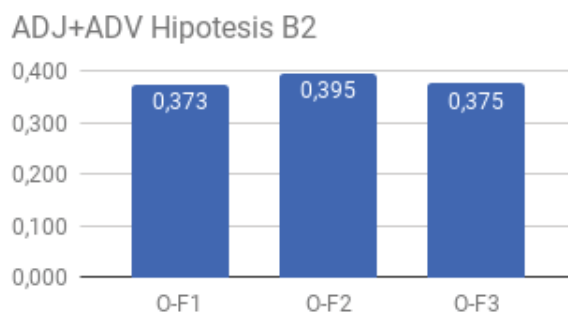
En la Figura 6 se contrastan los resultados agrupados según la cantidad de información en los hilos de discusión, considerando el agregado de sinónimos según lo establecido en la hipótesis B.2.

Table 3. F-Measure para cada corte de las pruebas realizadas

	2	3	4	5
O_aF_1	0.509	0.522	0.522	0.522
O_aF_2	0.573	0.581	0.568	0.572
O_aF_3	0.573	0.581	0.568	0.567
O_bF_1	0.4	0.39	0.402	0.401
O_bF_2	0.494	0.456	0.429	0.397
O_bF_3	0.433	0.408	0.394	0.367
O_cF_1	0.218	0.212	0.199	0.176
O_cF_2	0.167	0.183	0.166	0.54
O_cF_3	0.15	0.162	0.147	0.154

Al realizar una comparativa entre el primer caso de estudio (sin incorporar sinónimos) y el segundo caso de estudio (incorporando sinónimos de adjetivos y adverbios), se observa que utilizar sinónimos mejora la performance respecto a considerar el texto original de los hilos cuando se tiene en cuenta el título y la pregunta principal (caso F_2), pero contrariamente a lo planteado, al utilizar el hilo completo al agregar sinónimos disminuye la performance (caso F_3). Esto puede deberse a que agregar todos los sinónimos de adjetivos y adverbios podría ocasionar ruido ya que los cálculos que realiza Lucene se realizan sobre el total de palabras de la consulta, lo cual no parece verse afectado en los casos F_2 y F_1 donde la cantidad de palabras es considerablemente más baja que en el hilo completo. Por otro lado, la incorporación de sinónimos al considerar título y pregunta (caso F_2) principal consigue la mejor performance de todas las combinaciones posibles, lo que apoyaría la hipótesis B.2 planteada.

Respecto a las amenazas a la validez del experimento, se debe mencionar que, los documentos Oracle utilizados para clasificar los hilos en estos casos de

**Fig. 5.** Análisis de performance agrupado según cantidad de información en hilos de discusión (Hipótesis B2)

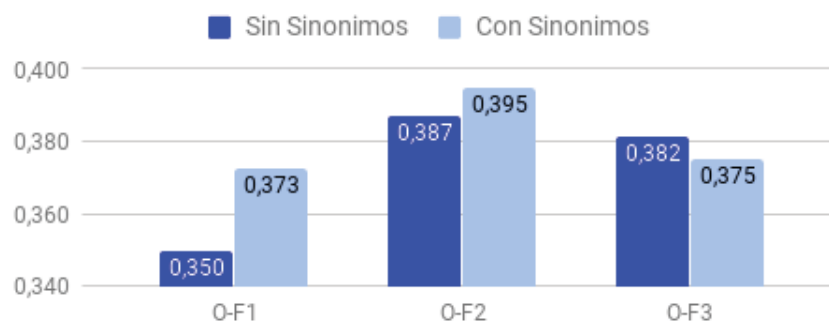


Fig. 6. Comparación de la performance al utilizar sinonimos de adjetivos y adverbios

estudio se restringieron a la versión Java 1.5, por lo que podría producirse alguna diferencia en los resultados al usar documentos de otra versión del lenguaje (por ejemplo por la inclusión o eliminación de un método de una versión a otra). Además, nuevas pruebas deben realizarse para evaluar estas hipótesis con un corpus de hilos mayor, recuperados de distintos foros.

6 Conclusiones y trabajo futuro

En este trabajo se presenta una estrategia para clasificar hilos de discusión recuperados de un foro de discusión técnico respecto a un conjunto de documentos de referencia, considerando la cantidad de información tomada de cada tipo de documento y la inclusión de sinónimos para lograr una mejor performance. Particularmente, en este artículo se presenta un caso de estudio con hilos recuperados del foro Stack Overflow, respecto al conjunto de documentos de especificaciones Oracle de las clases Java. La clasificación se realizó considerando las distintas secciones de ambos tipos de documentos y el agregado de sinónimos de los modificadores de sustantivos (adjetivos y adverbios) en los hilos de discusión.

De acuerdo a los resultados obtenidos con un corpus de 50 hilos de discusión, la clasificación tiene su mejor performance considerando el título y la pregunta principal del hilo e incluyendo sinónimos de los modificadores de sustantivos (adjetivos y adverbios).

Dado que este resultado proviene de un conjunto de hilos restringido, nuestro trabajo a futuro se enfocará en replicar estos experimentos con mayor cantidad de hilos así como con otras técnicas durante la fase de la recuperación de información, para asegurar la generalidad de estos resultados. Además, se pretende realizar nuevos experimentos controlados teniendo en cuenta las diferentes categorías gramaticales, utilizándolas individualmente y haciendo combinaciones de las mismas.

Posteriormente se plantea realizar resúmenes sobre los hilos de discusión, es decir, tener en cuenta las respuesta mas votadas, utilizar solo las respuestas

más relacionadas a la pregunta del usuario, etc., usando otras técnicas como la Similitud del Coseno, en lugar del hilo completo.

Finalmente, en este trabajo, las medidas utilizadas se aplican sobre un conjunto de documentos relacionados sin tener en cuenta su orden de relevancia. A futuro se planea ampliar el análisis a otros tipos de métricas que consideren el *ranking* de documentos relevantes retornado por Lucene.

Agradecimientos

Este trabajo está parcialmente soportado por el subproyecto “*Reuso de Conocimiento en Foros de Discusión - Parte II*”, correspondiente al Programa de Investigación 04/F009 “*Desarrollo Orientado a Reuso - Parte II*” de la Universidad Nacional del Comahue (Neuquén, Argentina).

References

1. G. Cong, L. Wang, C.-Y. Lin, Y.-I. Song, and Y. Sun, “Finding question-answer pairs from online forums,” in *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '08, (New York, NY, USA), pp. 467–474, ACM, 2008.
2. S. Gottipati, D. Lo, and J. Jiang, “Finding relevant answers in software forums,” in *26th IEEE/ACM International Conference on Automated Software Engineering (ASE 2011)*, Lawrence, KS, USA, November 6-10, 2011, pp. 323–332, 2011.
3. S. Bhatia and P. Mitra, “Adopting inference networks for online thread retrieval,” in *AAAI*, vol. 10, pp. 1300–1305, 2010.
4. Aranda, Gabriela, Martínez Carod, Nadina, Roger, Sandra, Faraci, Pamela, and Cechich, Alejandra, “Una herramienta para el análisis de hilos de discusión técnicos,” in *CACIC 2014, XX Congreso Argentino de Ciencias de la Computación*, (San Justo, Argentina), pp. 803 – 812, Oct. 2014.
5. V. Zoratto, G. N. Aranda, S. Roger, and A. Cechich, “Análisis de estrategias para clasificar contenidos en foros de discusión: Un caso de estudio,” in *Simposio Argentino de Ingeniería de Software (ASSE 2015)-JAIIO 44*, (Rosario), pp. p. 176–190, SADIO, 2015.
6. G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. J. Miller, “Introduction to wordnet: An on-line lexical database,” *International journal of lexicography*, vol. 3, no. 4, pp. 235–244, 1990.
7. C. D. Manning and H. Schütze, *Foundations of Statistical Natural Language Processing*. Cambridge, MA, USA: MIT Press, 1999.
8. V. Zoratto, G. N. Aranda, S. Roger, and A. Cechich, “Analyzing discussion forums threads about java programming language usage,” *Electronic Journal of Informatics and Operations Research*, vol. 15, no. 1, 2016.
9. M. Nicoletti, S. Schiaffino, and D. Godoy, “Mining interests for user profiling in electronic conversations,” *Expert Syst. Appl.*, vol. 40, pp. 638–645, Feb. 2013.
10. R. A. Baeza-Yates and B. Ribeiro-Neto, *Modern Information Retrieval*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1999.

Análisis de métricas de calidad en foros de discusión técnicos

Nadina Martínez Carod, Pamela Faraci, Gabriela Aranda

Grupo GIISCo

Facultad de Informática, Universidad Nacional del Comahue

Buenos Aires 1400 (8300) Neuquén, Argentina

{nadina.martinez|gabriela.aranda}@fi.uncoma.edu.ar,

pamelafaraci@yahoo.com.ar

Resumen Si bien los foros de discusión son consultados asiduamente en la Web para resolver problemas, muchas veces la información contenida en ellos no está bien organizada o no es la adecuada en términos de solución, razón por la cual se vuelve indispensable poder definir algunos criterios de calidad de la información existente en dichos foros. La propuesta es utilizar los foros como sistemas de recupero de información (SRI), con capacidad para almacenar, mantener y recuperar información. La información resultante estará ordenada de acuerdo a alguna prioridad generada considerando varias características de calidad. En esta propuesta se definen algunas métricas para medir la calidad de la información contenida en hilos de discusión técnicos, y se analizan las métricas propuestas para especificar un ranking en la información recuperada, se ejemplifica a través de un caso de estudio.

1. Introducción

Con el transcurso del tiempo la Web se ha convertido en un gran repositorio de datos donde se intercambia conocimiento, ideas y se producen debates. Las herramientas colaborativas son aplicaciones dinámicas que se caracterizan por tener comunidades donde el mayor énfasis se da a la contribución y participación de los usuarios; dentro de las herramientas existentes, los foros de discusión se destacan por ser generalmente utilizados para resolver inconvenientes que pudieran surgir en cualquier momento. No sólo son empleados por los usuarios que pertenecen a la misma comunidad como elemento de debate, sino que muchas veces pueden ser accedidos por usuarios extenos, como fuente de conocimiento. El crecimiento desmedido de la Web trajo como consecuencia algunas situaciones no deseables, en particular la existencia de información similar repetida en diferentes sitios, incluso muchas veces mal organizada. De esta manera, para poder acceder a la información almacenada en los foros de discusión, a menudo es necesario navegar por varios hilos hasta dar con una solución factible o con la solución adecuada, razón por la cual, se consideró la utilización de las características de calidad para evaluar las soluciones encontradas en diferentes foros de discusión [1].

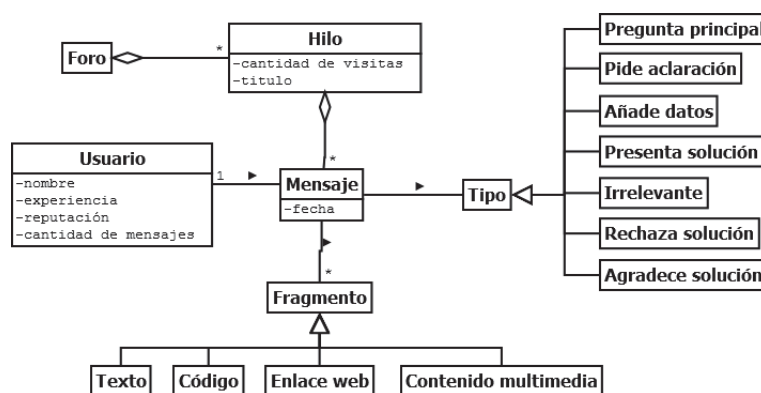


Figura 1: Modelo Conceptual de foros de discusión técnicos

De la revisión formal realizada sobre un grupo de foros de discusión, se desprende la metodología de búsqueda, donde una persona, perteneciente a la comunidad del mismo, realiza una pregunta, la cual llamamos pregunta principal, a partir de la misma se genera un hilo de discusión. Dicho hilo está formado por mensajes, los cuales a su vez están constituidos por fragmentos. Al ser un foro de discusión técnico, los fragmentos además de ser texto, gráficos, sonidos, videos, pueden contener código de lenguajes de programación. La Figura 1 muestra el modelo conceptual rectificado, diferenciado con el modelo presentado en [2], en el cual se identifican las componentes de un foro de discusión así como sus relaciones.

Una actividad típica en la Web es la consulta a través de algún motor de búsqueda multipropósito, la cual trae como resultado una lista de objetos relacionados a dicha consulta, motivo por el cual se desea que la lista resultante contenga hilos de discusión referidos al tema en cuestión. Luego de que una persona realiza una consulta, ésta necesitará visitar cada uno de los elementos de la lista resultante del motor antes de encontrar una propuesta de solución para su problema, para luego analizar si es la correcta. De esta manera la determinación de criterios de evaluación de la información contenida en los foros podría influir en el orden óptimo de visita de un individuo sobre la lista resultante.

Dentro de la herramienta propuesta en [3], uno de los procesos que la componen, realiza el análisis de los hilos de discusión existentes en foros técnicos para, dado un problema particular, sugerir un conjunto de soluciones exitosas en base al análisis de atributos en los hilos de discusión clasificados por tema, ordenado de mayor a menor probabilidad de éxito.

Del modelo de calidad definido para foros de discusión se desprende un conjunto de métricas. Tanto el modelo como las métricas propuestas son detallados en la Sección 2. En la Sección 3 se detalla la especificación del caso de estudio, cuyos resultados serán analizados en la Sección 4, finalizando con las conclusiones y lineamientos de posibles trabajos futuros en la Sección 5.

		Dimensiones de Calidad																
		Exactitud	Complejidad	Consistencia	Credibilidad	Actualidad	Accesibilidad	Conformidad	Confidencialidad	Eficiencia	Precisión	Trasabilidad	Entendibilidad	Disponibilidad	Portabilidad	Recuperabilidad	Valor Agregado	Adec. Represent.
		Atributos																
HILO	título	●	●															
	cantidad visitas	●	●							●								
USUARIO	nombre																	
	cantidad mensajes	●	●															
	experiencia	●	●		●	●				●	●							
	reputación	●	●		●					●	●		●					
fragmento	texto	●	●										●				●	
	código	●	●					●			●				●		●	
	enlace web	●	●									●		●			●	
	figura	●									●		●				●	
	fecha					●												
MENSAJE	pregunta principal	●	●															
	añade datos	●	●							●	●	●	●	●			●	●
	irrelevante																	
	agradece solución	●	●		●					●	●	●	●	●	●		●	●
	presenta solución	●	●		●					●	●	●	●	●	●		●	●
	pide aclaración																	
	rechaza solución																●	

Figura 2: Modelo conceptual relacionado con características de calidad

2. Modelo de Calidad de Datos

Debido a que se desea adquirir información a partir de los foros de discusión, fue necesario el estudio de las características de calidad para clasificar los datos. En [1] se presentó la matriz que relaciona ciertas características de calidad de datos a atributos del modelo conceptual planteado. Las dimensiones de calidad de la matriz se basan en el estándar ISO/IEC 25012:2008 [4] extendiendo dicho modelo con dos características propuestas por Rafique [5]. De esta manera surge un modelo de calidad de información para foros de discusión, que incluyen las dimensiones de calidad del modelo 25012 con dos dimensiones de calidad agregadas, que son el valor agregado y la adecuación representacional. En la Figura 2 se puede observar con un círculo negro para cada uno de los atributos del modelo conceptual, su posible relación con cada uno de los atributos de calidad definidos, por ejemplo en la fila correspondiente a la reputación se encuentra marcada como posible la relación con el atributo de calidad credibilidad.

En base a los cuadros seleccionados se proponen algunas métricas de calidad correspondientes a las dimensiones de exactitud, credibilidad y portabilidad. Las métricas de exactitud sugieren cuan acertado es un hilo con respecto a la cadena de búsqueda, las métricas de portabilidad aluden a si los enlaces expresados en un hilo, están operativos, y algunas métricas de credibilidad están enfocadas en el grado de expertitud de las personas que intervienen en un hilo.

A continuación se define un grupo de métricas de calidad.

Tabla 1: Plantilla de descripción de Métricas

Nombre (1)	Sigla (2)	Tipo (3)
Nombre en castellano (4)	Descripción de la métrica (5)	
Entidad relacionada (6)		
Atributo relacionado (7)	Descripción del atributo relacionado (8)	
Tipo de escala (9)	Tipo de valor (10)	Unidad (11)
Método de medición (12)	Tipo de método de medición (13)	

2.1. Métricas

Siguiendo la propuesta presentada por Olsina y Martín en [6], la Tabla 1 representa una plantilla para describir las características que posee una métrica de calidad para hilos de discusión técnicos. Una métrica recibe un nombre (1) que generalmente se presenta en inglés y en castellano (4).

Los ítems descriptivos de las métricas son:

- Sigla (2): Forma abreviada de mencionar a la métrica. Por convención se usa el símbolo # (hacen referencia a cantidad) y el símbolo % (porcentaje).
- Tipo (3): El tipo puede ser directa, si no depende de métricas de otros atributos, o indirecta en caso que sí dependa.
- Descripción (5): Explica lo que contabiliza la métrica.
- Entidad relacionada (6): Objeto sobre el cual se realiza la medición (por ejemplo, en nuestro caso la entidad relacionada es el hilo de discusión).
- Atributo relacionado (7): Propiedad física o abstracta medible de una entidad.
- Descripción del atributo relacionado (8): Describe las características del atributo.
- Tipo de escala (9): Puede ser *nominal* (con la especificación de la lista de valores), *ordinal*, *intervalo*, *proporción* o *absoluto*.
- Tipo de valor (10): En escalas nominales es la lista de valores posibles correspondientes a la categorización, en otros casos representará una lista o rango de valores numéricos.
- Unidad (11): Objeto referencia para dimensionar la medición (por ejemplo palabras, líneas de código, mensajes, etc.).
- Método de medición (12): Secuencia lógica de operaciones para realizar la medición que se describe.
- Tipo de método de medición (13): Puede ser de tipo subjetivo (implica juicio o criterio humano), u objetivo (basada en reglas numéricas).

2.2. Especificación de las Métricas

Se definieron 8 métricas, algunas correspondientes a la pertinencia de un hilo de discusión, otras sobre la disponibilidad de información y las últimas sobre expertitud dentro en un hilo de discusión. Todas ellas cumplen las mismas características con respecto a su entidad y a su escala, su entidad son los hilos de discusión y su escala es absoluta. El tipo de valor de las métricas directas

Tabla 2: Métricas directas

Nombre	Nombre castellano	Sigla	Descripción	Atributo relacionado	Descripción atributo relacionado	Unidad	Método de medición
Number of keywords in the title	Cantidad de palabras clave en el título	#KT	palabras clave que aparecen en el título	Pertinencia del hilo	cuan pertinente es el hilo de discusión para la cadena de búsqueda ingresada por un usuario externo en un buscador	palabra	Contar la cantidad de palabras clave que aparecen en el título
Number of keywords in the main question	Cantidad de palabras clave en la pregunta principal	#KMQ	palabras clave que aparecen en la pregunta principal	Pertinencia del hilo	cuan pertinente es el hilo de discusión para la cadena de búsqueda ingresada por un usuario externo en un buscador	palabra	Contar
Number of operational links	Cantidad de links operativos	#OL	links operativos presente en el hilo de discusión	Link operativo	Enlaces web que al referenciar no presentan errores del tipo 404, es decir su contenido puede ser observado	link	contar links operativos a lo largo de todo el hilo de discusión
Number of expert users	Cantidad de usuarios expertos	#EU	usuarios expertos que participan en el hilo de discusión	Expertitud del hilo	usuarios expertos pertenecientes a la comunidad del foro que participan en el hilo de discusión	mensaje	contar la cantidad de usuarios expertos que participan en el hilo de discusión

Tabla 3: Métricas indirectas

Nombre	Nombre castellano	Sigla	Descripción	Atributo relacionado	Descripción atributo relacionado	Método de cálculo
Percentage of keywords in the title	Porcentaje de palabras clave en el título	%KT	idem a descripción de #KT	Pertinencia del hilo	idem a descripción del atributo de #KT	(cantidad de palabras clave encontradas en el título/cantidad de palabras clave buscadas)*100
Percentage of keywords in the main question	Porcentaje de palabras clave en la pregunta principal	%KMQ	idem a descripción de #KMQ	Pertinencia del hilo	idem a descripción del atributo de #KMQ	(cantidad de palabras clave encontradas en la pregunta principal/cantidad de palabras clave buscadas)*100
Percentage of operational links	Porcentaje de links operativos	%OL	idem a descripción de #OL	Link operativo	idem a descripción del atributo de #OL	(cantidad de links operativos/cantidad de links en el hilo)*100
Percentage of expert users	Porcentaje de usuarios expertos	%EU	idem a descripción de #EU	Expertitud del hilo	idem a descripción del atributo de #EU	(cantidad de usuarios experto/cantidad de usuarios)*100

definidas es numérico, específicamente entero positivo, mientras que en todos los casos el tipo de medición es objetivo. Los otros detalles de las métricas directas visualizan en la Tabla 2, donde se puede observar por ejemplo que las siglas en las métricas directas comienzan con el símbolo # debido a que hacen referencia a cantidad de elementos.

Las métricas indirectas definidas, están en función de las directas por lo cual, tanto la descripción de la métrica como del atributo relacionado correspondiente son similares a las directas referenciadas en la Tabla 2. Se puede observar en la Tabla 3 que las siglas comienzan con el símbolo % y que en vez de método de medición se especifica el método de cálculo. Si bien comparten el tipo de valor numérico, a diferencia de las métricas directas, los métodos de cálculo corresponden a valores reales positivos, en estos casos la unidad es normalizada a porcentaje.

Existen algunos casos especiales en las métricas indirectas que son cuando la cantidad de lo que se busca es nula y la cantidad total de referencia del atributo relacionado también. En esos casos (%KT, %KMQ, %EU) se consideran con un valor de 0%. Salvo en la métrica correspondiente a la disponibilidad de información (%OL) que corresponde a los casos en los cuales no tienen links relacionados; en estas ocasiones se puede decir que la información está disponible en su totalidad, por lo que se define un valor del 100%.

La aplicación de las métricas sobre un conjunto de hilos de discusión ayudaría a establecer prioridades en la búsqueda de soluciones, para ello la propuesta es utilizar las cuatro métricas indirectas detalladas para determinar un orden de prioridad al adquirir información en debates.

2.3. Pesos relativos y Ranking

Con las métricas indirectas propuestas, se va a definir el ranking para cada hilo obtenido como resultado de una consulta; con el fin de lograr este objetivo se propone la asignación de un vector de pesos relativos, definido con un peso específico otorgado a cada una de las métricas indirectas. El vector se conforma por los siguientes pesos relativos:

- PRKT = Peso relativo otorgado a la métrica %KT
- PRKMQ = Peso relativo otorgado a la métrica %KMQ
- PROL = Peso relativo otorgado a la métrica %OL
- PREU = Peso relativo otorgado a la métrica %EU

Para un primer análisis, se considerará el mismo peso a cada una de las métricas, esto es un valor de 0.25. Entonces, en base a las métricas indirectas presentadas y al vector de pesos relativos, se determina el ranking que representa la relevancia de un hilo de discusión respecto a una cadena de búsqueda dada, al cual llamaremos $Ranking_{Tool}$. La forma de obtener dicho ranking se define con la siguiente fórmula:

$$Ranking_{Tool} = \%KT * PRKT + \%KMQ * PRKMQ + \%OL * PROL + \%EU * PREU$$

Por lo tanto, el valor $Ranking_{Tool}$ es el valor relativo asignado a un hilo de un foro de discusión respecto a una cadena de búsqueda. De esta manera, para obtener la proporción del ranking con respecto a la sumatoria de los valores obtenidos para todos los hilos que forman parte del conjunto bajo estudio, se utiliza la siguiente ecuación:

$$\%Relativo\ Ranking_{Tool} = \frac{Ranking_{ForumMet}\ del\ hilo\ i}{\sum_{j=1}^n Ranking_{ForumMet}\ del\ hilo\ j}$$

Para analizar el ranking dentro de un conjunto de hilos utilizaremos un ejemplo obtenido a partir de un caso de estudio.

3. Caso de Estudio

Un caso de estudio se desarrolla con el objetivo de investigar una entidad o fenómeno particular en el contexto de la vida real [7]. Los resultados al aplicar las métricas sobre 3 hilos de discusión recuperados de la Web utilizando la cadena de búsqueda: “*error en java nullpointerexception*” fueron los siguientes:

Hilo 1

- Las métricas retornadas fueron $\#KT = 1$, $\%KT = 33,33\%$, $\#KMQ = 1$ y $\%KMQ = 33,33\%$. De las 3 palabras clave de la cadena de búsqueda (error, java y nullpointerexception), en el título aparece solo la palabra *nullpointerexception*
- Este hilo presenta un link a contenido externo $\#OL = 1$ y $\%OL = 100\%$, dado que dicho enlace está operativo.
- Los resultados para estas métricas fueron $\#EU = 0$ y $\%EU = 0\%$.

Hilo 2

- Se obtuvieron $\#KT = 2$ y $\%KT = 66,66\%$. En la pregunta principal sucede algo similar y los resultados obtenidos fueron $\#KMQ = 2$ y $\%KMQ = 66,66\%$.
- En la Respuesta 3 aparece un link a las preguntas frecuentes de Java (FAQ's de Java), el cual está operativo. Luego las métricas retornaron los siguientes resultados $\#OL = 1$ y $\%OL = 100\%$.
- Dado que el foro no define roles para sus miembros, las métricas obtenidas fueron $\#EU = 0$ y $\%EU = 0\%$.

Hilo 3

- Se obtuvieron las mediciones $\#KT = 3$, $\#KMQ = 3$, $\%KT = 100\%$, $\%KMQ = 100\%$.
- El hilo no contiene links, por lo tanto las métricas para enlaces operativos presentaron los valores $\#OL = 0$ y $\%OL = 100\%$. Esto significa que está operativo en su totalidad.
- El foro define roles para sus miembros, en las respuestas participa un *moderador global avanzado* y el usuario que inició la pregunta. Por lo tanto se obtuvo $\#EU = 1$ y $\%EU = 100\%$.

En la Figura 3 se presenta los gráficos comparativos de las métricas del caso de estudio, en la Figura 3 (a) se comparan para los 3 hilos las métricas directas o de cantidad (#) y las indirectas o de porcentaje (%) en la Figura 3b(b).

4. Validación

Con el objetivo de evaluar los resultados obtenidos en el casos de estudio se elaboró una encuesta donde se solicitó a 10 informáticos, con amplios conocimientos en programación Java y usuarios frecuentes de foros de discusión, que dada una cadena de búsqueda establecieran el orden de relevancia (1, 2 o 3) de los 3 hilos tomados como base del caso de estudio, según el siguiente criterio:

1. Es el hilo más relacionado a la cadena de búsqueda y ofrece soluciones muy pertinentes (sin dudas, lo visitaría primero).

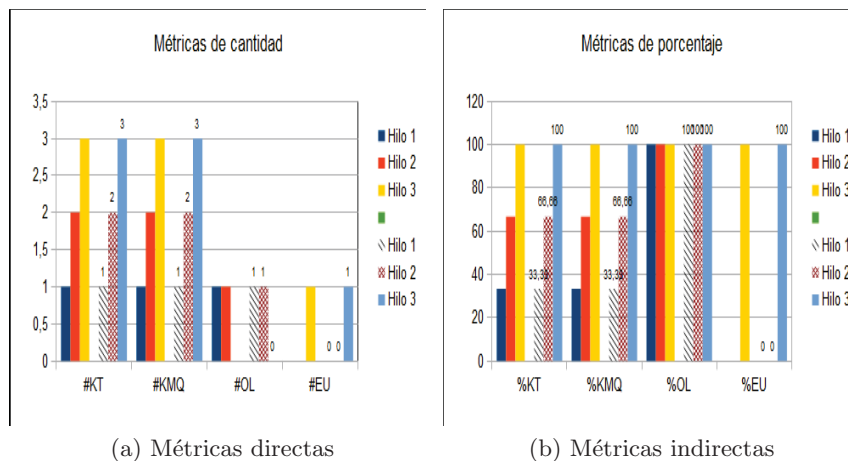


Figura 3: Métricas obtenidas en el caso de estudio

2. El hilo está muy relacionado a la cadena de búsqueda, pero hay otro que está más relacionado aún.
3. Es el hilo menos relacionado a la cadena de búsqueda de los tres.

Las opciones fueron ponderadas para poder calcular el orden de relevancia de cada hilo para cada uno de los expertos, dándole una puntuación de 20 a la opción 1, 10 puntos a la opción 2 y 0 puntos a la última opción.

En la Tabla 4 se presentan las respuestas que se obtuvieron por parte de los 10 expertos consultados. Por ejemplo en la Tabla 4 (a) se observa en la primera fila que el primer experto le dio el orden de mayor importancia al hilo 2, en segundo lugar el hilo 1 y por último el hilo 3 mientras que el segundo experto consideró con mayor importancia el hilo 3. Los valores ponderados para los órdenes de primero (1), segundo (2) y tercero (3) son 20, 10 y 0 respectivamente. En la Tabla 4 (b) se muestra los valores asignados ya ponderados, donde en la fila inferior se destaca la suma de los valores ponderados para cada hilo (por ejemplo, la suma de los valores ponderados para el Hilo 1 es 90, para el Hilo 2 es 60 y para el Hilo 3 es 150).

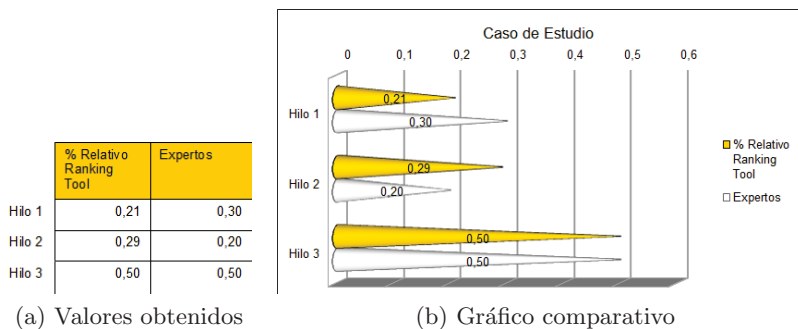
Orden hilo 1	Orden hilo2	Orden hilo3
2	1	3
2	3	1
2	3	1
3	2	1
3	2	1
1	2	3
3	2	1
1	3	2

Hilo 1	Hilo 2	Hilo 3
10	20	0
10	0	20
10	0	20
0	10	20
0	10	20
20	10	0
0	10	20
20	0	10
10	0	20
10	0	20
90	60	150

(a) Asignado (b) Ponderado

Tabla 4: Orden según encuesta a los expertos

Al analizar el caso de estudio, se destaca en la Tabla 4(a), la comparativa con los valores del %Relativo Ranking_{Tool} en la primer columna y el resultado de los encuestados en la columna titulada Expertos, mientras que en la Figura 4(b) se puede visualizar el gráfico comparativo de los datos de dicha tabla. Al comparar los resultados se puede observar que tanto el ranking relativo como los expertos sugieren el Hilo 3 en el primer lugar del ranking de visita. Aunque los Hilos 1 y 2 se encuentran en distinto orden de visita, la diferencia no sería significativa ($\leq 0,10$ en ambos casos).



(a) Valores obtenidos (b) Gráfico comparativo

Figura 4: Comparación Ranking vs. valoración de Expertos

5. Conclusiones y trabajo futuro

En este trabajo se presentan algunas métricas para medir la calidad de la información contenida en los hilos de discusión, se establece un criterio para determinar a partir de las métricas el orden de prioridad brindando un peso

específico a cada una de las métricas. Luego utilizando un caso de estudio enfocado en hilos de foros de discusión técnicos, y utilizando para validar criterios de expertos, se contrastaron los resultados obtenidos. Al comparar los datos de la encuesta realizada a un conjunto de expertos y el orden sugerido, se puede detectar que estos no difieren en gran medida. Por un lado, el orden de relevancia coincidió en el hilo más relevante y, aunque otorgó un orden distinto para el segundo y tercer hilo, la diferencia en ambos casos fue menor a 0,10 en una escala de 0 a 1, por lo que esta diferencia podría considerarse no significativa. Esta primer aproximación invita a contrastar resultados utilizando las mismas métricas en diferentes casos de estudio. A futuro se planea analizar bajo qué contexto una métrica podría ser más relevante que otra y ponderar de manera distinta el vector de pesos relativos. Si bien es un resultado parcial, en este caso existe coincidencia entre el ranking relativo y el otorgado por los expertos, es necesario entonces extender los casos de estudio y replicar los experimentos realizados.

Agradecimientos

Este trabajo está parcialmente soportado por el subproyecto “*Reuso de Conocimiento en Foros de Discusión, Parte II*”, correspondiente al Programa de Investigación 04/F009 “*Desarrollo Orientado a Reuso, Parte II*”, de la Universidad Nacional del Comahue (Neuquén, Argentina).

Referencias

1. G. Aranda, N. Martínez, P. Faraci, and A. Cechich, “Hacia un framework de evaluación de calidad de información en foros de discusión técnicos,” in *ASSE 2013-Simposio Argentino de Ingeniería de Software, JAIIO 42^o-Jornadas Argentinas de Informática*, (Córdoba, Argentina), SADIO, 2013.
2. Martínez Carod, Nadina, Aranda, Gabriela, Sagripanti, Mauro, Faraci, Pamela, and Cechich, Alejandra, “Análisis de la información presente en foros de discusión técnicos,” in *CACIC 2013, XIX Congreso Argentino de Ciencias de la Computación*, (Mar del Plata, Argentina), pp. 847–856, Oct. 2013.
3. Aranda, Gabriela, Martínez Carod, Nadina, Roger, Sandra, Faraci, Pamela, and Cechich, Alejandra, “Una herramienta para el análisis de hilos de discusión técnicos,” in *CACIC 2014, XX Congreso Argentino de Ciencias de la Computación*, (San Justo, Argentina), pp. 803 – 812, Oct. 2014.
4. “ISO/IEC 25012:2008, Software product quality requirements and evaluation (SQuaRE): Data quality model,” 2008.
5. I. Rafique, P. Lew, M. Q. Abbasi, and Z. Li, “Information quality evaluation framework: Extending ISO 25012 data quality model,” *International Journal of Computer and Information Sciences*, vol. 6, no. 1, pp. 1 – 6, 2012.
6. L. Olsina and M. d. l. A. Martín, “Ontology for software metrics and indicators,” *Journal of Web Engineering*, vol. 2, no. 4, pp. 262–281, 2004.
7. C. Wohlin, P. Runeson, M. Höst, M. Ohlsson, B. Regnell, and A. Wesslén, *Experimentation in Software Engineering*. Computer Science, Springer, 2012.

Asistente para la evaluación de características de calidad de producto de software propuestas por ISO/IEC 25010 basado en métricas definidas usando el enfoque GQM

Julieta Calabrese, Rocío Muñoz, Ariel Pasini, Silvia Esponda, Marcos Boracchia, Patricia Pesado

Instituto de Investigación en Informática LIDI (III-LIDI)*
Facultad de Informática – Universidad Nacional de La Plata
50 y 120 La Plata Buenos Aires

*Centro Asociado Comisión de Investigaciones Científicas de la Pcia. de Bs. As. (CIC)
{jcalabrese, rmunoz, apasini, sesponda, marcosb, ppesado}@lidi.info.unlp.edu.ar

Abstract. Se presenta un asistente para evaluar las características de un producto de software, propuestas por la ISO/IEC 25010 mediante el enfoque GQM (Goal, Question, Metric). Se definieron un conjunto de preguntas cuyas respuestas combinadas de forma lógica permiten obtener una métrica aplicable a las características que propone ISO/IEC 25010. Para este trabajo se tomó como caso de estudio la característica de Seguridad, se definieron las métricas y luego se muestran los resultados de la aplicación a tres casos de estudio.

Keywords: Calidad, Producto de software, GQM, ISO/IEC 25000

1 Introducción

El número de empresas desarrolladoras de software ha experimentado un fuerte crecimiento, juntamente con el incremento de la demanda de productos del sector. Para este tipo de empresas, la calidad del software tiene un papel fundamental, en particular como elemento diferenciador de competitividad y de imagen frente a sus clientes y porque consecuentemente, las pérdidas económicas que los problemas de la calidad en el software pueden ocasionar son considerables. En este contexto, las actividades relacionadas con la calidad de software y su evaluación, están cobrando cada vez más importancia. [1]

Una organización puede interesarse en evaluar su producto pues desea diferenciarse de los competidores, asegurando tiempos de entrega y reducción de fallos en el producto tras su implantación en producción; establecer acuerdos en el ámbito del servicio, definiendo parámetros de calidad que el producto debe cumplir antes de ser entregado; detectar los defectos en el producto software y proceder a su eliminación antes de la entrega; evaluar y controlar el rendimiento del producto software desarrollado, asegurando que podrá generar los resultados teniendo en cuenta las restricciones de tiempo y recursos establecidas; asegurar que el producto software

desarrollado respeta los niveles necesarios para las características de seguridad (*Confidencialidad, Integridad, Autenticidad, No-Repudio*, etc.) entre otras.

En este sentido, la familia ISO 25000 conocida como SQuaRE (Software Product Quality Requirements and Evaluation) nace para dar respuesta a estas necesidades. Su objetivo es la creación de un marco de trabajo común para evaluar la calidad del producto software, sustituyendo a las anteriores ISO/IEC 9126 e ISO/IEC 14598 Modelos de calidad y generación de métricas. [2]–[4]

En el presente trabajo se propone un asistente de evaluación de productos de software basado en las métricas definidas en ISO/IEC 25010 usando el enfoque GQM. [5]

En la sección 2 se describe brevemente la familia de la ISO/IEC 25000 y el enfoque que propone GQM. A continuación, apartado 3, se describe el modelo para realizar la evaluación de las características propuestas por ISO/IEC 25010 bajo el enfoque GQM, en particular de la característica de *Seguridad*. Luego se presentan tres casos de estudio, donde se aplica el modelo de evaluación junto con los resultados obtenidos de las mismas. Por último, las conclusiones del presente trabajo.

2 Modelos de calidad y generación de métricas

2.1 La familia de ISO/IEC 25000.

La gestión de la calidad se impone en las organizaciones por la importancia que alcanza en diferentes aristas: a nivel de sus productos, permitiendo establecer la calidad lograda y las características presentes en los mismos; a nivel de la organización, ocupándose de establecer un marco de procesos que permita obtener una mejora; como asimismo a nivel de proceso.

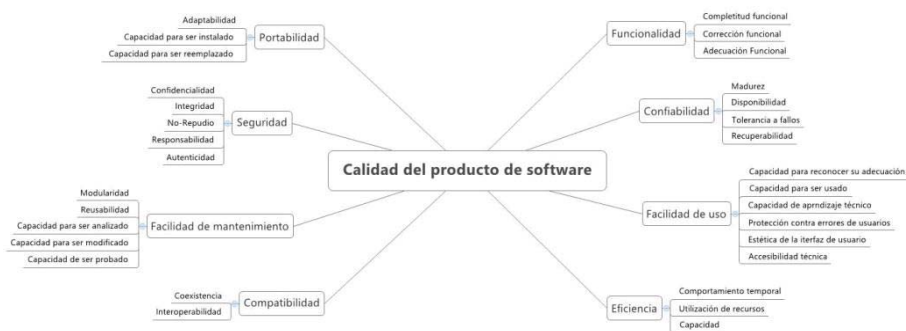
Con el objetivo de organizar y unir todas las normas relacionadas con la calidad de los productos de software, en el año 2005 ISO/IEC publica ISO/IEC 25000:2005 - SQuaRE (Requisitos y evaluación de la calidad del producto de software), también conocida como la familia ISO 25000. Dentro del conjunto del ISO/IEC 25000, se destacan ISO/IEC 25010 - *System and software quality models* e ISO/IEC 25040 - *Evaluation process* que se describen a continuación.

ISO/IEC 25010 - System and software quality models.

Reemplaza a la ISO/IEC 9126-1:2001. Incorpora nuevas características internas y externas, agrupándolas bajo el nombre de calidad del producto de software. La principal modificación es la incorporación de la característica *Compatibilidad* que se relaciona con la posibilidad de intercambio de información entre sistemas, y la característica *Seguridad* que se relaciona con los conceptos de confidencialidad y acceso a la información.[6]

Cada una de las características de la calidad del producto de software posee subcaracterísticas que las especifican más claramente según se muestra en la Figura 1.

Figura 1



ISO/IEC 25040 - Evaluation process.

Reemplaza a la ISO/IEC 14598-1:1999. La nueva versión define 13 procesos en cinco etapas:

- 1) Establecer los requisitos de la evaluación: a. Establecer el propósito de la evaluación. b. Obtener los requisitos de calidad del producto. c. Identificar las partes del producto que se deben evaluar. d. Definir el rigor de la evaluación.
- 2) Especificar la evaluación: a. Seleccionar los módulos de evaluación. b. Definir los criterios de decisión para las métricas. c. Definir los criterios de decisión de la evaluación.
- 3) Diseñar la evaluación: a. Planificar las actividades de la evaluación.
- 4) Ejecutar la evaluación: a. Realizar las mediciones. b. Aplicar los criterios de decisión para las métricas. c. Aplicar los criterios de decisión de la evaluación.
- 5) Finalizar la evaluación: a. Revisar los resultados de la evaluación. b. Crear el informe de evaluación. c. Revisar la calidad de la evaluación y obtener feedback. d. Tratar los datos de la evaluación. [7]

2.2 GQM. (Goal, Question, Metric)

GQM (Goal, Question, Metric) es un método orientado a lograr una métrica que mida cierto objetivo de una manera determinada. El modelo de medición tiene tres niveles:

- Nivel Conceptual (Goal/Objetivo): se define un objetivo para un objeto, el cual puede ser un producto, un proceso o un recurso, con respecto a varios modelos de calidad, desde varios puntos de vista y relativo a un entorno particular.
- Nivel Operativo (Question/Pregunta): se refina un conjunto de preguntas a partir del objetivo, con el propósito de verificar su cumplimiento. Las preguntas tratan de caracterizar el objeto de medición (producto, proceso o recurso) con respecto a una cuestión de calidad seleccionada y determinar su calidad desde el punto de vista seleccionado.

- Nivel Cuantitativo (Metric/Métrica): se asocia un conjunto de métricas, que pueden ser objetivas o subjetivas, para cada pregunta, de modo de responder a cada una de un modo cuantitativo.

Un modelo GQM se desarrolla identificando un conjunto de objetivos de calidad y/o productividad, a nivel corporativo, de división o de proyecto. A partir de esos objetivos y en base a modelos del objeto de medición, se elaboran preguntas que definen esos objetivos de la manera más completa posible. El siguiente paso consiste en especificar las medidas que deben ser tomadas para responder a esas preguntas y para realizar un seguimiento de la conformidad de los productos y procesos con los objetivos. Una vez especificadas las medidas, es necesario desarrollar los mecanismos de recopilación de información, incluidos los mecanismos de validación y análisis. [5]

3 Modelo de Evaluación de características de Calidad

El modelo desarrollado consiste en definir un conjunto de preguntas basadas en el enfoque GQM que luego, mediante conectivos lógicos, indicarán la satisfacción de los objetivos propuestos.

Como caso de estudio se tomó la característica **Seguridad** que contiene las subcaracterísticas: *Confidencialidad*, *Integridad*, *No-Repudio*, *Responsabilidad* y *Autenticidad*.

CONFIDENCIALIDAD: Evalúa la capacidad de protección contra el acceso de datos e información no autorizados, ya sea accidental o deliberadamente.

INTEGRIDAD: Evalúa la capacidad del sistema o componente para prevenir accesos o modificaciones no autorizados a datos o programas de ordenador.

NO-REPUDIO: Evalúa la capacidad de demostrar las acciones o eventos que han tenido lugar, de manera que dichas acciones o eventos no puedan ser repudiados posteriormente.

RESPONSABILIDAD: Evalúa la capacidad de rastrear de forma inequívoca las acciones de una entidad.

AUTENTICIDAD: Evalúa la capacidad de demostrar la identidad de un sujeto o un recurso.

3.1 Cuestionario

En función de las subcaracterísticas descriptas anteriormente, se definieron 33 preguntas para responder con verdadero/falso. Tabla 1

Tabla 1. Cuestionario para la característica de *Seguridad*

ID	PREGUNTA
P1	¿Se requiere que la contraseña posea al menos 8 caracteres?
P2	¿Se requiere que la contraseña posea letras mayúsculas y minúsculas?
P3	¿Se requiere que la contraseña posea números y letras?

P4	¿Se requiere que la contraseña posea caracteres especiales?
P5	¿El sistema utiliza conexión segura mediante HTTPS?
P6	¿La base de datos posee los datos encriptados?
P7	¿El sistema permite acceder a funcionalidades en las cuales no se tiene permiso?
P8	¿El sistema permite que cualquier persona tenga acceso a la base de datos?
P9	¿El sistema permite que cualquier persona tenga acceso al código del servidor de la aplicación?
P10	¿Cualquier persona tiene acceso al servidor físico?
P11	¿Cualquier persona tiene acceso al servidor remoto?
P12	¿El sistema posee redireccionamientos hacia sitios no seguros?
P13	¿El sistema solicita una confirmación de registro mediante un mail a la hora de registrarse?
P14	¿El sistema permite que cualquier persona pueda modificar la base de datos?
P15	¿El sistema permite que cualquier persona pueda modificar el código del servidor de la aplicación?
P16	¿El sistema permite inyecciones SQL?
P17	¿El sistema posee un historial de acciones realizadas?
P18	¿El sistema posee algoritmos de cifrado de datos?
P19	¿El sistema posee un mecanismo criptográfico, como firma digital?
P20	¿El sistema solicita confirmación a la hora de realizar una acción?
P21	¿El sistema posee una protección con certificados SSL?
P22	¿El sistema da aviso cuando se accede desde una ubicación desconocida?
P23	¿El sistema informa vía mail las operaciones realizadas?
P24	¿El sistema guarda un registro de fecha y hora de ingreso al mismo?
P25	¿El sistema registra el tipo de navegador y sistema de operación utilizado para ingresar al sitio?
P26	¿El sistema registra la dirección IP desde la cual se ingresa al sitio?
P27	¿El sistema realiza una comprobación de identidad mediante un certificado digital?
P28	¿El sistema posee un sistema de verificación en dos pasos?
P29	¿Es requerida una clave de segundo nivel para el ingreso al sistema?
P30	¿El sistema realiza una comprobación de identidad mediante datos biométricos?
P31	¿El sistema realiza una comprobación de identidad mediante tarjeta de coordenadas?
P32	¿El sistema realiza una comprobación de identidad mediante credenciales?
P33	¿El sistema realiza una comprobación de identidad mediante una firma electrónica?

3.2 Descripción de criterios de evaluación (CE)

Con el fin de lograr el objetivo, las respuestas de las preguntas fueron combinadas de forma lógica estableciendo un puntaje a cada uno de los CE.

Tabla 2. Descripción de criterios de evaluación (CE)

<i>ID</i>	<i>Nombre</i>	<i>Descripción</i>	<i>Fórmula</i>	<i>Ptos</i>
C-1	Conexiones seguras	Una conexión se considera segura si se utiliza HTTPS y si no se tienen redireccionamientos hacia sitios no seguros	$P5 \ \& \ \sim P12 = V$	1
C-2	Control de acceso	Se debe controlar que no se permita acceder a funcionalidades sin autorización, tampoco a la base de datos, al código de la aplicación ni a los servidores, físico ni remoto	$si \ P7 \ \ P8 \ \ P9 \ \ P10 \ \ P11 = F$	1
C-3	Encriptación de datos	Los datos de la base de datos deben estar encriptados	$P6 = V$	1
C-4	Contraseña de bajo nivel	La contraseña se considera de bajo nivel si posee menos de 8 caracteres, no posee letras mayúsculas y minúsculas, no posee letras y números y no posee caracteres especiales	$P1 \ \ P2 \ \ P3 \ \ P4 = F$	0
	Contraseña de medio nivel	La contraseña se considera de medio nivel si posee al menos 8 caracteres o letras mayúsculas y minúsculas o letras y números o	$P1 \ \ P2 \ \ P3 \ \ P4 = V$	0.5
	Contraseña de alto nivel	La contraseña se considera de alto nivel si posee al menos 8 caracteres, letras mayúsculas y minúsculas, letras y números y caracteres especiales	$P1 \ \& \ P2 \ \& \ P3 \ \& \ P4 = V$	1
I-5	Prevención de accesos	Se debe prevenir que no se permita acceder a funcionalidades sin autorización, tampoco a la base de datos ni al código de la aplicación, y que no se permitan inyecciones SQL	$P7 \ \ P8 \ \ P9 \ \ P16 = F$	1
I-6	Prevención de modificaciones	Se debe prevenir que no se permita modificar datos de la base de datos ni modificar el código de la aplicación sin autorización	$P14 \ \ P15 = F$	1
I-7	Confirmación de datos	Se debe realizar una confirmación de registro por mail	$P13 = V$	1
NR-8	Operaciones realizadas	Se debe poseer un historial de acciones realizadas o las mismas deben ser enviadas por mail	$P17 \ \ P23 = V$	1
NR-9	Mecanismos de cifrado	Se debe poseer un algoritmo de cifrado de datos o un mecanismo criptográfico, como firma digital, o una protección con certificados SSL	$P18 \ \ P19 \ \ P21 = V$	1
NR-10	Confirmación de acciones	Se debe solicitar una confirmación al realizar una determinada acción	$P20 = V$	1
NR-11	Registro de ubicación	Se debe informar si se accedió al sistema desde una ubicación desconocida	$P22 = V$	1

R-12	Registro de acciones y datos	Se debe poseer un historial de acciones realizadas, o un registro de fecha y hora de ingreso al sistema o de la dirección IP desde la cual se ingresa o del tipo de navegador y sistema de operación utilizado	P17 P24 P25 P26 = V	1
R-13	Control de ubicación	Se debe dar aviso cuando se accede al sistema desde una ubicación desconocida	P22 = V	1
A-14	Comprobación de identidad	El sistema debe realizar una comprobación de identidad mediante alguno de los siguientes métodos: datos biométricos, tarjeta de coordenadas, credenciales, firma electrónica o certificado digital	P27 P30 P31 P32 P33 = V	1
A-15	Comprobación adicionales	Se debe poseer un sistema de verificación en dos pasos, o se debe requerir una clave de segundo nivel para el ingreso al sistema o una confirmación de registro mediante un mail	P28 P29 P13 = V	1

3.3 Métricas para cada subcaracterística.

Se combinaron los CE para definir las métricas que satisfacen los objetivos de las subcaracterísticas. Para cada una se definió un nombre, un propósito, un método de aplicación, valores de entradas y fórmula aplicada.

Confidencialidad.

Métrica: *Confidencialidad*

Propósito: *¿Cuán eficiente es el sistema a la hora de proteger el acceso de datos e información no autorizados, ya sea accidental o deliberadamente?*

Método de aplicación: *Contestar las preguntas de los CE correspondientes a la subcaracterística "Confidencialidad" y calcular la puntuación obtenida, sumando los puntajes de los CE referenciados que cumplan con la meta esperada. "Puntaje total" hace referencia al máximo puntaje que se puede obtener.*

Entradas: *A = Puntaje obtenido. B = Puntaje total.*

Fórmula: $X = A/B$

Observaciones: *Los CE a utilizar son: C-1, C-2, C-3 y C-4.*

Integridad.

Métrica: *Integridad*

Propósito: *¿Cuán capaz es el sistema a la hora de prevenir accesos o modificaciones no autorizados a datos o programas de ordenador?*

Método de aplicación: *Contestar las preguntas de los CE correspondientes a la subcaracterística "Integridad" y calcular la puntuación obtenida, sumando los puntajes de los CE referenciados que cumplan con la meta esperada. "Puntaje total" hace referencia al máximo puntaje que se puede obtener.*

Entradas: $A = \text{Puntaje obtenido}$. $B = \text{Puntaje total}$.

Fórmula: $X = A/B$

Observaciones: Los CE a utilizar son: I-5, I-6 e I-7.

No-Repudio.

Métrica: *No-Repudio*

Propósito: *¿Cuán capaz es el sistema de demostrar las acciones o eventos que han tenido lugar, de manera que dichas acciones o eventos no puedan ser repudiados posteriormente?*

Método de aplicación: *Contestar las preguntas de los CE correspondientes a la subcaracterística "No-Repudio" y calcular la puntuación obtenida, sumando los puntajes de los CE referenciados que cumplan con la meta esperada. "Puntaje total" hace referencia al máximo puntaje que se puede obtener.*

Entradas: $A = \text{Puntaje obtenido}$. $B = \text{Puntaje total}$.

Fórmula: $X = A/B$

Observaciones: Los CE a utilizar son: NR-8, NR-9, NR-10 y NR-11.

Responsabilidad.

Métrica: *Responsabilidad*

Propósito: *¿Cuán capaz es el sistema de rastrear de forma inequívoca las acciones de una entidad?*

Método de aplicación: *Contestar las preguntas de los CE correspondientes a la subcaracterística "Responsabilidad" y calcular la puntuación obtenida, sumando los puntajes de los CE referenciados que cumplan con la meta esperada. "Puntaje total" hace referencia al máximo puntaje que se puede obtener.*

Entradas: $A = \text{Puntaje obtenido}$. $B = \text{Puntaje total}$.

Fórmula: $X = A/B$

Observaciones: Los CE a utilizar son: R-12 y R-13.

Autenticidad.

Métrica: *Autenticidad*

Propósito: *¿Cuán capaz es el sistema de demostrar la identidad de un sujeto o un recurso?*

Método de aplicación: *Contestar las preguntas de los CE correspondientes a la subcaracterística "Autenticidad" y calcular la puntuación obtenida, sumando los puntajes de los CE referenciados que cumplan con la meta esperada. "Puntaje total" hace referencia al máximo puntaje que se puede obtener.*

Entradas: $A = \text{Puntaje obtenido}$. $B = \text{Puntaje total}$.

Fórmula: $X = A/B$

Observaciones: Los CE a utilizar son: A14 y A15.

Las fórmulas aplicadas para cada subcaracterística se presentan en la Tabla 3.

Tabla 3. Fórmula para cada subcaracterística

MÉTRICA	FÓRMULA
CONFIDENCIALIDAD	$(C1+C2+C3+C4)/4$
INTEGRIDAD	$(I5+I6+I7)/3$
NO REPUDIO	$(NR8+NR9+NR10+NR11)/4$
RESPONSABILIDAD	$(R12+R13)/2$
AUTENTICIDAD	$(A14+A15)/2$

4 Casos de Estudio

Se realizó el proceso de evaluación según la estructura de la ISO/IEC 25040 en tres aplicaciones web, con el propósito de evaluar la característica de *Seguridad*.

Caso a) Se encuentra en producción desde hace aproximadamente 18 meses, posee más de 3200 usuarios, con una frecuencia promedio de 500 accesos diarios.

Caso b) Se encuentra en producción desde hace aproximadamente 30 meses, posee más de 160 usuarios, con una frecuencia promedio de uso de 75 accesos diarios

Caso c) Se encuentra en etapa de testing, posee diez usuarios con una frecuencia mínima por los usuarios que realizan el testing de la aplicación.

4.1 Establecer los requisitos de la evaluación

El propósito de la evaluación es medir, analizando diferentes aspectos, la seguridad de tres sistemas web. En base al propósito se selecciona la característica “*Seguridad*” definida en la norma ISO/IEC 25010.

Dos de los sistemas a evaluar se encuentran en su versión final y están siendo utilizados por diferentes usuarios. El sistema restante se encuentra en una versión de prueba, y está bajo el uso de diferentes personas encargadas del testing.

4.2 Especificar la evaluación

Las métricas para las subcaracterísticas, son las definidas en el apartado 3). Los criterios de aceptación para las subcaracterísticas son:

Inaceptable: $0 \leq X < 40$

Mínimamente aceptable: $40 \leq X < 60$

Rango objetivo: $60 \leq X < 90$

Excede los requerimientos: $90 \leq X \leq 100$

El propósito se considerará aceptado si todas las subcaracterísticas se encuentran entre los rangos mínimamente aceptables y excede los requerimientos

4.3 Diseñar la evaluación

Para realizar la evaluación se les solicitó a tres personas encargadas del desarrollo de los sistemas web a evaluar (un desarrollador por cada sistema) que respondieran con V/F las diferentes preguntas planteadas en el apartado 3.1. Además, se les brindó una

planilla de Excel preparada para el ingreso de las respuestas, y en la misma se obtuvieron automáticamente los valores A y B de las métricas.

4.4 Ejecutar la evaluación

Se ejecutó la evaluación según lo planificado y se obtuvieron los siguientes resultados:

Caso a) *Confidencialidad* 88%, *Integridad* 67%, *No-Repudio* 50%, *Responsabilidad* 50% y *Autenticidad* 0%

Caso b) *Confidencialidad* 75%, *Integridad* 100%, *No-Repudio* 75%, *Responsabilidad* 50% y *Autenticidad* 50%

Caso c) *Confidencialidad* 0%, *Integridad* 33%, *No-Repudio* 50%, *Responsabilidad* 0% y *Autenticidad* 0%

4.5 Finalización de la evaluación

El caso a) posee las subcaracterísticas *Confidencialidad* e *Integridad* en el rango aceptable, *No-Repudio* y *Responsabilidad* mínimamente aceptable y *Autenticidad* inaceptable.

El caso b) posee las subcaracterísticas *Integridad* en el rango excede las expectativas, *Confidencialidad* y *No-Repudio* en el rango aceptable, *Responsabilidad* y *Autenticidad* mínimamente aceptable.

El caso c) posee las subcaracterísticas *No-Repudio* en el rango mínimamente aceptable y *Responsabilidad*, *Autenticidad*, *Integridad* y *Confidencialidad* inaceptable.

En la figura 2 se presenta la comparación de las subcaracterísticas evaluadas de cada uno de los casos.

Análisis de la característica – Seguridad

El caso a) no cumple con el propósito de la evaluación ya que la subcaracterística *Autenticidad* se encuentra en un rango inaceptable: el sistema no realiza una comprobación de identidad mediante ningún método ni posee un sistema de verificación en dos pasos, una clave de segundo nivel o una confirmación de registro mediante correo electrónico.

El caso b) se considera que cumple con el propósito de la evaluación ya que todas sus subcaracterísticas se encuentran en un rango de aceptación.

El caso c) no cumple con el propósito de la evaluación ya que sólo la subcaracterística *No-repudio* se encuentra en un rango aceptable. La subcaracterística *Autenticidad* es inaceptable por los mismos motivos que en el caso a). En el caso de la *Responsabilidad*, la subcaracterística se considera inaceptable debido a que no se posee un historial de acciones o un registro de acceso, ni tampoco se da aviso cuando

se accede desde una ubicación desconocida. En cuanto a la subcaracterística *Confidencialidad*, el sistema no posee conexiones seguras ni control de acceso, tampoco se cuenta con una encriptación de los datos de la base de datos ni se establecen criterios para la creación de contraseñas seguras. Por último, la subcaracterística *Integridad*: si bien cuenta con una prevención de modificaciones no autorizadas de la base de datos y del código, no es suficiente para que alcance un nivel de aceptación porque no presenta mecanismos para prevenir el acceso a funcionalidades sin autorización ni realiza confirmaciones de datos vía correo electrónico.

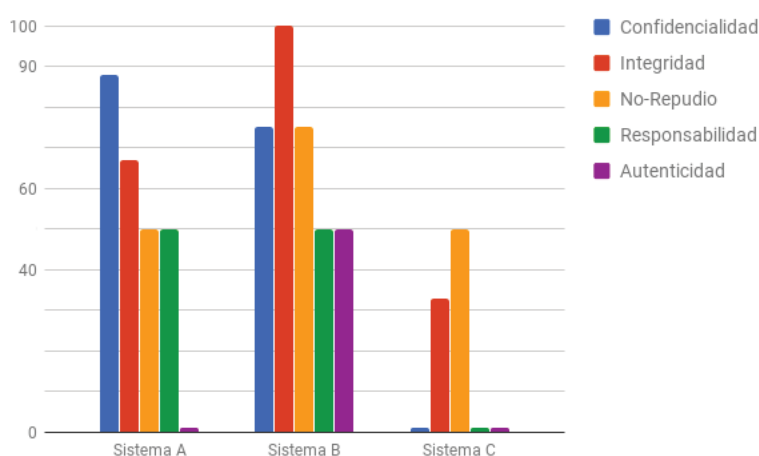


Fig. 1 Subcaracterísticas de cada caso de estudio

5 Conclusiones

La norma ISO/IEC 25010 brinda un modelo de calidad para la evaluación de un conjunto de características aplicables a un producto de software. Se presentó un modelo de evaluación para las características y subcaracterísticas basado en el enfoque GQM, el cual parte de un objetivo concreto para luego crear preguntas asociadas a dicho objetivo, y mediante la combinación de las mismas obtener la métrica en cuestión. Se generaron preguntas para la subcaracterística de la característica *Seguridad* y un conjunto de reglas de evaluación para las respuestas a las preguntas, que combinadas generaron las métricas a cada subcaracterística y, en consecuencia, las métricas de la característica.

Se realizó la evaluación en tres sistemas web, donde sólo uno pasó la evaluación de forma positiva. La evaluación del resto de los sistemas fue de mucha utilidad para detectar falencias en los mismos.

Se proyecta ampliar el modelo, generando las preguntas y los criterios de evaluación para las restantes características de la norma ISO/IEC 25010.

6 Referencias

1. S. Esponda, P. Pesado, “Ambiente para la ayuda a la mejora de procesos en las PyMEs”, 2013.
2. ISO, “ISO/IEC 25000:2014 Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- Guide to SQuaREtle,” 2014.
3. IRAM and ISO, “IRAM-NM-ISO IEC 9126-1 Information technology. Software engineering. Product quality. Part 1 - Quality model.,” 2009.
4. IRAM;ISO, “IRAM-ISO-IEC 14598-1 Information technology. Software engineering. Software product evaluation. Part 1: General overview,” 2006.
5. V. R. Basili, G. Caldiera, and H. D. Rombach, “The goal question metric approach,” vol. 2, pp. 1–10.
6. ISO, “ISO/IEC 25010:2011 Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- System and software quality models,” 2011.
7. ISO, “ISO/IEC 25040:2011 Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- Evaluation process,” 2011.

Dificultades de los “recién llegados” a proyectos software en ejecución

Gerardo Maturro^{1,a}, Karina Barrella^{1,b}, Patricia Benitez^{1,c}, María Inés Lund^{2,d}

¹ Departamento de Ingeniería de Software, Facultad de Ingeniería,
Universidad ORT Uruguay, Montevideo, Uruguay

² Instituto de Informática, Facultad de Ciencias Exactas, Físicas y Naturales,
Universidad Nacional de San Juan, San Juan, Argentina

^amaturro@uni.ort.edu.uy, ^bkbarrella@gmail.com, ^cpatobenitez@gmail.com,

^dmlund@iinfo.unsj.edu.ar

Abstract. No es poco frecuente que, en los proyectos software, sea necesario incorporar nuevos desarrolladores en una etapa avanzada de su ejecución. En estas circunstancias, estos “recién llegados” enfrentan varias dificultades y desafíos que les impiden comenzar rápidamente a contribuir, con sus conocimientos y experiencia previos, a la marcha del proyecto. Este artículo reporta los resultados de un estudio exploratorio-descriptivo dirigido a identificar las dificultades a las que se enfrentan los nuevos miembros del equipo de proyecto al unirse a un proyecto en ejecución, así como identificar las acciones que usualmente se adoptan para mitigar estos problemas y dificultades. El estudio revela que la escasa o nula documentación y la necesidad de conocer el producto en construcción son las principales dificultades, mientras que la asignación de un referente y la provisión de capacitación se mencionan como las principales acciones que las organizaciones suelen tomar para mitigar esos problemas.

Keywords: newcomers, recién llegados, proyectos software, dificultades

1 Introducción

Según Fairley, un factor particular que distingue los proyectos de software de otros tipos de proyectos es que los proyectos de desarrollo de software son esfuerzos orientados al equipo e intensivos en conocimientos [1]. Para este autor, los equipos son necesarios porque tomaría demasiado tiempo para que una persona sola desarrolle un sistema de software moderno y porque es improbable que una única persona posea el rango necesario de habilidades requeridas para desarrollarlo.

Cuando un desarrollador se incorpora a un proyecto de desarrollo de software en ejecución, se enfrenta a una situación en la que debe familiarizarse con diversos aspectos técnicos, tecnológicos, organizacionales y sociales relativos al proyecto en sí y al equipo que ya está participando en la ejecución de ese proyecto [2].

Estos “recién llegados” enfrentan diversas dificultades y desafíos para encontrar su lugar en el proyecto y poder comenzar a contribuir rápidamente, con sus conocimientos y experiencia previos, a la marcha del proyecto.

En este sentido, Rastogi y colegas opinan que, si bien las empresas esperan contribuciones tempranas al proyecto, los nuevos miembros a menudo requieren varias semanas para alcanzar el mismo nivel de productividad que los empleados existentes [3].

Agregar gente a un equipo de desarrollo de software no suele aumentar la productividad global de manera lineal porque, entre otros factores, el aumento de la sobrecarga de comunicación y coordinación de las actividades de trabajo entre las personas añadidas disminuye la productividad del equipo existente [1]. Es bien conocida la ley de Brooks relativa a este aspecto, que establece, de forma un tanto simplificada que, cuando se incorpora personal a un proyecto atrasado, éste se atrasa más [4].

A pesar de la importancia de esta cuestión, poco se sabe acerca de las barreras que enfrentan los recién llegados cuando participan en un proyecto, un proceso que, según Wolff-Marting y colegas, todavía presenta muchos problemas abiertos [5]

El propósito de este artículo es reportar el proceso seguido y los resultados obtenidos de un estudio exploratorio-descriptivo orientado a: a) determinar los problemas y dificultades a las que suelen enfrentarse los nuevos miembros de un equipo de proyecto cuando se incorporan a un proyecto que ya está en ejecución, y b) conocer las acciones y estrategias a niveles individual, de equipo y organizacional que se suelen adoptar para mitigar esos problemas y dificultades.

El resto de este artículo está organizado de la siguiente manera. En la sección 2 se hace un repaso de los principales conceptos y definiciones relativos a equipos de trabajo, normas y reglas de comportamiento, y se exponen algunos trabajos relacionados con el tema de investigación. La sección 3 explica el diseño metodológico del estudio, donde se exponen las preguntas de investigación y los procedimientos de recolección y preparación de datos. En la sección 4 se presentan los principales resultados obtenidos, mientras que en la sección 5 se expone el análisis de estos resultados. La sección 6 aborda las amenazas a la validez de este estudio y en la sección 7 se exponen las conclusiones y se esbozan los trabajos futuros.

2 Marco teórico y trabajos relacionados

Uhl-Bien y colegas definen *equipo* como un grupo de personas reunidas para usar sus habilidades complementarias para lograr un propósito común para el cual son colectivamente responsables [6].

El éxito del equipo depende de la cooperación de todos los miembros. Humphrey y colegas sostienen que los equipos suelen ser necesarios para las tareas que implican más trabajo, la disponibilidad de una variedad de habilidades, o alguna otra capacidad que una única persona no podría proporcionar por sí sola [7].

Usualmente, los equipos se conforman en la etapa inicial de los proyectos que deberán ejecutar. Sin embargo, y por diversas circunstancias, puede ocurrir la necesidad de incorporar nuevos miembros a un equipo ya formado, y en una etapa eventualmente avanzada de ejecución del proyecto.

Definimos entonces el concepto de “recién llegado” (*newcomer*) como el de la persona que se incorpora a un equipo de proyecto en un momento posterior al de la conformación del equipo inicial.

De acuerdo con George y Jones, la capacidad de un grupo para controlar los comportamientos de sus miembros depende de la medida en que los recién llegados aprendan los roles, reglas y normas del grupo. Los recién llegados no saben inicialmente lo que se espera de ellos y lo que pueden y no pueden hacer [8].

Diversos trabajos han abordado el estudio de los problemas y dificultades que enfrentan los recién llegados a un proyecto software en ejecución, pero poniendo el foco particularmente en proyectos de tipo *open source*.

Este es el caso, por ejemplo, de Steinmacher y colegas que llevaron a cabo una revisión sistemática de literatura [9]. En su estudio, que implicó la revisión de 21 trabajos seleccionados, identificaron una serie de barreras que los autores clasificaron en cinco categorías que han denominado: Interacciones sociales, Encontrar la manera de comenzar, Problemas de documentación, Cuestiones de código, y Conocimientos previos del recién llegado.

Otros estudios sobre la misma temática, enfocados a proyectos *open source*, reportados por este autor con los mismos u otros colegas, son: [10], [11] y [12]. Este último, en particular, aborda también los proyectos software del tipo *crowdsourcing*.

Panichella y colegas también abordan el estudio de los recién llegados a proyectos software, y lo hacen también enfocados en proyectos *open source*, en: [13] y [14].

En una línea similar, Fagerholm y colegas reportan estudios sobre lo que denominan “proceso de integración” (*onboarding*) de nuevos miembros, también relativos a proyectos del tipo *open source*: [15] y [16].

Por otra parte, Dagenais y colegas [2], así como Begel and Simon [17] presentan estudios sobre el proceso de integración de los recién llegados a proyectos software, pero con un enfoque exclusivo en entornos industriales (no *open source*) de producción de software.

En este sentido, el estudio que se reporta en este trabajo se enfoca también en el ámbito industrial del desarrollo de software.

3 Diseño metodológico

El diseño metodológico consistió en la definición de las preguntas de investigación, la selección de los informantes claves, la elaboración de los instrumentos de recolección de datos, su aplicación en entrevistas personales, la lectura crítica y codificación de las respuestas obtenidas, y el análisis y la interpretación de los datos obtenidos.

3.1 Preguntas de investigación

Para este estudio se plantearon las siguientes preguntas de investigación:

- RQ1: ¿Cuáles son los problemas y dificultades que suelen encontrar los recién llegados al incorporarse a un proyecto de desarrollo de software en ejecución?

- RQ2: ¿Qué acciones y estrategias se implementan a niveles de equipo y organizacional para mitigar esos problemas y dificultades?

3.2 Selección de los informantes claves

Para contactar informantes claves que pudieran dar su visión sobre los problemas de los recién llegados y sobre las estrategias usadas para mitigarlos, se contactaron 16 empresas de desarrollo de software basadas en Montevideo, de las cuales nueve accedieron a participar en el estudio. De estas nueve, cinco mencionaron que gestionan sus proyectos aplicando algunas prácticas ágiles (sin especificar detalles), mientras que las otras cuatro lo hacen siguiendo una metodología más tradicional.

Para recopilar datos para este estudio se realizaron entrevistas a doce líderes de equipos o jefes de proyecto y a ocho miembros de equipos pertenecientes a esas nueve empresas participantes.

En relación con los líderes de equipos, el criterio de selección para entrevistarlos fue que durante su dirección de algún proyecto haya ocurrido la incorporación de un nuevo integrante al equipo con posterioridad al inicio del proyecto. Estos líderes de equipo son profesionales con entre 4 y 17 años de experiencia general en desarrollo de software, y con antecedentes de haber liderado equipos de proyecto en al menos tres oportunidades.

Respecto a los miembros de equipo, el criterio de selección fue que, en alguna oportunidad, hayan vivido la experiencia personal de haberse incorporado a un equipo de proyecto en una instancia posterior al inicio de su ejecución. Todos los miembros de equipo entrevistados indicaron que desempeñaron el rol de programador, y que tenían poca o nula experiencia profesional de desarrollo de software en el momento en que se incorporaron al equipo de proyecto. De los ocho entrevistados, seis están cursando una carrera universitaria, y los otros dos son graduados de una carrera terciaria no universitaria.

3.3 Instrumentos de recolección de datos

Para la realización de las entrevistas se diseñaron dos cuestionarios: uno para líderes de equipo o jefes de proyectos, y otro para miembros de equipos. Ambos cuestionarios contienen preguntas cerradas y preguntas abiertas, y su propósito fue el de guiar a los autores en la conducción de las entrevistas y asegurar la recolección de los datos necesarios para responder a las preguntas de investigación.

3.4 Lectura y codificación de respuestas

Para la codificación de las respuestas a las preguntas abiertas se siguió el procedimiento propuesto en [18]; esto es, las preguntas abiertas se codifican una vez que se conocen todas las respuestas de los participantes a los cuales se les aplicaron, o al menos las principales tendencias de respuestas en una muestra de los cuestionarios

aplicados. Con esta codificación se obtienen ciertas categorías que representan los resultados finales.

El procedimiento consistió, entonces, en identificar estos patrones generales de respuesta (respuestas similares o comunes), listarlos, y asignarles un nombre o valor textual. De este modo, cada patrón constituye una categoría particular de respuestas.

4 Resultados obtenidos

En esta sección se presentan los principales resultados obtenidos a partir de las entrevistas, y luego de realizada la codificación y categorización de respuestas.

De las varias preguntas incluidas en los cuestionarios, se presentan aquí las directamente relacionadas con las preguntas de investigación formuladas en 3.1.

4.1 Respuestas obtenidas de los Miembros de equipo

Para los miembros de equipo que en algún momento fueron “recién llegados” a un proyecto software en ejecución, la pregunta específica relativa a la RQ1 fue: ¿Cuáles fueron las dificultades a las que se enfrentó al incorporarse al proyecto?

La Tabla 1 lista las categorías de dificultades identificadas en sus respuestas, y las veces (entrevistados) que fueron mencionadas.

Tabla 1. Dificultades mencionadas por los miembros de equipo y frecuencia de menciones

Dificultades	Frecuencia
Documentación	3
Entender la solución / Conocer el producto	3
Comunicación	1
Metodología de trabajo	1
Tecnología / Herramientas utilizadas	1

Los significados de las categorías de dificultades mencionadas son los siguientes:

- **Documentación:** Refiere a la poca o nula documentación del código fuente ya desarrollado, ya sea en forma de comentarios en el código o en forma de “apuntes”. También refiere a la falta de documentación sobre los requerimientos software iniciales y de los cambios ocurridos desde el comienzo de proyecto.
- **Entender la solución / Conocer el producto:** Refiere a tener una visión global del sistema en desarrollo, de los problemas de negocio que se pretenden resolver con el nuevo sistema, de las generalidades de la solución propuesta y de la organización y estado actual general del sistema que se está construyendo.
- **Comunicación:** Refiere a la poca e incompleta comunicación de parte del resto del equipo para explicar los cambios, inconvenientes y decisiones de proyecto que se realizaron durante el período previo a su incorporación al equipo.

- **Metodología de trabajo:** Refiere principalmente a conocer y familiarizarse con el proceso software en uso por parte del equipo, con la metodología de gestión del proyecto, con el ciclo de vida de desarrollo software que se está aplicando, así como con las rutinas de trabajo y de reuniones, y con las formas y criterios para asignar y delegar tareas.
- **Tecnología / Herramientas utilizadas:** Refiere a conocer y aprender a usar las herramientas software que se están utilizando en el desarrollo, y también con las diversas tecnologías de implementación de la solución, particularmente las relativas a la arquitectura software.

4.2. Respuestas obtenidas de los Líderes de equipo o Jefes de proyecto

A continuación, se presentan las preguntas del cuestionario utilizado para los líderes de equipo o jefes de proyecto relacionadas específicamente con las preguntas de investigación, y el resumen de las respuestas obtenidas.

Pregunta: ¿Qué dificultades suelen enfrentar los nuevos miembros de un equipo cuando se incorporan a un proyecto ya en ejecución?

La Tabla 2 lista las categorías de dificultades de los recién llegados desde el punto de vista de los líderes de equipo o jefes de proyecto, y las veces mencionadas.

Tabla 2. Dificultades mencionadas por los Team Leaders y frecuencias de mención

Dificultades	Frecuencia
Entender la solución / Conocer el producto	9
Metodología de trabajo	6
Conocer la empresa	5
Integración al grupo de trabajo	5
Entender el proyecto	3
Documentación	3
Tecnología / Herramientas utilizadas	3

Los significados de las categorías de dificultades mencionadas son los siguientes:

- **Entender la solución / Conocer el producto:** Refiere a tener una visión panorámica del sistema en desarrollo, de los problemas de negocio que se pretenden resolver, de las generalidades de la solución propuesta y de la organización y estado actual del sistema que se está construyendo.
- **Metodología de trabajo:** Refiere principalmente a conocer y familiarizarse con el proceso software en uso por parte del equipo, con la metodología de gestión del proyecto, con el ciclo de vida que se está aplicando, así como con las rutinas de trabajo y de reuniones, y con las formas y criterios para asignar y delegar tareas.
- **Conocer la empresa:** Refiere a que cuando el “recién llegado” no solo es nuevo en el equipo sino también en la empresa (recién contratado), siente la carencia de saber sobre la propia empresa, su estructura y cultura organizacionales, su misión, los tipos de clientes o de mercado a los que dirige sus operaciones o servicios de desarrollo de software.

- **Integración al equipo:** Refiere a los pasos iniciales en el equipo, saber “quien es quien”, a quién preguntar o pedir ayuda o consejo, relacionarse con los demás miembros del equipo, conocer sus roles y sus características personales y técnicas.
- **Entender el proyecto:** Refiere a conocer, a grandes rasgos, de qué se trata el proyecto, cuáles son sus objetivos, tiempo planificado de implementación, grado de avance actual, principales funcionalidades y “urgencias” de desarrollo.
- **Documentación:** Refiere a la poca o nula documentación del código fuente ya desarrollado, ya sea en forma de comentarios en el código o en forma de “apuntes”. También refiere a la falta de documentación sobre los requerimientos software iniciales y de los cambios ocurridos desde el comienzo de proyecto.
- **Tecnología / Herramientas utilizadas:** Refiere a conocer y aprender a usar las herramientas software que se están utilizando en el desarrollo, y también con las diversas tecnologías de implementación de la solución, particularmente las relativas a la arquitectura software.

Pregunta: ¿Qué acciones o mecanismos ha implementado para mitigar estas dificultades?

La Tabla 3 lista las categorías de acciones o mecanismos que implementan los líderes de equipo o jefes de proyecto para lidiar con las dificultades de los recién llegados, y las veces (entrevistados) que fueron mencionadas.

Tabla 3. Acciones implementadas por los Team Leaders y frecuencias de mención

Acciones	Frecuencia
Asignar un referente	10
Supervisión y seguimiento	6
Capacitación	5
Documentación	5
Apoyo a la persona	1
Asignar funcionalidades simples	1
Darles libertad	1
Definir un plan de integración	1
Programación de a pares	1

Los significados de las categorías de acciones implementadas mencionadas son los siguientes:

- **Asignar un referente:** Refiere a asignar a un miembro experiente del equipo a la labor de guía y *coach* del recién llegado, para orientarlo en los aspectos técnicos del proyecto y también de integración al equipo.
- **Supervisión y seguimiento:** Refiere a hacer un seguimiento continuo de las tareas asignadas al recién llegado, con espacios para consultar dudas.
- **Capacitación:** Refiere a brindar capacitación relativa al rol que ha de desempeñar el recién llegado, y también sobre los aspectos generales del proyecto.
- **Documentación:** Refiere a darle acceso a la documentación que pueda existir sobre el proyecto, y asignarle un tiempo para que lea, se interiorice en sus detalles técnicos y también formule preguntas.

- **Apoyo a la persona:** Refiere a brindar apoyo emocional para motivar y guiar a la persona, y para facilitar su integración personal y técnica al equipo.
- **Asignar funcionalidades simples:** Refiere a asignar al recién llegado tareas de proyecto simples que le permitan ir ganando confianza en su rol, y también empezar a conocer el producto en desarrollo.
- **Darle libertades:** Refiere a permitir y alentar a que el recién llegado opine, y proponga cambios y puntos de vista personales que lo haga sentir cómodo.
- **Plan de integración para cada uno:** Refiere a establecer un plan personalizado de asignación gradual de tareas y responsabilidades.
- **Programación de a pares:** Aplicar la técnica en la que se programa en parejas.

5 Análisis de los resultados

En esta sección se presenta el análisis de los datos recolectados, y se responde a las preguntas de investigación planteadas al comienzo.

5.1 Dificultades de los recién llegados

La pregunta de investigación RQ1 establecida al comienzo refiere a cuáles son los problemas y dificultades que suelen encontrar las personas que se incorporan a un proyecto software que ya está en marcha.

La Tabla 4 resume las categorías de dificultades mencionadas por líderes de equipo y por los propios “recién llegados”. Las dificultades percibidas en común son: a) escasa o nula documentación, principalmente del código desarrollado hasta el momento, b) desconocimiento de las características y funcionalidades principales del producto en desarrollo, c) necesidad de conocer y aprender la metodología de trabajo que utiliza el equipo al que se incorpora, y d) falta de conocimiento o experiencia en la tecnología de implementación o en el uso de las herramientas software en uso.

Tabla 4. Comparativo de dificultades percibidas por Líderes de equipo y miembros de equipo

Dificultades	Líderes de equipo	Miembro de equipo
Comunicación	---	X
Conocer la empresa	X	---
Documentación	X	X
Entender la solución / Conocer el producto	X	X
Entender el proyecto	X	---
Integración al equipo	X	---
Metodología de trabajo	X	X
Tecnología / Herramientas utilizadas	X	X

5.2 Acciones implementadas para mitigar las dificultades

La pregunta de investigación RQ2 refiere a conocer, desde el punto de vista de líderes de equipo o jefes de proyecto, cuáles son las acciones y mecanismos que suelen implementar para reducir o eliminar las dificultades y problemas que enfrentan los “recién llegados” a un proyecto en marcha.

De las diversas acciones mostradas en la Tabla 3, la asignación de un integrante del equipo como referente y guía del nuevo miembro es la opción más usual, a la que le siguen la realización de un seguimiento y supervisión de las tareas asignadas al nuevo integrante, y la provisión de capacitación y de documentación (si existe) sobre el proyecto y el producto en construcción.

6 Amenazas a la validez

Las amenazas más relevantes tienen que ver principalmente con los informantes claves seleccionados. Dado su escaso número, no constituyen una muestra representativa del total de profesionales que en algún momento de sus vidas fueron “recién llegados” a un proyecto. Esta población es, incluso, difícil de cuantificar. Por otra parte, los miembros de equipo entrevistados tenían escasa o nula experiencia en el momento de haber sido “recién llegado”, lo cual dejó fuera a profesionales que vivieron esa situación en una instancia avanzada de sus vidas profesionales, es decir, los que podemos denominar “recién llegados con experiencia profesional previa”.

7 Conclusiones y trabajos futuros

El propósito de este trabajo fue reportar los primeros resultados de un estudio orientado a identificar las dificultades y problemas iniciales que enfrentan los “recién llegados” a un proyecto software que ya está en ejecución, y también conocer las acciones que se suelen adoptar para mitigar esas dificultades iniciales.

Del análisis anterior de las respuestas obtenidas de los miembros de equipo entrevistados, los problemas más destacados refieren a la falta de documentación del producto en desarrollo, y el desconocimiento de la metodología de trabajo y de las tecnologías y herramientas de desarrollo utilizadas por el equipo.

Como forma de mitigar estos problemas, los jefes de equipo mencionan como la principal estrategia la asignación de un referente que acompañe al recién llegado en sus primeros pasos en el proyecto. Asimismo, mencionan que se les brinda capacitación y se les provee de la documentación que pueda existir.

Como trabajos futuros se plantea profundizar este estudio, particularmente sobre las siguientes interrogantes: a) ¿qué tipo de documentación y sobre qué aspectos del proyecto o producto los recién llegados consideran que deberían disponer? b) ¿cuán efectivas son las estrategias adoptadas por los jefes de proyecto para reducir o eliminar las dificultades de los nuevos integrantes de un equipo de proyecto software en ejecución?, c) ¿qué características personales y técnicas debería poseer la persona que se asigne como referente para acompañar al recién llegado a un proyecto?

Referencias

- 1 R. Fairley, *Managing and leading software projects*. Hoboken, NJ: John Wiley & Sons, 2009.
- 2 B. Dagenaisy, H. Ossherz, R. K. E. Bellamy, M. P. Robillard, and J. P. de Vries, Moving into a New Software Project Landscape, 32nd International Conference on Software Engineering (ICSE 2010), Cape Town, 2010.
- 3 A. Rastogi, S. Thummalapenta, T. Zimmermann, N. Nagappan, and J. Czerwonka, Ramp-up Journey of New Hires: Do strategic practices of software companies influence productivity?, in *ISEC 2017: Innovations in Software Engineering Conference*, Jaipur, India, 2017.
- 4 F. Brooks, *The Mythical Man-Month. Essays on Software Engineering. Anniversary edition*. Boston, MA: Addison Wesley, 1995.
- 5 V. Wolff-Marting, C. Hannebauer, and V. Gruhn, Patterns for tearing down contribution barriers to FLOSS projects, 12th Intl. Conf. on Intelligent Software Methodologies, Tools and Techniques (SoMeT 2013), Budapest, 2013, pp. 9–14.
- 6 M. Uhl-Bien, J. R. Schermerhorn, and R. Osborn, *Organizational behavior*, 13th ed. Hoboken, NJ: John Wiley & Sons, 2014.
- 7 W. S. Humphrey, T. Chick, W. R. Nichols, and M. Pomeroy-Huff, *Team Software Process (TSP) Body of Knowledge (BOK), SEI Technical report CMU/SEI-2010-TR-020*. Hoboken, NJ: John Wiley & Sons, 2010.
- 8 J. M. George and G. Jones, *Understanding and managing organizational behavior*, 6th ed. Upper Saddle River, NJ: Pearson, 2010.
- 9 I. Steinmacher, M. A. Graciotto Silva, and M. A. Gerosa, Barriers Faced by Newcomers to Open Source Projects: A Systematic Review, 10th IFIP WG 2.13 International Conference on Open Source Systems, OSS 2014, San José, Costa Rica, 2014.
- 10 I. Steinmacher, T. Uchoa Conte, Ch. Treude, and M. A. Gerosa, Overcoming Open Source Project Entry Barriers with a Portal for Newcomers, 38th IEEE International Conference on Software Engineering (ICSE 2016), Austin, TX, 2016.
- 11 I. Steinmacher, I. Scaliante Wiese, and M. A. Gerosa, Recommending Mentors to Software Project Newcomers, 3rd. International Workshop on Recommendation Systems for Software Engineering (RSSE 2012), Zurich, 2012.
- 12 A. Lazaretti, I. Steinmacher, L. Santos Machado, C. R.B. de Souza, and R. Prikladnicki, Barriers Faced by Newcomers to Software-Crowdsourcing Projects, *IEEE Softw.*, vol. 34, no. 2, pp. 37–43, 2017.
- 13 G. Canfora, M. Di Penta, R. Oliveto, and S. Panichella, Who is going to mentor newcomers in open source projects?, 20th International Symposium on the Foundations of Software Engineering, Cary, North Carolina, 2012.
- 14 S. Panichella, Supporting Newcomers in Software Development Projects, 31st International Conference on Software Maintenance and Evolution (ICSME2015), Bremen, 2015.
- 15 F. Fagerholm, P. Johnson, A. Sánchez Guinea, J. Borenstein, and J. Münch, Onboarding in Open Source Software Projects: A Preliminary Analysis, 8th International Conference on Global Software Engineering Workshops (ICGSEW 2013), Bari, 2013.
- 16 F. Fagerholm, A. Sánchez Guinea, J. Borenstein, and J. Münch, Onboarding in Open Source Projects, *IEEE Softw.*, vol. 31, no. 6, pp. 54–61, 2014.
- 17 A. Begel and B. Simon, Novice Software Developers, All over Again, ICER '08: 4th Intl. Workshop on Computing Education Research, Sydney, 2008.
- 18 P. H. Sampieri, R. Fernández, and C. Baptista, *Metodología de la investigación*, 5th ed. México: McGraw Hill, 2010.

Diseño de Aplicaciones Móviles basadas en Posicionamiento: un Framework Conceptual

Cecilia Challiol^{1,2}, Alejandra B. Lliteras^{1,3} and Silvia E. Gordillo^{1,3}

¹ UNLP, Facultad de Informática, LIFIA. Calle 50 y 120, La Plata, Argentina.

² CONICET, Argentina

³ CICPBA, Buenos Aires, Argentina

{ceciliac, lliteras, gordillo}@lifia.info.unlp.edu.ar

Resumen. Este trabajo presenta un framework conceptual para el diseño de Aplicaciones Móviles basadas en Posicionamiento. Se usa el concepto de separación de concerns para identificar aquellas características relevantes de este tipo de aplicaciones. Esta separación, facilita la adaptación y evolución de las mismas. Los conceptos del framework propuesto pueden ser útiles cuando se deben diseñar no solo nuevas aplicaciones sino también herramientas de autor para este tipo de aplicaciones. Se desea que este trabajo sirva para crear un espacio de discusión de aquellas características que podrían cambiar en este tipo de aplicaciones, y de cómo poder anticiparse en una etapa temprana de diseño a estos cambios. Se discutirán algunas características relacionadas a este tipo de aplicaciones.

Palabras Claves: Aplicaciones Móviles basadas en Posicionamiento; Framework Conceptual; Diseño; Separación de Concerns.

1 Introducción

Actualmente es indudable la masividad que tienen los dispositivos móviles, generando que se incrementen día a día las aplicaciones para los mismos. Al contar con mecanismos de sensado incorporados (por ejemplo, GPS) se puede crear gran variedad de aplicaciones, en particular aquellas que se basan en la posición del usuario; estas aplicaciones son conocidas como basadas en posicionamiento y cuentan con la particularidad de brindar, a partir de la posición actual del usuario, alguna información o servicio. A medida que el usuario se mueve por el espacio físico, estos servicios van variando. En este tipo de aplicaciones, la posición es un contexto relevante [1].

Las Aplicaciones Móviles basadas en Posicionamiento pueden cubrir diferentes dominios, por ejemplo, en [2] se mencionan: tour guiados, aplicaciones o herramientas educativas, ficciones sensibles a la posición o juegos sensibles a la posición. Otro tipo de dominio podría ser el de las experiencias urbanas vivenciales [3]. Cada dominio requiere que se diseñen características propias del mismo, por ejemplo, información del tour o actividades educativas. Sin embargo, todas estas aplicaciones tienen características en común relacionadas al posicionamiento, como

pueden ser, por ejemplo, el sensado de la posición, mostrar la posición en un mapa (lo cual implica representar de alguna manera el espacio físico), brindar información (o servicios) al llegar a una posición. Es decir, el diseño de este tipo de aplicaciones implica no solo considerar las características propias de cada dominio, sino también todo lo relacionado con el posicionamiento. La complejidad del diseño aumenta si se consideran además otros contextos [4].

El diseño de cualquier aplicación debe pensarse para soportar adaptación y evolución, como se menciona en [5]. Para estos autores, la adaptación consiste en poder agregar nuevos requerimientos pero manteniendo el objetivo original para el que fue creada la aplicación (por ejemplo, mediante nuevos servicios). Mientras que la evolución consiste en realizar cambios modificando el objetivo original para el que fue creada la aplicación (por ejemplo, que una aplicación móvil de tour guiado agregue características de juego). En el caso de las Aplicaciones Móviles basadas en Posicionamiento, por la gran demanda del mercado, estas están en constante cambio, no solo se agregan nuevos requerimientos (adaptación) sino que también, muchas veces, cambia el objetivo original (evolución). Con lo cual es fundamental considerar esto en una etapa temprana de diseño.

Existen metodologías para abordar diferentes características de las aplicaciones sensibles al contexto [6], incluyendo a las aplicaciones móviles basadas en posicionamiento. Sin embargo no hay una solución unificada o sistemática para diseñar este tipo de aplicaciones como se menciona en [4] y [6]. Es decir, esto aún es un área abierta de investigación.

Muchas veces, por desconocimiento, los desarrolladores pasan directamente a la implementación sin considerar la etapa de diseño. Por ejemplo, si un desarrollador quiere usar la API de posicionamiento de Android cuenta con documentación de consulta [7] como así también ejemplos de código [8]. Sin embargo, estos ejemplos solo se focalizan en mostrar cómo usar la funcionalidad de la API, y tienen acoplado generalmente todo el comportamiento de la aplicación en una sola clase. Si los desarrolladores siguen solo estos ejemplos sin tener en cuenta buenas prácticas de diseño, es muy posible que su aplicación no pueda adaptarse fácilmente a nuevos requerimientos y que sea muy compleja la evolución de la misma. Esta es una de las motivaciones de este trabajo, poder brindar un framework conceptual que sirva no solo en la etapa de diseño, sino también como guía para que los desarrolladores puedan identificar características relevantes y puntos de variabilidad en las Aplicaciones Móviles basadas en Posicionamiento; y que a esto lo puedan considerar cuando implementan este tipo de aplicaciones. Es decir, que los puntos de variabilidad se implementen de manera desacoplada, facilitando así la evolución de las aplicaciones desarrolladas.

El objetivo de este trabajo es presentar un framework conceptual para el diseño de Aplicaciones Móviles basadas en Posicionamiento. Este framework surge a partir del trabajo que hemos venido haciendo en el área (por ejemplo: [9], [10], [11] y [12]). Usamos el concepto de separación de concerns para identificar aquellas características relevantes de este tipo de aplicaciones, esto facilita luego la adaptación y evolución de las mismas. Esperamos que los conceptos del framework conceptual propuesto puedan ser útiles no solo a la hora de diseñar nuevas aplicaciones sino también al diseñar herramientas de autor para este tipo de aplicaciones.

Se desea que este trabajo sirva para crear un espacio de discusión de aquellas características que podrían evolucionar en este tipo de aplicaciones, y cómo poder anticiparse a esto en una etapa temprana de diseño.

Este trabajo se estructura de la siguiente manera. En la Sección 2 se presentan algunos trabajos relacionados. En la Sección 3 se describe un ejemplo motivador para ayudar al lector a identificar las características de evolución que podrían tener las Aplicaciones Móviles basadas en Posicionamiento. En la Sección 4 se presenta el framework conceptual propuesto para diseñar Aplicaciones Móviles basadas en Posicionamiento. Además, se discuten algunas características relevantes a considerar cuando se está diseñando este tipo de aplicaciones. Las conclusiones y algunos trabajos futuros son descriptos en la Sección 5.

2 Trabajos Relacionados

Las Aplicaciones Móviles basadas en Posicionamiento, son un subconjunto de las aplicaciones sensibles al contexto [6], donde la posición es un contexto relevante. Estas posiciones se vuelven lugares con significado [3]. Estos lugares se pueden estructurar de diferentes maneras dentro de las aplicaciones, por ejemplo, linealmente [2]. Cada forma de estructurar o tipo de estructura, determina como se va a comportar luego la aplicación móvil, por ejemplo, haciendo que el usuario recorra los lugares de uno en uno, siguiendo un orden preestablecido. En [2] se identifican diferentes estructuras pero no se contempla una separación entre el contenido de la aplicación y las posiciones. Este acoplamiento no permite que se reúsen ambos conceptos (contenidos y posiciones) en diferentes estructuras, en la Sección 4 se brindan más detalles de cómo desacoplar estos conceptos.

En [1] se presenta una taxonomía de guías móviles sensibles al contexto, donde se especifica que el contexto más usado por las aplicaciones relevadas, es la posición. Esto se debe generalmente a la disponibilidad de APIs para obtener la posición del usuario. En [6] se presenta un resumen de las características involucradas en las aplicaciones sensibles al contexto, se detallan distintas características relacionadas con la implementación de este tipo de aplicaciones, lo que podría considerarse como un complemento al framework conceptual presentado en este trabajo.

En [4] se propone un proceso de diseño de sistemas inteligentes (o sensibles al contexto) que tiene como objetivo apoyar a los diseñadores de sistemas en su tarea. Los autores consideran todos los pasos de la ingeniería de software, desde el análisis del estado del arte hasta la implementación. Dado que los pasos son descriptos de forma muy general, nuestro framework conceptual podría ser útil para poder abordar con más profundidad los pasos que, en [4], se denominan “*functional requirements*” y “*fine-turning*”.

3 Motivación

En esta sección se presenta un ejemplo para ayudar a los lectores a entender algunos de los diferentes potenciales que podrían surgir relacionados a la evolución de las

Aplicaciones Móviles basadas en Posicionamiento, los cuales deberían ser considerados en una etapa temprana de diseño.

Supongamos que se desean diseñar Aplicaciones Móviles basadas en Posicionamiento para ser usadas dentro de un museo (espacio físico). Puede ser que la necesidad de contar con algunas de estas aplicaciones surja a lo largo del tiempo y en diferentes momentos (por ejemplo, a partir de requerimientos de los usuarios), por eso se considera importante prever la posibilidad de brindar nuevas aplicaciones a ser usadas en el mismo espacio físico, anticipándolo en la etapa de diseño.

Podríamos suponer, que la primera aplicación que una persona se imagina, relacionada a un museo, es aquella que brinda información relacionada con cada elemento del museo. En este caso, cada elemento del museo podría tener, por ejemplo, un código QR, y cuando el usuario lee el mismo con su dispositivo móvil, recibe la información de ese elemento (un ejemplo de este tipo de funcionalidad es presentado en [13]).

Luego, el museo podría querer proporcionar Aplicaciones Educativas Móviles basadas en Posicionamiento, dado que se reciben muchas visitas de escuelas. En este caso, se plantean actividades educativas asociadas a diferentes elementos del museo. Podrían surgir diferentes aplicaciones educativas móviles según los perfiles de los alumnos, planteando cada una de ellas diferentes actividades educativas. Estas aplicaciones podrían, por ejemplo, brindar preguntas cuando los alumnos leen con sus dispositivos móviles los códigos QR provistos por el museo. Cabe destacar que en el caso de diseñar aplicaciones educativas móviles basadas en posicionamiento la participación de los docentes es fundamental [14].

Pero también, el museo podría querer proveer Juegos Móviles basados en Posicionamiento, donde se involucren elementos del museo. Por ejemplo, que los chicos que visitan el lugar, al leer un código QR (de un elemento del museo) puedan identificar que nombre tenía ese animal en la película “La era de hielo”.

Además, el museo podría querer proporcionar una Aplicación Móvil para asistir en la movilidad a personas no videntes que visiten el lugar. Este tipo de aplicaciones requiere más nivel de precisión en la representación del espacio físico del lugar [15]. Esta última aplicación podría integrarse con cualquiera de las otras mencionadas anteriormente.

De esta manera se puede apreciar cómo pueden ir surgiendo diferentes aplicaciones móviles dentro del museo, cada una de ellas con contenidos distintos pero todas relacionadas a las posiciones de los elementos del museo. Es decir, que estas posiciones del museo (posiciones relevantes) se podrían reutilizar, esto es una característica a considerar en la etapa de diseño. Para esto, las posiciones relevantes deben ser especificadas sin acoplarse con características de los dominios [10]. Más aun, los mecanismos usados para determinar la posición actual del usuario, también se podrían compartir entre las distintas aplicaciones. En el caso de los ejemplos mencionados anteriormente se reutilizaban los códigos QR. En la etapa de diseño se deberían plantear los mecanismos de sensado de posiciones de forma genérica, con el fin de desacoplar la manera en la que se obtiene una posición, de las características propias de los dominios de las aplicaciones en las cuales serán usados.

Supongamos ahora, que en el museo ya se cuenta con las aplicaciones descritas anteriormente, y que con el auge de los *Beacons* [16], el museo decide incorporarlos como su nuevo mecanismo de sensado, evitando que los usuarios tengan que leer

códigos QR empleando un dispositivo móvil. Para que esto impacte lo menos posible en las aplicaciones existentes, las posiciones se deberían de haber especificado de manera independiente del mecanismo de sensado usado para obtener el valor de la posición. De ser así, cambiar el mecanismo de sensado o combinarlos, impactaría mínimamente en las aplicaciones existentes. Esta es otra de las características que se puede considerar al diseñar las posiciones, previendo que los mecanismos de sensado de posiciones pueden variar en el tiempo [9]. Para esto, las posiciones relevantes se deberían definir de manera independiente al mecanismo de sensado; y adicionalmente tener una forma de "interpretar" el valor obtenido con el mecanismo de sensado para identificar con cual posición relevante se corresponde.

La representación del espacio físico, que el usuario generalmente visualiza en su dispositivo móvil como un mapa, también se podría reusar entre las distintas aplicaciones. El museo podría tener su mapa indoor ad-hoc, sin embargo este también podría evolucionar en el tiempo y pasar a estar digitalizado como un mapa indoor de Google [17]. Si en la etapa de diseño se establece una interfaz común a cualquier representación del espacio físico [18], el impacto de usar una nueva representación sería mínimo.

En la Fig. 1 se muestra cómo desacoplar para poder reusar tanto la representación del espacio físico como las posiciones relevantes. Por simplicidad, se muestra en detalle como una misma posición relevante podría combinarse en varias aplicaciones móviles, en este caso son una aplicación educativa, un tour y un juego. Se podrían reusar todas las posiciones relevantes o un subconjunto de ellas. El contenido de cada aplicación está relacionado con lo que cada una debe definir, como puede ser, para este caso: actividades educativas, información de los elementos del museo o información del juego respectivamente.

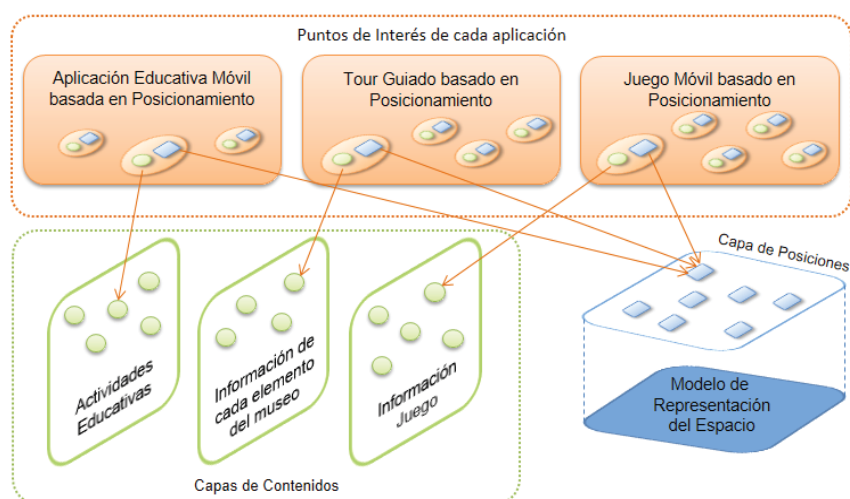


Fig. 1. Aplicaciones Móviles basadas en Posicionamiento dentro de un museo.

En la Fig. 1 se puede apreciar que los puntos de interés (PoI) definidos para cada aplicación no se relacionan entre sí, esto es por simplicidad del ejemplo. Estos PoI podrían estar estructurados de diferente manera, algunos ejemplos de estructuras se mencionan en [2]. Cabe mencionar que en [2] no se identifica el concepto de PoI, ni la separación entre contenido y posiciones, solo se detallan los tipos de estructuras. Para el diseño, estas estructuras podrían estar planteadas en forma genérica y luego instanciarse con diferentes PoI. Especificar los PoI independientes de la estructura que los contenga permite que los mismos puedan ser reusados en más de una estructura. Más aun, considerando que cada uno de los tipos de estructuras se pueden definir para que se comporten siempre igual independientemente de los PoI que contengan. Por ejemplo, a una estructura lineal el usuario siempre la recorre de la misma manera independientemente del contenido que se muestra en cada lugar. Si estas estructuras se diseñan genéricas se pueden reusar en diferentes aplicaciones.

En la Fig. 1, se pueden apreciar contenidos para tres aplicaciones particulares, sin embargo estos contenidos podrían ser reusados en otras aplicaciones. Por ejemplo, si se considera a una actividad educativa como un contenido, ésta podría reusarse en varias aplicaciones educativas móviles.

De esta manera, se pudo apreciar como las aplicaciones pueden ir cambiando en el tiempo o bien pueden ir surgiendo nuevas aplicaciones. Contar con una interfaz común para las representaciones del espacio, definir los mecanismos de sensado de posiciones genéricos como así también las capas de posiciones (independiente de los dominios y de los mecanismos de sensado), agiliza el diseño de nuevas Aplicaciones Móviles basadas en Posicionamiento. Y además permite prever puntos de variabilidad en una etapa temprana, como es el diseño, facilitando tanto la adaptación y evolución de las aplicaciones.

4 Framework conceptual para el diseño de Aplicaciones Móviles basadas en Posicionamiento

La definición de nuestro framework conceptual surge a partir de diferentes trabajos que venimos realizando en el área. Por ejemplo, en [9] presentamos un modelo de solución para la variabilidad de las características de contexto en aplicaciones sensibles al contexto. Hemos combinado una aplicación móvil basada en posicionamiento con actuaciones en vivo de actores [10]. En [11] definimos una herramienta de creación in-situ de experiencias móviles basadas en posicionamiento, la cual extendimos para el dominio educativo [12]. Estos trabajos nos han permitido explorar y aprender distintas características relacionadas al diseño de este tipo de aplicaciones.

En la Fig. 2 se puede apreciar el framework conceptual propuesto, en el cual usamos el concepto de separación de concerns (capas) para desacoplar cada una de las características que potencialmente podría tener variabilidad. Se puede apreciar que se desacoplan tanto los mecanismos de sensado como la representación del espacio.

Realizar una registración genérica a mecanismos de sensado permite estar recibiendo notificaciones del cambio de posición, sin importar de cual mecanismos de sensado se obtuvo dicho valor. Para esto, además cada mecanismos de sensado debe

estar definido genérico (en [9] se puede profundizar en una solución de modelado para abordar esta característica). Cada vez que se sensa un nuevo valor (por alguno de los mecanismos de sensado), las aplicaciones registradas reciben dicho valor, y le dan una interpretación acorde a las posiciones relevantes que tienen definidas.

Desacoplar la representación del espacio, permite que ésta pueda ir variando en el tiempo, ésto se presentó en el ejemplo motivador donde se contaba con una representación del espacio ad-hoc, la cual como se mencionó podría ser reemplazada por un mapa indoor de Google. Al diseñar desacoplado, y plantear una interfaz de acceso común, reemplazar una representación del espacio por otra, sería transparente para la aplicación.

También se puede apreciar en la Fig. 2 que se plantean por separado, en diferentes capas, los aspectos de contenido, posiciones y PoI, esto permite el reuso de cada aspecto en distintas Aplicaciones Móviles basadas en Posicionamiento. Cabe mencionar que cada capa de posiciones relevantes debe ser especificada de forma genérica sin estar acoplada a la representación del espacio, esto permite que cualquier cambio en la representación no impacte en los PoI ya definidos por la aplicación.

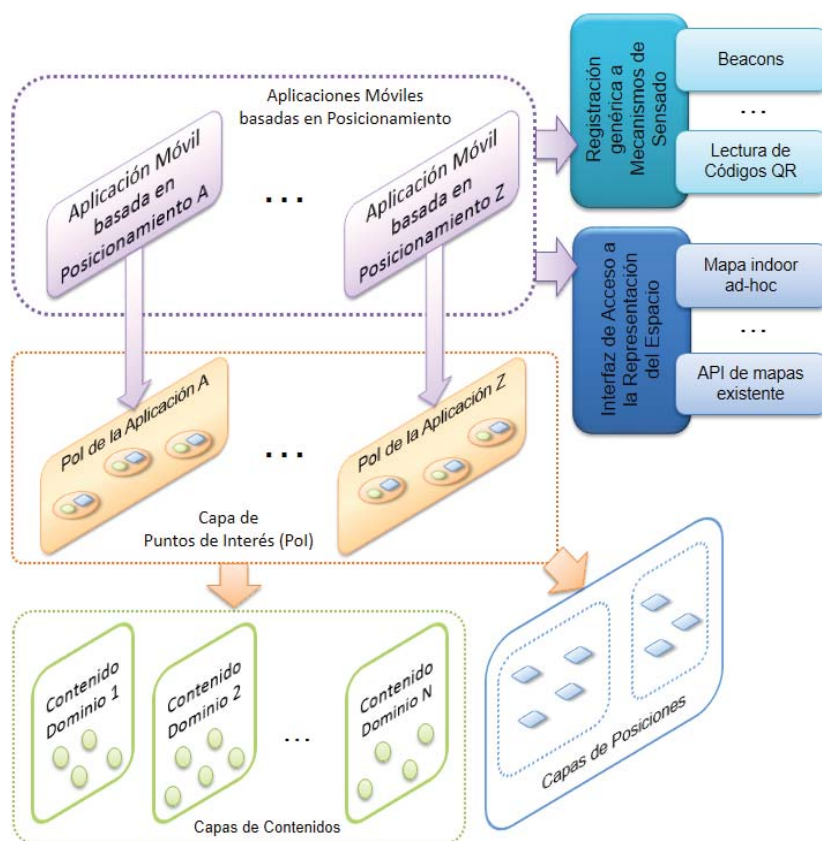


Fig. 2. Framework conceptual propuesto para diseñar Aplicaciones Móviles basadas en Posicionamiento.

Podría considerarse otra capa desacoplada que agrupe los PoI según diferentes estructuras, más detalles de posibles estructuras se puede encontrar en [2]; tener en cuenta que, en el trabajo citado, estas estructuras no se plantean desacoplando contenidos de posiciones algo que se puede complementar con nuestro framework. Para que estas estructuras sean reutilizables deberían plantearse genéricas, y ser instanciadas con los PoI de la aplicación, como se mencionó en la Sección 3.

De esta manera se presentó el framework conceptual propuesto para diseñar Aplicaciones Móviles basadas en Posicionamiento. Donde se pueden apreciar los puntos de variabilidad respecto de los mecanismos de sensado y la representación del espacio. Además, se presenta una separación en capas del contenido, posiciones y PoI fomentando el reuso de los mismos, y de esta manera, agilizar los futuros diseños.

4.1 Discusión

En esta subsección se presentan algunas discusiones relacionadas al framework conceptual propuesto para aportar así otras consideraciones que se deben tener en cuenta cuando se diseñan Aplicaciones Móviles basadas en Posicionamiento.

Cabe mencionar que el framework conceptual propuesto es un punto de partida para poder seguir explorando que consideraciones hay que tener en cuenta para lograr que las Aplicaciones Móviles basadas en Posicionamiento sean adaptables y puedan evolucionar de una forma sencilla.

Notar que los conceptos presentados en el framework conceptual podrían ser utilizados para llevarlos a la práctica en cualquier entorno de desarrollo, ya sea nativo, web móvil o híbrido.

En [19] se menciona que el diseño de las Aplicaciones Móviles basadas en Posicionamiento podría ser in-situ. Esto requiere que se visite el lugar físico, y definir acorde a esa visita que contenidos brindar y cuáles podrían ser posiciones relevantes dentro de ese espacio físico. Por ejemplo, cuanto se tarda estimativamente en ir caminando de un lugar a otro, esto puede afectar que decisiones de diseño se toman. Crear in-situ Aplicaciones Móviles basadas en Posicionamiento es un nuevo campo y requiere más investigación como se menciona en [20]. Usar como guía un framework conceptual, como el propuesto en este trabajo, permitiría que todo lo que se especifica in-situ sea luego reusable y pueda evolucionar.

Muchas veces el diseño de este tipo de aplicaciones requiere de la participación de un equipo multidisciplinario, el cual podría incluir a los usuarios finales en este proceso de diseño [21]. En [22] se explora la creación de narrativas basadas en posicionamiento, usando técnicas de co-diseño. Hay que considerar que las personas involucradas en el diseño pueden requerir de un entrenamiento para comprender las características de este tipo de aplicaciones [23].

En [24] se analizan ciertas características que debería tener una herramienta de autor para crear narrativas basadas en posicionamiento. Los autores, hacen un análisis a nivel conceptual con el objetivo de sentar las bases para la construcción de una herramienta de autor que están desarrollando.

En [11] y [12] se realiza una exploración inicial de las características que deberían tener las herramientas de autor para creación in-situ de Aplicaciones Móviles basadas en Posicionamiento. Lo presentado en [11] y [12] sigue los lineamientos de

separación de concerns presentados en el framework conceptual propuesto. Acorde a lo aprendido a partir de estas herramientas definidas, podemos decir que al definir herramientas de este estilo, se deberían tener en cuenta las características que tienen los productos finales generados por las herramientas. Sería deseable que cambiar de mecanismo de sensado o de representación del espacio fuera fácil para el usuario final que usa dichas herramientas. También sería deseable que estas herramientas permitieran reusar contenidos, posiciones o PoI, previamente definidos.

5 Conclusiones y Trabajos Futuros

En este trabajo se presentó un framework conceptual para el diseño de Aplicaciones Móviles basadas en Posicionamiento. Se usa el concepto de separación de concerns para identificar aquellas características relevantes de este tipo de aplicaciones, pudiendo detectar así los puntos de variabilidad en una etapa temprana de diseño. Se describieron distintas consideraciones que podrían ser útiles no solo para diseñadores sino también para desarrolladores de este tipo de aplicaciones.

Deseamos que este trabajo ayude a crear un espacio de discusión sobre cómo diseñar Aplicaciones Móviles basadas en Posicionamiento. Los conceptos que se describen en este trabajo podrían utilizarse no solo para crear aplicaciones sino también herramientas de autor para este tipo de aplicaciones. Más aun, estas herramientas de autor podrían generar aplicaciones basadas en el framework conceptual propuesto. En [10] se pueden ver más detalles de una aplicación concreta que usa separación de concerns entre posiciones y contenido. En [11] o [12] se pueden ver más detalles de herramientas de autor concretas construidas también sobre la base de la separación de concerns.

Seguimos trabajando en extender la herramienta de autor presentada en [12], para luego poder realizar a futuro pruebas con docentes. También se realizarán a futuro co-diseño, por ejemplo, con docentes. A partir de este aprendizaje se espera poder seguir aprendiendo y ajustando las características que deben considerarse en un framework conceptual considerando además co-diseñar con un equipo multidisciplinario.

Referencias

1. Emmanouilidis, C., Koutsiamanis, R.A., Tasidou, A.: Mobile guides: Taxonomy of architectures, context awareness, technologies and applications. *Journal of Network and Computer Applications* 36 (1), 103-125 (2013)
2. Hargood, C., Hunt, V., Weal, M.J., Millard, D.E.: Patterns of sculptural hypertext in location based narratives. In: 27th ACM Conference on Hypertext and Social Media, pp. 61-70. ACM, New York (2016)
3. Stals, S., Smyth, M., IJsselsteijn, W.: Walking & talking: probing the urban lived experience. In: 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational, pp. 737-746. ACM, New York (2014)
4. Bauer, C., Dey, A.K.: Considering context in the design of intelligent systems: Current practices and suggestions for improvement. *Journal of Systems and Software* 112, 26-47 (2016)

5. Weyns, D., Caporuscio, M., Vogel, B., Kurti, A.: Design for Sustainability = Runtime Adaptation \cup Evolution. In: the 2015 European Conference on Software Architecture Workshops, pp. 62-69. ACM, New York (2015)
6. Alegre, U., Augusto, J.C., Clark, T.: Engineering context-aware systems and applications: A survey. *Journal of Systems and Software* 117, 55-83 (2016)
7. Location-Aware en Android, <https://developer.android.com/training/location/index.html>
8. Ejemplos de Google Play Location, <https://github.com/googlesamples/android-play-location>
9. Fortier, A., Rossi, G., Gordillo, S.E., Challiol, C.: Dealing with variability in context-aware mobile software. *Journal of Systems and Software* 83 (6), 915-936 (2010)
10. Alconada Verzini, F.M., Tonelli, J.I., Cecilia, C., Llitas, A.B., Gordillo, S.E.: Combing Location-Aware Applications with in-situ Actors Performances. In: the 2015 Workshop on Narrative & Hypertext, pp. 27-31. ACM, New York (2015)
11. Alconada Verzini, F.M., Tonelli, J.I., Cecilia, C., Llitas, A.B., Gordillo, S.E.: Authoring Tool for Location-Aware Experiences. In: the 2015 Workshop on Narrative & Hypertext, pp. 21-25. ACM, New York (2015)
12. Zimbello, A.M., Alconada Verzini, F.M., Challiol, C., Llitas, A.B., Gordillo, S.E.: Authoring for location-based learning experiences. In: 4th International Conference on Mobile Software Engineering and Systems, pp. 211-212. IEEE Press (2017)
13. Pérez-Sanagustín, M., Parra, D., Verdugo, R., García-Galleguillos, G., Nussbaum, M.: Using QR codes to increase user engagement in museum-like spaces. *Computers in Human Behavior* 60, 73-85 (2016)
14. Gilman, E., Milara, I.S., Cortés, M., Rieki, J.: Towards user support in ubiquitous learning systems. *IEEE Transactions on Learning Technologies* 8(1), 55-68 (2015)
15. Li, B., Munoz, J.P., Rong, X., Xiao, J., Tian, Y., Arditi, A.: ISANA: wearable context-aware indoor assistive navigation with obstacle avoidance for the blind. In: European Conference on Computer Vision, pp. 448-462. Springer International Publishing (2016)
16. Beacons Estimote, <https://estimote.com>
17. Maps indoor de Google, <https://www.google.com/maps/about/partners/indoormaps>
18. Llitas, A.B., Challiol, C., Mostaccio, C.A., Gordillo, S.E.: Representaciones enriquecidas para la navegación indoor-outdoor en aplicaciones móviles. In: XVII Congreso Argentino de Ciencias de la Computación, pp. 867-876 (2011)
19. Santos, P., Hernández-Leo, D., Blat, J.: To be or not to be in situ outdoors, and other implications for design and implementation, in geolocated mobile learning. *Pervasive and Mobile Computing* 14, 17-30 (2014)
20. Brundell, P., Koleva, B., Wetzell, R.: Supporting the design of location-based experiences by creative individuals. In: 11th International Workshop on Semantic and Social Media Adaptation and Personalization, pp. 112-116. IEEE Press (2016)
21. Sanders, E.B.N., Stappers, P.J.: Co-creation and the new landscapes of design. *Co-design* 4 (1), 5-18 (2008)
22. Millard, D.E., Hargood, C.: A Research Framework for Engineering Location-Based Poetics. In: the 2015 Workshop on Narrative & Hypertext, pp. 13-16. ACM, New York (2015)
23. Fidas, C., Sintoris, C., Yiannoutsou, N., Avouris, N.: A survey on tools for end user authoring of mobile applications for cultural heritage. In: 6th International Conference on Information, Intelligence, Systems and Applications, pp. 1-5. IEEE Press (2015)
24. Millard, D.E., Hargood, C.: Location Location Location: Experiences of Authoring an Interactive Location-Based Narrative. In: 9th International Conference on Interactive Digital Storytelling, pp. 419-422. Springer International Publishing (2016)

Early aspects in “Aspect-Oriented Process for a Smooth Transition”

Fernando Pinciroli¹, José L. Barros-Justo²

¹ Research Institute of the Faculty of Informatics and Design, Champagnat University,
Mendoza, Argentina

pincirolifernando@uch.edu.ar

² School of Informatics (E.S.E.I.), University of Vigo, Ourense, Spain

jbarros@uvigo.es

Abstract. The aspect-oriented paradigm brought new expectations about building software in a modular way and improving many quality attributes, but it also came with new challenges. One of them is the lack of casuistry of its use in the industry. Software development project leaders are already facing enough risks and should not add more, such as incorporating an immature approach. Furthermore, there are no aspect-oriented proposals using standard notations and covering the full Software Development Life Cycle (SDLC). We have elaborated an alternative called Aspect-Oriented Process for a Smooth Transition (AOP4ST), that allows the smoothly incorporation of the aspect-oriented paradigm in the current industrial projects and offers a complete homogenous proposal for the phases of the SDLC. In this paper, we present the first stages of AOP4ST, seeking the incorporation of the aspect-oriented paradigm in the industry with the least impact, but trying to take full advantage of the possibilities it offers.

Keywords: business process modeling; requirement modeling; aspect-oriented software development; crosscutting concerns; UML; BPMN; AOP4ST.

1 Introduction

The emergence of the aspect-oriented paradigm brought new expectations about the possibility of building software in a more modular way and improving its quality attributes, such as maintainability, flexibility, comprehensibility, reusability, etc.

However, the new paradigm also came with new challenges, since there is not enough casuistry of its use in the industry guiding towards a way to apply it properly, particularly with regard to techniques, tools, notations and good practices.

Software development projects must deal with a large number of risks. It is not advisable to add new ones, such as incorporating an approach that is not sufficiently mature and that requires training people in poorly known tools and techniques, applying not well-tested methods, lacking support from vendors, and many more.

Similarly, there are no methodological proposals employing standard notations and covering the full SDLC, so in case of applying the aspect-oriented paradigm, we are

obliged to compose a method by picking up parts from different authors, who worked each phase of the SDLC in isolation [1].

In this context, we decided to move forward with an alternative that allows us to incorporate the aspect-oriented paradigm in the current projects in the industry, and offering a complete proposal by unifying homogeneously the different phases of the SDLC.

In this paper, we present the models of AOP4ST corresponding to the early stages of the SDLC. AOP4ST is a framework process for software development whose aim is the incorporation of the aspect-oriented paradigm in a smooth way, causing the least possible negative impact in the industry, but trying to make the most with all of the advantages that current aspect-oriented paradigm offers.

In section 2 we briefly present the AOP4ST's schema. The subsequent sections explain the different parts of the process: in Section 3 we offer a description of the business model and Section 4 describes the user requirement model. Next, in Section 5, we present the three views that make up the system requirement model: functional, static and state views. Finally, Section 6 presents the conclusions, and highlights some open issues and future work.

2 About AOP4ST

AOP4ST is a framework process, it is not a method nor a methodology. It covers the whole SDLC, but we are presenting here its structure for the early phases, commonly known as “early aspects” [2][3].

AOP4ST's name highlights two main concepts: a) the AOP, “Aspect-oriented Process”, indicates that it is truly aspect-oriented, ensuring that can be reached the widely known benefits of this paradigm; b) the 4ST, “for a Smooth Transition”, points to the possibility of applying this process in the industry immediately, because it employs widespread techniques, notations, standards, tools, etc. and allows to move to an aspect-oriented reality, taking advantage of the current state of the paradigm, until their own tools, techniques, etc. were imposed and completely accepted on the market.

The problem that AOP4ST tries to solve is how to bring the benefits of the aspect-oriented paradigm to the whole SDLC at the same time that are being used techniques, tools and standards currently widespread in the industry. In addition, the use of well-known techniques and tools allows to incorporate this paradigm gradually, until the different existing proposals have sufficient diffusion and maturity to warrant their employment in real and complex projects.

The whole SDLC include the business model, not considered for the authors offering aspect-oriented approaches for the early stages of the SDLC. There are few incomplete proposals about aspect-oriented business modeling [4].

Our approach arises from several factors, which we have to face in the industry and in the adoption of new technologies for software development.

First, it is well known that the different software development paradigms initially appear in the programming phase and then continue their definitions upstream, along the SDLC [5][6]. The aspect-oriented paradigm follows the same pattern, that is why

we can find more proposals for the programming phase than for the early phases of the SDLC.

Second, many proposals about software development sound promising and offer benefits difficult to refuse, but their massive use in industry depends on many factors. A well-known case is that of the object-oriented databases, that beyond the benefits they offered and the enormous popularity of the object-oriented languages and development tools today, they have not at all achieved the leading role in industry that could be expected [7].

Finally, software development projects have to deal with many risks, and the main function of project leaders is to minimize the damage that these risks can cause. The use of immature technologies, tools newcomers to the market, techniques that have not been tested enough, etc., would be very risky decisions to take by who has the responsibility to carry out a successful software development project. On the other hand, the availability of well-known tools and techniques and the adherence to standards and best practices will help professionals to make good estimates and to take better decisions.

AOP4ST is based on the hypothesis that it is possible to design an aspect-oriented software development process that encompass techniques, tools, notations and standards of widespread use in the current practice. This development process is suitable for the early stages of the SDLC, making the most with the benefits of the aspect-oriented paradigm in real-world settings. Under certain circumstances, it is possible to make use of existing techniques, tools and standard notations, right now, while specific theoretical and practical instruments are developed and introduced in the market, achieving enough dissemination and support to justify its use in real software development projects.

AOP4ST's basic structure for early aspects is composed of three models: business model, user requirements model and software requirements model. The last one is divided into three views: functional, static and state views. Concerns can be progressively discovered along these models and views.

3 The business model in AOP4ST

The business model starts the process in AOP4ST, since we seek to have it as the first layer of the enterprise architecture [8]. In an enterprise architecture, this layer has to pull on the rest of the models downstream in order to meet the business goals.

Besides this, an early concern detection can help to improve the software robustness. If concerns are detected, separated and encapsulated away from the elements of the problem domain, we may obtain:

1. Cleaner business models, since they are not tangled with unrelated issues to the problem domain.
2. Business models totally focused on the needs of the very specific business processes.
3. Crosscutting concerns modeling and encapsulating specific processes not belonging to the problem domain.

4. The possibility to focus each analyst on each party's specific goals: the functional perspective is modeled separately of the quality perspective for the business models.

Five main aspect-oriented activities must be done in AOP4ST's business model: concern detection, concern separation and modeling, composition rules checking, conflict resolution, aspect-oriented modeling.

3.1 Concern detection

The business model can be structured in several ways. A typical way is to divide it into primary processes, support processes and management processes [9], and besides to them, the reusable processes. The latter are processes that are not instantiated in themselves, but are done from any of the first kinds of processes and can be shared. Regardless of the type of process to which they belong, each process must be located within a specific package. These packages correspond to the concerns that are detected in this first model, and will host the concerns throughout the whole SDLC.

Reusable processes can be of two types: belonging to the domain of the problem and not belonging to it. In the first case, these are typical activities of the domain of the problem, that are repeated in several processes and which we, normally, could associate with functional requirements.

In the second case, these are activities that are independent of the problem domain and can be found even in different problem domains, e.g. access control, security, audit, logging, etc. They are, typically, quality attributes, also known as non-functional requirements (Figure 1), that are tangled with the activities belonging to the problem domain and scattered along all the processes.

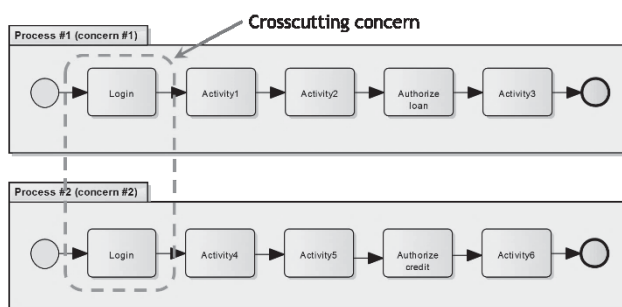


Fig. 1. Tangled and scattered crosscutting concerns.

Concern detection is done with the deliberate intention of doing so. In order to achieve a better success, we try to detect separately, the concerns corresponding to functional requirements and to non-functional ones. The former concerns are more difficult to be detected, because they depend on the wording of the modeler. The latter are simpler, because they are clearly distinguishable from the activities belonging to the problem domain and, besides, it is possible to have a list of standard categories of non-functional requirements to follow in a systematic way. Manual or automated aspect mining techniques can be used in order to detect concerns [10].

3.2 Concern separation and modeling

Concern separation is done by placing the detected ones in a zone of reusable processes in the model, so as to be able to form a catalog of reusable processes that is accessible by all modelers, in a kind of catalog of "reuse with" resources.

Concerns must be encapsulated into packages, along with the rest of the common elements of the model, and using a notation based on the proposal of Charfi et al. [11] although adjusted to use only elements belonging to the standard BPMN 2.0. This notation is also used to specify pointcuts, that are conditions that explicitly indicate join points. Join points are the points where the concerns will be composed again. In the base process, we use an annotation element to indicate the join point (Figure 2) and the separated concern is modeled inside of a pool element. This pool includes a "Proceed" activity, that represents the join points and that indicates if the composition must be done before, after or around the join point (Figure 3).

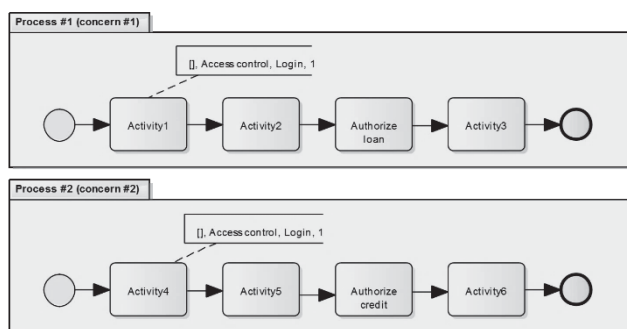


Fig. 2. Pointcut represented with annotation elements and indicating the join points.

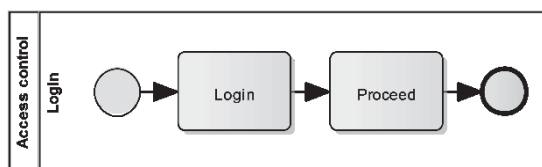


Fig. 3. Concern modeled and encapsulated, with a "Proceed" activity indicating the join points.

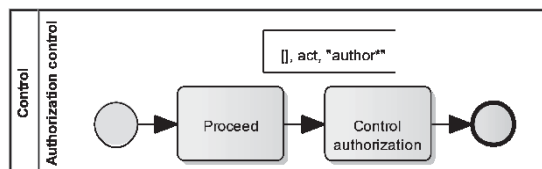


Fig. 4. Implicit concern composition.

We have incorporated a notation to indicate join points implicitly and to add concerns into the processes whenever a particular condition is met. The concern

presented in Figure 4 will be composed after all the activities called “Control authorization” in the model of Figure 2.

3.3 Composition rules checking and conflict resolution

When concerns are separated, we must be sure that they will be able to be re-composed correctly. The evolution of models over time could also produce new concerns that, at the time of composing, might not be adequately integrated.

To improve the concern composition, we have designed a set of composition rules that allows to reduce possible errors when integrating again the concerns into the processes [12]. We have analyzed the possible consequences of integrating the different types of concerns (“after”, “before”, and “around”) with the different elements of BPMN 2.0, and produced a set of recommendations that lead to a syntactically correct composition. This will also help to improve the semantics of the processes and to reduce potential conflicts.

3.4 Aspect-oriented modeling

Since base processes and concerns are modeled within packages, the composition relationships and the relationships among concerns can be represented with a “concern model”, build with UML package diagrams (Figure 5). In addition, it is possible to present a more detailed application of the concerns in the different join points by means of a “join point model”, where the packages are presented as "white boxes", showing the join points inside (Figure 6).

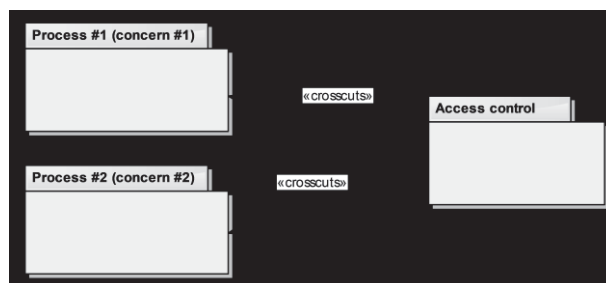


Fig. 5. Concern model.

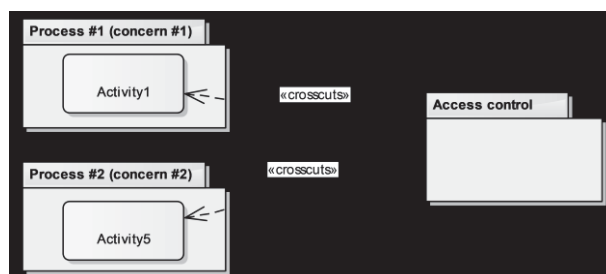


Fig. 6. Join point model.

4 The user requirements model in AOP4ST

The business model and the user requirements model crosscut the systems belonging to the organization. Each business process can describe activities that are supported by different systems. Similarly, the implementation of a user requirement could impact several systems, so this model of user requirements does not belong to a particular system but to the global solution.

In the business model, the processes were described placing them into specific packages that correspond to concerns. The same packages existent in the business model are re-created in this model of user requirements, so that the user requirements that are now detected are perfectly delimited to the process that requires them.

Functional user requirements will be easier to locate within a specific concern, while non-functional user requirements and business rules are more likely to be global, that is, to apply to several or even all concerns. They are usually referred to as crosscutting concerns. In the business model, several crosscutting concerns had already been detected when we modeled processes not belonging to the domain of the concerns, but in this model, will arise new crosscutting concerns (Figure 7). These concerns also have an important influence among them, due to the positive and negative contribution relationships [13].

In addition to modeling the user requirements in the corresponding packages, it will be necessary to specify the relationships among them, so that the aspect-oriented models mentioned in the previous section (concern and join point models) are used again, now including the relationships among user requirements.

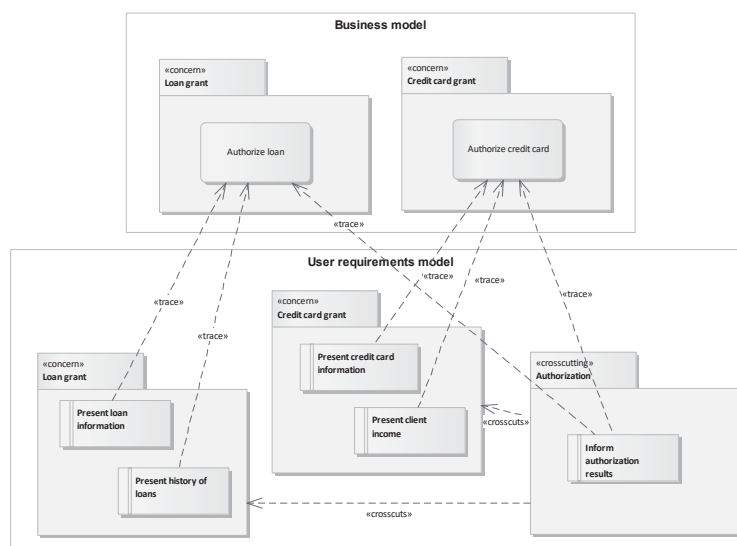


Fig. 7. Traceability between business and user requirements models.

5 The requirement model in AOP4ST

When implementing a user requirement, the solution could impact one or more computer systems. Thus, it will be necessary to divide this model of system requirements for each of the computer systems that are involved in the solution. We must develop a “system requirements model”, or simply “requirement model”, for each of the computer systems, and divide it into three views: functional, static and state views.

5.1 Requirement model: functional view

The functional view is elaborated using the use case technique, mainly based on the proposal of Jacobson and Ng [14], with some adjustments that we have found necessary, especially in the specification of the use cases, as well as in their models, in order to be able to automatically create test cases from the scenarios.

Again, we must create the same packages that exist in the previous model, which correspond to the concerns found so far, to place within them the use cases that we will find in this model. Since each requirement of the previous model might require an implementation in some of the computer systems of this model, we will only create the packages for the computer systems that will be impacted by the solution. If the solution of each user requirement could involve more than one use case, we will create a new package for each use case. It is important to keep in mind that each use case corresponds to a concern.

5.2 Requirement model: static view

In the static view of the requirements model we will maintain the division of computer systems and create a package for each of the packages in the use case view. In this way, we will build a logical class diagram for each use case by using UML.

Clarke y Baniassad [15] show how to elaborate these diagrams and how to compose the final classes later on. We have extended these models and incorporated practices that allow better traceability among models, for purposes of impact analysis for maintenance, regression testing, etc.

5.3 Requirement model: state view

With this third view, we consider that it will be possible to have a complete model of the solution. Since the state diagrams describe the life cycle of objects belonging to a single class, there is no need of creating the package structure again in this view, since

the state diagrams will be appended to their respective classes. We use the UML notation to build the state diagrams.

A state diagram presents the potential states of the objects of a class and their possible transitions. The state of an object is the abstraction of the values of its attributes and their relationships at a given time, any changes in the attributes and relations of a class in the static view can cause a strong change in its corresponding state diagram. We have included the “revised” tagged value for all the states that should change to “false” when a state could be affected for a change on the static view.

6 Conclusions

This article presents the main ideas about AOP4ST, a framework process for aspect-oriented software development, focusing on the early stages of the SDLC. With this approach, we expect to obtain a double gain: to reduce the negative impact of employing a paradigm that is not yet mature enough on real projects in the industry and, to take advantage of the benefits of the current state of the aspect-oriented paradigm.

The main virtue of this proposal lies in the use of widely disseminated standards in the industry, and with good practices, software tools, suppliers, human resources, etc. that are enough to allow to face a software development project in real-world settings.

Another very important outcome is the homogeneity and cohesion among the models, which allow for a coherent transition from one to another, enabling pre and post-requirement specification traceability and impact analysis.

The third outstanding feature is that concerns are emerging naturally and progressively throughout the models.

So far, we have been able to perform some theoretical and practical validations of AOP4ST. From the theoretical point of view, we have submitted it for consideration to a symposium of doctoral theses [16], from which we have received very rich feedback. We have also applied the criteria established by Jalali [4] for the measurement of AOP4ST's business model, and the results placed it in a privileged position.

Regarding the practical validation, we were able to test separately the AOP4ST's models in several companies. AOP4ST was fully applied to re-modeling a whole model of one of the most important biochemical laboratories in Argentina.

We believe that there is still a lot of effort needed to validate AOP4ST and to achieve a greater maturity in the architectural and test models. We are now working on these issues.

For Spanish-speaking readers, it is possible to find more information about AOP4ST in [16] and [17], and about AOP4ST's business model in [18] and [19].

References

1. Magableh, A., Shukur, Z. and Ali, N. M. "Systematic review on aspect-oriented UML modeling: A complete aspectual UML modeling framework," *J. Appl. Sci.*, vol. 13, no. 1, pp. 1–13 (2013).
2. Bakker, J., Tekinerdogan, B. and Aksit, M. "Characterization of Early Aspect Approaches," *Early Asp. Asp. Requir. Eng. Archit. Des. Work.*, p. 7 (2005).
3. Rashid, A., Moreira, A. and Tekinerdogan, B. "Early aspects: aspect-oriented requirements engineering and architecture design," *Software, IEE Proc.*, vol. 151, no. 4, pp. 153–155 (2004).
4. Jalali, A. "Assessing Aspect Oriented Approaches in Business Process Management," in *Perspectives in Business Informatics Research*, 13th International Conference, BIR 2014, pp. 231–245 (2014).
5. Capretz, L. F. "A brief history of the object-oriented approach," *SIGSOFT Softw. Eng. Notes*, vol. 28, no. 2, p. 6– (2003).
6. "History of Programming Languages," in *History of Programming Languages Conference*.
7. Leavitt, N. "Whatever Happened to Object-Oriented Databases?" *Computer (Long Beach, Calif.)*, vol. 33, no. 8, pp. 16–19 (2000).
8. Greefhorst, D. and Proper, E. *Architecture Principles. The Cornerstones of Enterprise Architecture*, vol. 6, no. 3 (2011).
9. *ABPMP, BPM CBOK V.3.0 - Business Process Management BPM Common Body of Knowledge* (2013).
10. Pincirolí, F. "Considerações acerca da mineração de aspectos," *Perspect. em Ciências Tecnológicas*, vol. 5, no. 5, pp. 83–101 (2016).
11. Charfi, A., Müller, H. and Mezini, M. "Aspect-oriented business process modeling with AO4BPMN," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 6138 LNCS, pp. 48–61 (2010).
12. Pincirolí, F. "Aspect-oriented business process composition rules in AOP4ST," in *35th International Conference of the Chilean Computer Science Society (SCCC 2016) held in conjunction with the 42th Latin American Computing Conference (CLEI 2016)*, pp. 1–6 (2016).
13. Pincirolí, F. "Improving software applications quality by considering the contribution relationship among quality attributes," *Procedia Comput. Sci. 3rd Int. Work. Comput. Antifragility Antifragile Eng. (ANTIFRAGILE 2016)*, vol. 83, pp. 970–975 (2016).
14. Jacobson, I. and Ng, P. *Aspect-oriented software development with use cases*. Addison-Wesley (2005).
15. Clarke, S. and Baniassad, E. *Aspect-oriented analysis and design. The Theme approach*. Boston: Addison-Wesley (2005).
16. Pincirolí, F. "Aspect-Oriented Process for a Smooth Transition," in *Ph.D. Symposium of the IEEE 11 Congreso Colombiano de Computacion* (2016).
17. Pincirolí, F. "AOP4ST – Aspect-Oriented Process for a Smooth Transition," in *WICC 2015 - XVII Workshop de Investigadores en Ciencias de la Computación* (2015).
18. Pincirolí, F. and Zeligueta, L. "El modelo de negocio en AOP4ST," in *WICC 2016 - XVIII Workshop de Investigadores en Ciencias de la Computación* (2016).
19. Pincirolí, F. and Zeligueta, L. "Modelado de negocios orientado a aspectos con AOP4ST," in *WICC 2017 - XIX Workshop de Investigadores en Ciencias de la Computación* (2017).

Elicitación de requerimientos de stakeholders distribuidos¹

Fernanda Cuadrado¹, Alejandro Oliveros²

¹NEC Argentina cuadrado.fernanda@gmail.com,²Universidad de Tres de Febrero, aoliveros@untref.edu.ar

Abstract. La Ingeniería de Requerimientos no ha prestado suficiente atención la elicitación de requerimientos con stakeholders distribuidos, las técnicas disponibles suponen el contacto directo, personal, con los stakeholders del dominio del problema. El caso de organizaciones ad-hoc constituidas por entidades teóricamente iguales, carentes de una superior jerárquico y distribuidos geográficamente, requiere disponer de enfoques para atacar esa realidad. En el este artículo se comunica un estudio de caso en el que se identificó lo fundamental del proceso actual de requerimientos aplicando *Grounded Theory* para analizar los datos obtenidos: de ellos mediante el uso de GT se logró que emergieran una serie de conceptos relevantes para formular el modelo de gestión de requerimientos de un caso de stakeholders distribuidos.

Keywords: Elicitación, Requerimientos distribuidos, Estudio de caso, Teoría Fundada

1 Introducción

La Ingeniería de Requerimientos dispone de numerosas indicaciones acerca de los métodos o metodologías para una correcta obtención y formulación de requerimientos. Sin embargo, existen escenarios que no son cubiertos por esas metodologías o que su utilización puede ser demasiado compleja o no efectiva. Tal es el caso de los Stakeholders distribuidos, que se ha propuesto mejorar la comunicación y la negociación mediante el uso de herramientas de computación que medien entre los stakeholders o hacer reuniones en momentos particulares de los proyectos.

El objetivo del presente *caso de estudio* es establecer la metodología de elicitación de requerimientos utilizada en proyectos de software orientados a un grupo de instituciones que trabajan en forma descentralizada. Se trata de entender qué proceso de recolección de requerimientos es efectivo para involucrar stakeholders de las distintas instituciones y cómo se instrumenta la validación de los requerimientos definidos ante un amplio universo de interesados. El caso en consideración es un

¹ El trabajo de campo de la investigación que se informa y el comienzo del análisis de la información se desarrolló en el INTEC/Facultad de Ciencias Exactas e Ingeniería de la Universidad Argentina de la Empresa (UADE) como parte del Proyecto de Investigación y Desarrollo (PID) P15T01,

ejemplo de esquemas organizacionales distribuidos sin constituir estrictamente una organización. En la Argentina se presenta con organizaciones estatales que se proponen un desarrollo de software en conjunto. Tal sería el caso de un proyecto que involucre a provincias, o a los municipios de una provincia.

Ejemplos en la Argentina serían proyectos que involucren a provincias o a municipios de una provincia en los que se toman algunas decisiones por consenso. Algunas características de este tipo de proyectos son: usuarios geográficamente distribuidos, igualdad de derechos de los organismos participantes, marco legal/normativo común (por ej. ley de contabilidad de la Nación). Con diferenciación de: funcionalidades dentro de las pautas comunes, culturas organizacionales/funcionales diferentes, madurez en la experiencia organizacional, el uso de tecnología, poder político, dinámica de adhesión al proyecto.

Las Universidades son entidades complejas, descentralizadas con múltiples stakeholders y que poseen autonomía pero deben seguir ciertas normas legales que regulan las actividades. Como caso se estudió la metodología del tratamiento de requerimientos por el SIU, una organización que desarrolla aplicaciones para la Universidades de la Argentina.

Las organizaciones se identifican en tres dimensiones básicas [1]: (a) la parte de la organización que desempeña el papel principal en el éxito o fracaso, (b) el método de coordinación de las actividades, (c) el grado con el que la organización involucra a los subordinados en el proceso de toma de decisiones. En el caso considerado el papel clave lo desempeña la organización de desarrollo.

La Ingeniería de Requerimientos (IR) se propone especificar el conjunto de requerimientos que satisfacen las necesidades y objetivos de los stakeholders del dominio del problema. El proceso de IR se despliega a lo largo de la *elicitación*, *especificación* y *validación* de los requerimientos [2]. Estas actividades requieren interacción de los desarrolladores con los *stakeholders* que se encuentran en el dominio del problema. Esta interacción es más estrecha y profunda en el momento de la *elicitación de los requerimientos*. En esta actividad los desarrolladores intercambian con los restantes *stakeholders* un volumen de información en una magnitud que no volverá a repetirse. La elicitación de requerimientos cuenta con una extensa serie de técnicas [3], que han sido clasificadas en diferentes formas [2], [4],[5].

Estos enfoques se han orientado a contextos en el que todos los stakeholders se encuentren en una única localización (*co-located*). Las entrevistas, focus groups, reuniones de grupo, etc. interactúan con los stakeholders que están accesibles porque comparten la localización entre ellos y con los desarrolladores. Las técnicas orientadas a entrevistas, por ejemplo, no mencionan ningún problema derivado de que el stakeholder se encuentra a distancia [2]. Una lista exhaustiva de técnicas de elicitación [4] menciona solamente una aplicable al caso de estar los stakeholders distribuidos (el *survey*). La única posibilidad de atacar la distribución de stakeholders consiste en utilizar en diferentes localizaciones una técnica desarrollada para uno solo: por ejemplo la observación participante [6] en diferentes localizaciones.

En la década de 1990 tomó un fuerte impulso el Desarrollo Global de Software (GSD), introduciendo nuevos desafíos al desarrollo de software en cuanto la comunicación y coordinación [7]. Inicialmente se propuso utilizar facilidades de desarrollo distribuidas en el mundo y luego se presentó la problemática de la

distribución de los stakeholders. Esto es, una organización, u organizaciones, distribuidas en diversas localizaciones y se trata de desarrollar software desde una o varias organizaciones de desarrollo [8]. En este contexto se han identificado una serie de desafíos: comunicación inadecuada, gestión de conocimiento, diversidad cultural y diferencias de tiempo.

Se han estudiado alternativas organizacionales que debe enfrentar en la *Ingeniería de Requerimientos Distribuida* [9] en el contexto de una gran multinacional. Considerando las localizaciones del Análisis, el Diseño de alto nivel y la Implementación (incluyendo diseño de bajo nivel), se identifican diferentes estructuras organizacionales. Estos diseños organizacionales generan una serie de problemas que han sido atacados exitosamente con varias técnicas. Este tipo de estructura presenta fuerte similitudes con nuestro caso, aunque nuestro objeto de estudio carece de una autoridad jerárquica única.

Se encaró la siguiente **pregunta de investigación**: ¿cuáles son las técnicas y enfoques utilizados para la elicitación de requerimientos en el caso de una asociación voluntaria de diferentes organizaciones con la meta de compartir el resultado de un desarrollo de software de utilidad para cada uno de los integrantes de la asociación?

2 Metodología de investigación

2.1 Caso de estudio

Para el desarrollo del *Caso de estudio* se siguieron las pautas establecidas en la ingeniería de software [10]. Este es un *estudio exploratorio* que se propone obtener un conocimiento para encarar investigaciones con hipótesis claramente formuladas. Es un estudio *cualitativo* en el que la posibilidad de realizar una adecuada *triangulación* de observadores se presentó mediante el trabajo de los dos investigadores y que con respecto a los datos se limitó el acceso a fuentes. Se trata de un estudio *inductivo*. Como metodología de análisis de los datos se utilizó *Grounded Theory*.

2.2 Grounded Theory

Para el análisis de los datos obtenidos se utilizó el enfoque de la Teoría Fundada (*Grounded Theory*, GT de sus siglas en inglés). La GT es la generación sistemática de teoría a partir de datos analizados mediante un método riguroso de investigación.

El proceso de GT [11] se puede resumir en los siguientes pasos (Figura 1):

1. Marcar *puntos clave* en cada documento: las frases que se consideran relevantes para ser consideradas
2. A cada punto clave se le asigna un *código* (frase corta), debe tenerse en cuenta:
 - El investigador debe mantener la mente libre de códigos preconcebidos
 - Cada código introducido se compara con los restantes, este es el método de comparación constante (*GT's constant comparison method-CCM*)
3. Los códigos se agrupan en *conceptos* (unidades de mayor nivel de abstracción)

- Los conceptos introducidos se comparan con los restantes del documento y los otros documentos (aplicando de nuevo el CCM)
 - Este agrupamiento de conceptos no se hace después de la asignación de códigos, se realiza al mismo tiempo que la codificación
4. Se introducen las *categorías* mediante el agrupamiento de los conceptos.
 5. Codificación selectiva (*Selective coding*). Se reitera CCM aplicada a un conjunto limitado de categorías: la categoría core y las estrechamente relacionadas con ella.

Hay dos codificaciones. La *codificación sustantiva* identifica los códigos teóricos que integraran la teoría sustantiva. Los códigos de categorías tienen un significado muy preciso. La *codificación teórica* es el modelo relacional a través del cual todos los códigos / categorías sustantivos están relacionados con la categoría principal.

Durante todo el proceso se producen *Memos* que registran el conocimiento que el investigador va acumulando y posteriormente son útiles para construir teoría.

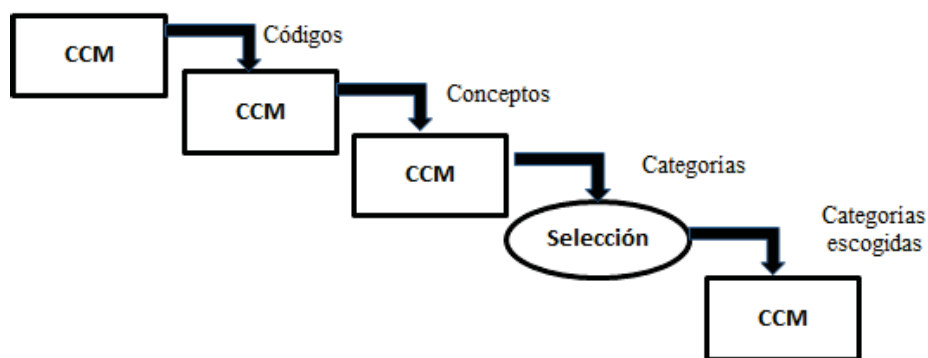


Fig. 1. Proceso GT

2.3 Diseño del Caso de Estudio

En lo que sigue se detalla el diseño del *Caso de estudio* siguiendo el enfoque de Runeson [10]. El *fundamento* de este estudio ha sido conocer las prácticas de elicitación de requerimientos en el contexto de una estructura organizacional y de proyecto poco tratada en la literatura aunque de importancia en ciertos esquemas (especialmente gubernamentales) de la Argentina. El *propósito* consistió en obtener claridad sobre el proceso actual de elicitación a fin de formular hipótesis de investigación precisas en el campo de la Ingeniería de Requerimientos Distribuida. El *caso* en consideración ha sido el servicio de desarrollo de software para Universidades y la *unidad de análisis* un proyecto específico. En este caso no se trató de indagar *relaciones causales*, por el carácter exploratorio del estudio. Las *datos* provienen de entrevistas en profundidad y el *análisis de datos* se hizo utilizando GT.

En el presente *Caso de estudio* se realizó una colección de información, la cual fue analizada utilizando la metodología de GT. A continuación se describe en detalle el proceso de investigación completo

3 Proceso de investigación.

3.1 Recolección de Datos

Para la recolección de la información principal se elaboró una guía de entrevistas personales. El cuestionario se concentró en recopilar información acerca de la ejecución de la elicitación de requerimientos y su gestión, herramientas tecnológicas que utilizaban (si utilizaban alguna) y quiénes participaban de la elicitación. Se indagó acerca del contexto de la organización, los clientes y el tipo de proyectos. El cuestionario fue probado con un desarrollador con experiencia en desarrollo y gestión de proyectos. Después de esta prueba se hicieron ajustes finales al instrumento,

La selección de los candidatos a entrevistar se basó en personas conocidas por los investigadores, que trabajaban en instituciones que desarrollaran aplicaciones para entidades descentralizadas, y que estén dispuestas a colaborar en esta investigación. Finalmente se seleccionó una persona, a la cual se le efectuaron dos extensas entrevistas (en total unas cuatro horas de entrevista y más de 26.000 palabras la desgrabación). La primera entrevista se enfocó en las consultas diseñadas previamente y solamente se pidieron aclaraciones a las respuestas. Previamente a la segunda entrevista se revisó cuidadosamente la primera y se consultó algunos temas que generaron dudas o se pidió más información.

Las entrevistas se grabaron, con consentimiento del entrevistado. La desgrabación la hizo un investigador y el otro la revisó contrastándola con la grabación original.

3.2 Análisis de Datos

A continuación se detallan los pasos dados para el análisis de la información obtenida.

Codificación.

En primer lugar se codificó utilizando palabras simples o frases que sugerían la idea a la que hacía referencia el fragmento que se asociaba al código. Por ejemplo: “Mercado es un código simple y “Cantidad de Recursos” es un código frase.

En segundo lugar, se realizaron varias iteraciones en las que se introdujeron nuevos códigos y modificaron algunos existentes porque no se ajustaban a lo que se refería el texto. Se iteró hasta no encontrar nuevos códigos

La tercera tarea comenzó a analizar los códigos obtenidos. Los 99 códigos sintácticamente distintos pero semánticamente coincidentes se unificaron en 18.

Ejemplos:

- Los códigos “Tipo Aplicación” y “Tipo de Sistema”, hacían referencia a lo mismo, y se decidió dejar “Tipo de Aplicación” para ambos.
- “Comunidad de Prácticas”, “Comunidad de Desarrollo” ó “Comité de Usuarios” también hacían referencia a lo mismo y se decidió dejar “Comité de Usuarios” por ser lo más utilizado generalmente en los procesos de desarrollo.
- “Validación Requerimientos”, “Validación de los solicitado vs. Lo desarrollado” y “Prototipo”, se unificaron en “Validación de Requerimientos”.

La cuarta actividad analizó a qué hacían referencias los códigos definitivos e identificó los tramos de la entrevista a que se referían. Se establecieron los códigos que hacían referencia al proceso de requerimientos. Estos códigos fueron extraídos para continuar con las siguientes etapas de análisis de los mismos.

La quinta etapa comenzó el análisis de las líneas de la entrevista que hacían referencia a los requerimientos, para ello se identificó la necesidad de cambiar la forma de expresar los códigos, mejorando las frases explicativas. Por ejemplo:

- el código “Priorización de Requerimientos”, luego del análisis se definió como “Los requerimientos se ingresan por medio de un texto y se agrega una prioridad”. Esta tarea de recodificación se realizó para todas las líneas que hacían referencia al tratamiento de los requerimientos.

Asociación de Conceptos y Categorías

Los conceptos, según la GT, son elementos que “surgen/emergen” a partir de algunos códigos y que luego se pueden asociar a códigos que hacen referencia a lo mismo.

- Por ejemplo: los códigos “Refinamiento de la Definición de un Requerimiento” y “Cierre de la Formulación de un Requerimiento” hacen referencia a “Definición de los Requerimientos”, es por ello que se asoció éste último como concepto de esos códigos.

Las categorías, que tienen la misma dinámica respecto a los conceptos que estos con los códigos, asocian conceptos que hagan referencias a una definición mayor,

- Por ejemplo: “Definición de los Requerimientos” y “Validación de los Requerimientos”, dos conceptos hallados, hacen referencias a “Desarrollo de los Requerimientos”, que definimos como una categoría.

A partir del relacionamiento de códigos, conceptos y categorías se armó un árbol de relaciones, en donde varios códigos hacen referencia a un mismo concepto y varios conceptos a una misma categoría. En la Tabla 1 se muestra un ejemplo.

Tabla 1. Relacionamiento códigos, conceptos y categorías

Códigos	Concepto	Categoría
Refinamiento de la definición del requerimientos	Definición de los requerimientos	Desarrollo de los requerimientos
Cierre de la formulación del requerimiento		

Memos y *Sorting*

Al comenzar a trabajar con los conceptos y categorías, se hizo necesario agregar los memos (pequeñas anotaciones de la razón de ser del código identificado, y por consiguiente del concepto y categoría).

- Por ejemplo: el código de “Validación de Requerimiento” se concluía a partir del siguiente memo “Cuando se tiene duda de entender bien el requerimiento se escribe, y lo envían al usuario para buscar su acuerdo “...*Te parece razonable...*”, en donde parte del memo es conclusión nuestra y parte textual del entrevistado. Ver Tabla 2.

Posteriormente los memos se agruparon, haciendo referencia a la parte de *sorting* de la aplicación de la GT y con el criterio de aquellas partes (códigos y conceptos) que hacían referencia al mismo tópico. Así todos los códigos y conceptos que hacían referencias a una misma categoría quedaron ubicados juntos.

Table 2. Uso de memos

Texto de la entrevista	Código	Concepto	Categoría
Un requerimiento pedido por un usuario se trata de generalizar para que aplique a varios. Cuando se tiene duda de entender bien el requerimiento se escribe, y lo envían al usuario para buscar su acuerdo «...Te parece razonable» En algunos casos se desarrolla y se envía al usuario para su validación «...entrá y fijate si este te sirve o no te sirve»	Validación de Requerimiento	Validación de Requerimiento	Desarrollo de los Requerimientos

Análisis de los resultados

El objetivo es analizar la relación entre los códigos, conceptos y categorías para poder obtener una conclusión del análisis llevado a cabo a partir de la información obtenida. El objetivo de ese análisis era poder determinar qué nos había mencionado el entrevistado en cuanto a su proceso de Requerimientos de las aplicaciones que desarrollan. El análisis de los resultados se detalla en la sesión siguiente.

4 Resultados

Luego de finalizado el análisis de la entrevista se concluyó que el tratamiento de los requerimientos que se lleva a cabo en el SIU posee un proceso de “Gestión de Requerimientos” bien definido. Gestión abarca las actividades comprendidas desde que se inicia la elicitación hasta la finalización del tratamiento del mismo. La finalización puede ser por la decisión de no implementar el requerimiento o por el desarrollo y aceptación por parte del usuario del mismo.

El proceso comienza con el registro de un requerimiento en forma de un texto y con una priorización dada por el usuario, ambos surgen del comité de usuarios o de pedidos individuales de las universidades. Ese registro es llevado a cabo en una herramienta, que se utiliza a lo largo del proceso de gestión. Esos requerimientos registrados luego son gestionados a través de una actividad de “homologación” por parte del grupo de desarrollo para decidir su tratamiento o no.

La decisión de tratarlos no tiene un proceso formalizado, aunque puede estimarse que depende el impacto y esfuerzo de desarrollo. Los requerimientos son clasificados en “grandes” y “chicos”, sin un criterio preciso de inclusión de los requerimientos en

una u otra categoría. Posiblemente la clasificación es función del impacto y esfuerzo de desarrollo que implique ese requerimiento. Los requerimientos “chicos” son priorizados por desarrollo e incluidos en una nueva versión en el “Delivery”. Los requerimientos “grandes” son enviados al comité de usuarios que analiza y se decide la “Prioridad” del momento del desarrollo de cada uno de ellos. El comité de usuarios está conformado por personas de las universidades nacionales, de diferentes áreas administrativas y académicas.

En todos los casos los requerimientos aceptados pasan por una actividad de “Definición” en la que se analiza lo que se debe construir, interactuando con el usuario que lo pidió, en caso de requerirse detalle sobre el mismo, y dejando registro de esta actividad en la herramienta de gestión de requerimientos. La actividad de definición finaliza cuando desarrollo entiende que ya el requerimiento se encuentra especificado como para pasar a ser desarrollado y planificado para una nueva versión de software a liberarse (“Delivery”). Como último paso del proceso se puede determinar que los requerimientos incluidos en una versión ya liberada de software son “Validados” por el usuario para determinar si lo desarrollado coincide con lo que fue solicitado, lo cual se puede expresarse en otras palabras diciendo si lo desarrollado aplica a las necesidades por las cuales fue pedido el requerimiento.

Si bien todo lo mencionado anteriormente se incluyó como el proceso de gestión de requerimientos, también se identificó un subproceso que denominamos “Desarrollo de Requerimientos” y que incluyen las actividades “Definición”, “Delivery” y “Validación” descriptas anteriormente. La consideración que el subproceso de Desarrollo de Requerimientos está contenido en el proceso de Gestión de Requerimiento está dada porque si bien es una actividad propia de desarrollo, tiene una gestión del requerimiento durante esas etapas de su ciclo de vida.

En la Figura 2 se representan los diferentes conceptos descriptos anteriormente, las relaciones entre ellos, y el agrupamiento en las categorías.

5 Conclusiones

En el presente caso de estudio se analizó el tratamiento de requerimientos en estructuras descentralizadas, para la recolección de datos se utilizó una herramienta diseñada especialmente para tales fines. El análisis de datos fue realizado por medio de la aplicación de la GT. A partir del refinamiento de la codificación y posterior clasificación en categorías y conceptos se pudo establecer el proceso de administración de requerimientos descripto por el entrevistado.

Se pudo constatar que una metodología a utilizar en entidades descentralizadas con diferentes stakeholders consiste en comenzar con sesiones de brainstorming, continuando con comités de usuarios más específicos, los cuales son seleccionados de forma que representen las institucionales más importantes y aquellos usuarios que se encuentren más comprometidos con el proyecto. Por otra parte, es importante contar con una herramienta en la que se registre la información recolectada y las validaciones llevadas a cabo por el usuario, así como tener un seguimiento de los requerimientos a través de todo su ciclo de vida, desde la definición hasta la implementación del mismo.

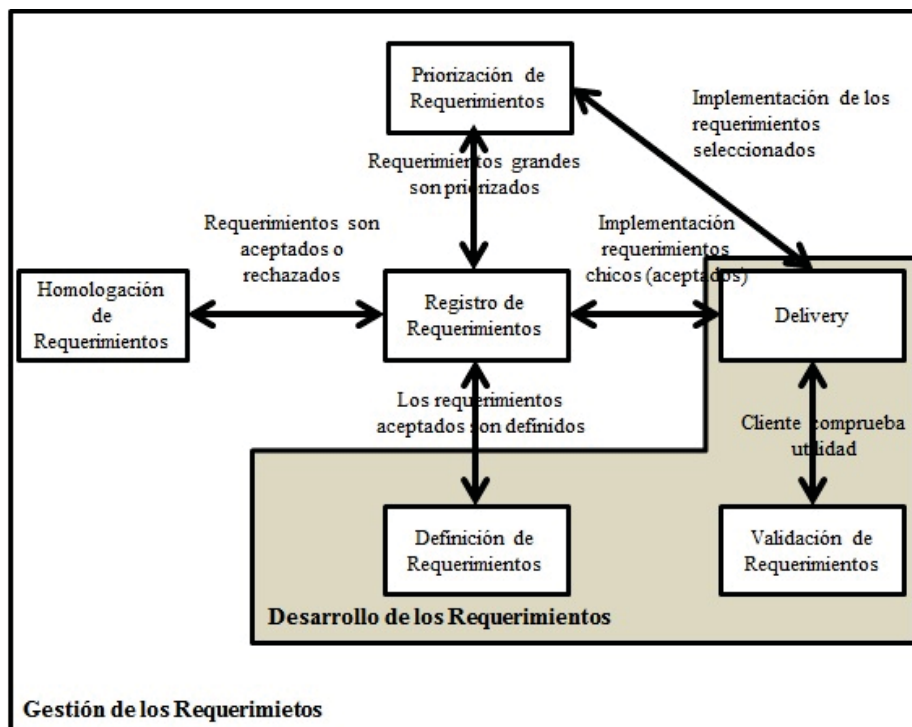


Fig. 2. Proceso global

La reuniones con los stakeholders son presenciales, excepto para algún refinamiento particular de los requerimientos que se llevan a cabo por otros medios (teléfono, correo electrónico, etc.). En cada reunión hay una gran actividad para encausar la discusión por parte del coordinador para que se puedan tratar los temas específicos y que no invertir mucho tiempo en tópicos no relacionados a los requerimientos que se están tratando. Tal vez es de importancia poder incorporar reuniones a través de videoconferencias con el objetivo de optimizar tiempos de viajes y traslados ya que en general las diferentes instituciones, en nuestro caso Universidades, se encuentran distribuidos en diferentes ciudades del país. La incorporación de las videoconferencias puede colaborar en mantener mayor cantidad de reuniones, con menor duración y mayor efectividad.

Metodológicamente se logró utilizar GT para el análisis de los datos obtenidos, aún sin contar con una herramienta específica, sino software de colaboración.

Los próximos pasos se orientarán en primer lugar a contrastar las conclusiones obtenidas con otros stakeholders y otras fuentes de datos. También se profundizarán aspectos clave como el proceso de clasificación de los requerimientos de acuerdo al “tamaño” y las dinámicas de la priorización y la validación.

6 Referencias

- [1] F. C. Lunenburg, “Organizational Structure: Mintzberg’s Framework,” *INTERNATIONAL JOURNAL OF SCHOLARLY, ACADEMIC, INTELLECTUAL DIVERSITY*, vol. 14, no. 1, p. 8, 2012.
- [2] P. Loucopoulos and V. Karakostas, *Systems Requirements Engineering*. McGraw-Hill, 1995.
- [3] R. R. Young, *Effective Requirements Practices*. Upper Saddle Street, NJ, USA: Addison-Wesley, 2001.
- [4] B. Nuseibeh and S. Easterbrook, “Requirements Engineering: A Roadmap,” in *ICSE ’00 Proceedings of the Conference on The Future of Software Engineering*, Limerick, Ireland, 2000, pp. 35–46.
- [5] B. H. C. Cheng and J. M. Attle, “Research Directions in Requirements Engineering,” in *Proceeding FOSE ’07 2007 Future of Software Engineering*, Minneapolis, MN, USA, 2007, pp. 285–303.
- [6] S. Viller and I. Sommerville, “Social Analysis in the Requirements Engineering Process: from ethnography to method,” in *RE ’99 Proceedings of the 4th IEEE International Symposium on Requirements Engineering*, Limerick, Ireland, 1999, pp. 6–13.
- [7] J. Herbsleb and D. Moitra, “Global Software Development,” *IEEE Software*, no. 2, pp. 16–20, Apr. 2001.
- [8] D. Damian, “Stakeholders in Global Requirements Engineering: Lessons Learned from Practice,” *IEEE Software*, vol. 24, no. 2, pp. 21–27, Apr. 2007.
- [9] B. Berenbach, “Impact of Organizational Structure on Distributed Requirements Engineering Processes: Lessons Learned,” in *Proceedings of the 2006 international workshop on Global software development for the practitioner*, New York, NY, USA, pp. 15–19.
- [10] P. Runeson, M. Höst, A. Rainer, and B. Regnell, *Case Study Research in Software Engineering*. Hoboken, New Jersey: John Wiley and Sons, 2012.
- [11] R. Hoda, J. Noble, and S. Marshall, “Grounded Theory for Geeks,” in *PLoP ’11 Proceedings of the 18th Conference on Pattern Languages of Programs*, Portland, Oregon, 2011, p. Art No 24.

Evaluación de performance de engine 3D para dispositivos móviles

Federico Cristina¹, Sebastián Dapoto¹, Pablo Thomas¹, Patricia Pesado^{1,2}

¹ Instituto de Investigación en Informática LIDI,
Universidad Nacional de La Plata – Argentina

² Comisión de Investigaciones Científicas de la Provincia de Buenos Aires – Argentina

{fcristina, sdapoto, pthomas, ppesado}@lidi.info.unlp.edu.ar

Resumen. En la actualidad existen diversos frameworks para el desarrollo de aplicaciones móviles 3D, con un denominador común en todos ellos: la performance es un aspecto crítico incluso más decisivo que en los equipos de escritorio, los cuales en general cuentan con mayor capacidad de cómputo. Las herramientas de análisis de rendimiento o *profiling* con las que cuentan estos frameworks pueden ayudar en cierta medida a determinar la posible existencia de cuellos de botella en la ejecución de aplicaciones. Sin embargo, este tipo de herramientas tiene ciertas limitaciones tales como abarcar solo un espectro de las posibles causas del problema o acotar el análisis a ciertos escenarios en particular. El presente artículo propone una evaluación y medición de la incidencia de las principales características en la performance de las aplicaciones móviles 3D.

Palabras Clave: Unity engine, dispositivos móviles, aplicaciones 3D, performance

1 Introducción

En la actualidad los dispositivos móviles son cada vez más sofisticados y su evolución tecnológica permite ejecutar aplicaciones cada vez más complejas y con rigurosos requerimientos de hardware.

Sin embargo, cuando estas aplicaciones son visuales e incluyen gráficos tridimensionales, es posible notar cierta degradación en la fluidez de ejecución. Esta pérdida en la eficiencia de ejecución es debida a las características propias del dispositivo en el cual se ejecuta la aplicación, pero también a las características de la herramienta de desarrollo y/o de su implementación.

El número de frameworks para la creación de aplicaciones móviles 3D interactivas es cada vez mayor. Cada uno de estos frameworks posee diferentes características que los hacen adecuados para distintos tipos y magnitudes de proyecto.

Al elegir un framework en particular es posible basarse en diversos criterios tales como: la comunidad existente, los lenguajes de codificación permitidos, la facilidad

de uso, la calidad de los gráficos 3D resultantes, entre otros. Sin embargo, uno de los principales puntos de interés es obtener una buena performance en cuanto a visualización y fluidez.

En dispositivos móviles toma mayor relevancia la limitada capacidad de procesamiento comparada con – por ejemplo – la de los equipos de escritorio. Por este motivo, es importante conocer las capacidades y limitaciones del framework a utilizar en lo que respecta a la performance de visualización.

A fin de lograr este objetivo, el presente trabajo propone una evaluación que permite aislar, analizar y dimensionar la incidencia de las principales características relacionadas con la performance visual de las aplicaciones 3D móviles.

Esta evaluación brinda un soporte al ingeniero de software de aplicaciones móviles 3D, posibilitando la identificación de los factores que ralentizan la performance visual de las aplicaciones que desarrolla, y permitiendo ajustar estos puntos críticos hasta lograr la fluidez deseada.

El resto del paper se organiza del siguiente modo: la sección 2 describe la motivación del análisis propuesto; la sección 3 presenta en detalle la evaluación; la sección 4 exhibe la experimentación realizada y, finalmente, en la sección 5 se exponen las conclusiones y el trabajo a futuro.

2 Motivación

La gestación del análisis de performance que propone este trabajo se inicia durante el proceso de desarrollo de dos aplicaciones móviles: R-Info3D [1] e InfoUNLP3D [2]. La primera herramienta es un entorno 3D de aprendizaje de algoritmia básica, mientras que la segunda es un escenario virtual 3D de la Facultad de Informática de la Universidad Nacional de La Plata. Ambos proyectos resultaron en aplicaciones inmersivas, las cuales por su naturaleza presentan un elevado costo computacional, principalmente en lo que refiere a visualización [3].

Tanto en el proceso de implementación como en su posterior ejecución en distintos dispositivos móviles, se encontraron limitaciones en cuanto a la performance obtenida. Se procedió a reconocer los puntos críticos que incidían en la fluidez visual y se determinaron ciertos umbrales que no debían ser superados. Este análisis llevó a modificar las aplicaciones para lograr un mejor funcionamiento en los diferentes tipos de dispositivos móviles.

Previamente al inicio del desarrollo de las herramientas mencionadas, se realizó un análisis de los principales engines de uso libre para el desarrollo de aplicaciones móviles 3D. Los engines considerados fueron: Unity [4], uno de los más populares y sencillos de utilizar; Unreal Engine [5], en comparación con otros engines, algo más complejo de usar y con requerimientos de hardware más elevados; CryEngine [6], pensado principalmente para desarrollos 3D en primera persona, su instalación y uso no son procesos triviales.

R-Info3D e InfoUNLP3D fueron desarrolladas con el framework Unity. Dado que no existe un framework que sea óptimo en todos los aspectos posibles, la elección de Unity por sobre el resto de los frameworks de desarrollo móvil 3D analizados estuvo basada en una serie de factores, entre los que se destacan:

- Tutoriales: existen una gran cantidad de tutoriales y ejemplos que guían al nuevo usuario en el proceso de aprendizaje del uso del framework. Estos tutoriales están categorizados según el tipo de desarrollo, son audiovisuales y están provistos de todos los elementos necesarios para realizarlos (objetos, audios, imágenes, scripts, etc).
- Documentación: el manual de usuario [7] es amplio, de fácil comprensión y está subdividido debidamente en distintas categorías. Contiene un buscador que facilita la búsqueda de un tema o funcionalidad en particular.
- Requerimientos de software/hardware: los requisitos de sistema son considerablemente más bajos en comparación con otros frameworks similares. Esto, acompañado de una instalación simple y rápida, incentiva al usuario/desarrollador a comenzar rápidamente con la utilización del framework.
- Componentes disponibles: cuenta con un repositorio (*Asset Store*) en donde es posible encontrar una gran variedad de componentes que pueden acelerar el desarrollo de las aplicaciones. Contiene un buscador de componentes que permite realizar búsquedas complejas mediante distintos tipos de filtro.
- Facilidad de aprendizaje / uso: el framework es versátil y permite trabajar en dos lenguajes diferentes: C# y javascript. Esto, sumado a la documentación, los tutoriales, el repositorio y la comunidad existente, generan un escenario ideal para aprender rápidamente a usar el framework y resolver cualquier problema que pueda surgir en el proceso.
- Comunidad: la gran popularidad de Unity es un factor preponderante al momento de elegirlo. Es el framework 3D más utilizado y su comunidad está compuesta por más de dos millones de usuarios [8]. Contiene un foro subdividido en categorías en donde es posible plantear las situaciones o problemas que pueden surgir durante el desarrollo de una aplicación.

Existen diferentes propuestas de análisis de performance de engines 3D para aplicaciones móviles. Una de estas propuestas [9] evalúa la performance de las aplicaciones teniendo en cuenta el consumo de CPU y/o GPU que éstas generan. Otros trabajos [10, 11] basan su evaluación en un análisis de las principales características y funcionalidades de los frameworks, dando como resultado una lista o tabla comparativa.

Ninguno de los trabajos encontrados adopta el enfoque aquí propuesto, es decir, que den algún tipo de guía para el desarrollo de aplicaciones móviles 3D basándose en la performance final de visualización lograda en dichas aplicaciones.

3 Evaluación propuesta

El objetivo principal de la evaluación propuesta consiste en aislar cada una de las principales características que abarca una aplicación 3D; especialmente aquellas que tienen impacto directo en la performance, tiempo de respuesta y fluidez de ejecución de las aplicaciones generadas con el engine [12].

Características tales como el número de polígonos, la aplicación de luces y sombras, la utilización de texturas y/o transparencias, la visualización de sistemas de partículas y el cálculo de la física de objetos que componen la escena, son ejemplos de los principales ítems a evaluar.

Si bien la incidencia de estas características sobre la performance varía de acuerdo al software y al hardware sobre el cual se ejecuta la aplicación, las pruebas realizadas demuestran que existe un patrón común respecto a la degradación de performance en relación al aumento de los requerimientos visuales.

En función del conjunto de características previamente mencionado, se define una serie de pruebas independientes para evaluar la performance. Dichas pruebas se enumeran a continuación:

1. Renderización de objetos simples (*Basic Mesh Rendering*): se presentan progresivamente en pantalla objetos simples sin textura en movimiento en una escena sin iluminación ni sombras. Los objetos deben rotar continuamente a velocidad constante. El número de objetos en pantalla irá creciendo acorde transcurre el tiempo. Se contabiliza el número de fotogramas por segundo o *frame rate* (FPS) a lo largo de la simulación en función del número de objetos.
2. Renderización de objetos complejos (*Complex Mesh Rendering*): consiste en visualizar un objeto complejo en movimiento, el cual debe contener un elevado número de polígonos. La distancia de renderizado (*clipping plane*) se va incrementando a medida que avanza la prueba. Se contabiliza el FPS a lo largo de la simulación en función de la distancia de renderizado.
3. Luces y sombras (*Lights & Shadows*): se realiza una simulación similar a la renderización de objetos simples, pero en este caso la escena contiene iluminación y objetos con proyección y recepción de sombras. Se contabiliza el FPS en función del número de objetos.
4. Texturas (*Textures*): se realiza una simulación similar a la renderización de objetos simples, pero en este caso los objetos poseen texturas complejas, como puede ser transparencias, reflejos, etc. Se contabiliza el FPS en función del número de objetos.
5. Sistemas de partículas (*Particle Systems*): se crea una escena en donde se presenta progresivamente nuevas instancias de un sistema de partículas (por ejemplo humo, chispas, explosión, etc.). Se contabiliza el FPS en función del número de sistemas de partículas.
6. Física (*Physics*): se realiza una simulación similar a la renderización de objetos simples, pero en este caso a los objetos se les aplica reglas de física, como por ejemplo *gravedad*. Se contabiliza el FPS a lo largo de la simulación en función del número de objetos.

Aplicando este conjunto de pruebas se simplifica considerablemente la tarea de determinar los puntos críticos a optimizar en las aplicaciones desarrolladas, posibilitando así la identificación del punto de equilibrio en la calibración de las características analizadas.

En particular, para el caso de estudio de este trabajo se utilizaron cubos como objetos simples, mientras que para la prueba de renderización de un objeto complejo se eligió el modelo del edificio de la Facultad de Informática, que es utilizado en la aplicación InfoUNLP3D. Dicho objeto contiene más de 500.000 polígonos y un

elevado número de ventanas. Por esto último, para las pruebas de texturas se aplicó transparencia de tipo vidrio a los elementos. En cuanto al sistema de partículas, se generó uno similar al utilizado por el robot en R-Info3D al momento de ejecutar una instrucción de reposicionamiento (Pos).

4 Experimentación

El experimento consistió en realizar las pruebas enumeradas en el punto anterior en un conjunto de dispositivos móviles con diferentes características. Los dispositivos utilizados fueron tres smartphones y dos tablets: Samsung Galaxy S2 (smartphone), Samsung Galaxy J5 (smartphone), LG L5 II (smartphone), Asus MemoPad FHD10 (tablet) y Acer B1-730 (tablet). Estos dispositivos presentan diferencias considerables en sus prestaciones y cuentan con diferentes arquitecturas de hardware, como ARM y x86.

Las figuras 1 y 2 muestran los valores medios de los resultados obtenidos en las pruebas realizadas a lo largo de todos los dispositivos, considerando los dos extremos posibles en lo que respecta a calidades de renderizado en la plataforma Unity: *Fastest* (la más baja) y *Fantastic* (la más alta) [13].

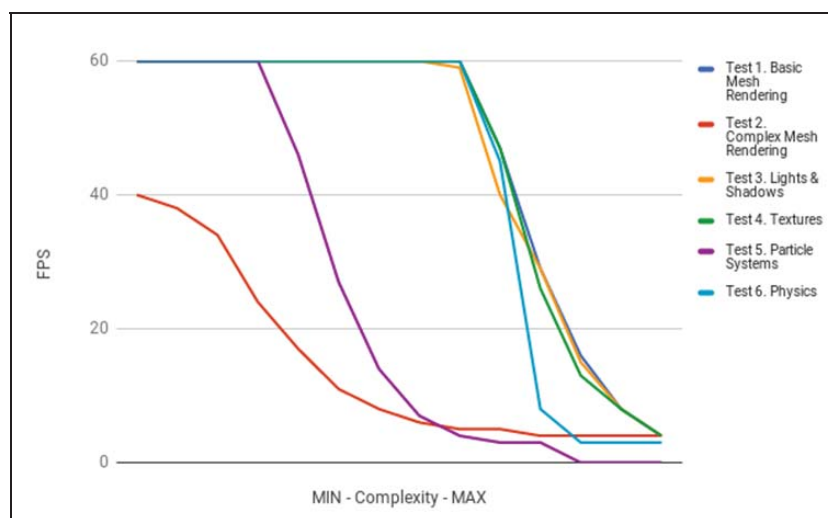


Fig. 1. FPS. Evolución de cada prueba en calidad *Fastest*.

En todos los casos, la información se encuentra normalizada en función del rango de valores mínimos y máximos a renderizar dependiendo del tipo de prueba realizada. En las pruebas 1, 3, 4, 5 y 6 el número de objetos en pantalla inicia en 1 y se incrementa gradualmente a lo largo de la simulación hasta llegar a un valor elevado, como puede ser 10000 objetos. Para la prueba 2, la evolución en este caso es la distancia de renderizado, que puede ir desde un mínimo de 200 hasta un máximo de

2000 unidades de distancia. Adicionalmente, el *frame rate* ideal considerado es de 60 FPS.

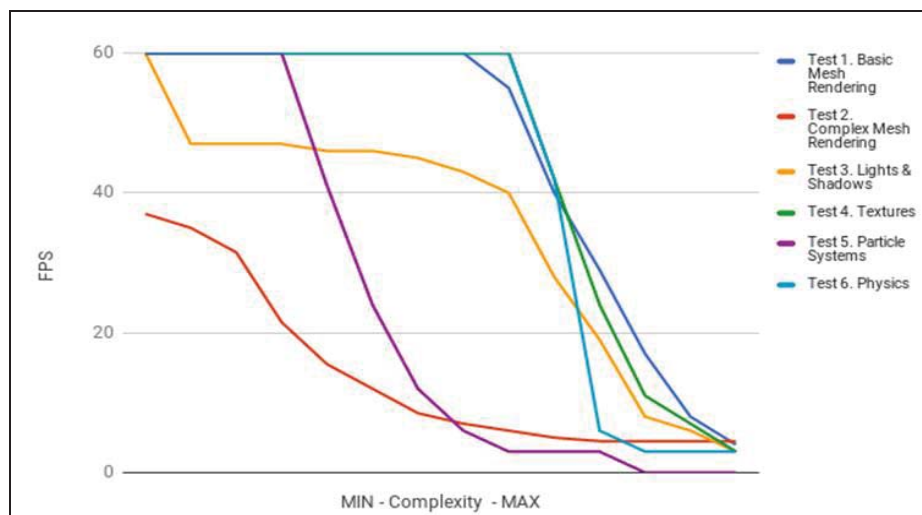


Fig. 2. FPS. Evolución de cada prueba en calidad *Fantastic*.

Los resultados de las pruebas permiten obtener un conjunto de conclusiones de interés, posibilitando la optimización correspondiente de las aplicaciones. Las observaciones a continuación son el resultado de aplicar la evaluación propuesta específicamente para el caso de estudio en cuestión.

- El render básico tanto en *Fastest* como en *Fantastic* no presenta diferencias en su rendimiento. Ambas curvas presentan prácticamente la misma evolución a lo largo de la simulación.
- La luz afecta la performance en calidad *Fantastic* pero no así en *Fastest*, en donde parece ser directamente ignorada en lo que respecta a cálculos y renderización. Tal como se detallará posteriormente, la aplicación de luces/sombras es uno de los factores determinantes en lo que respecta a performance visual.
- El número de polígonos en el modelo de la Facultad de Informática (+500k) es excesivamente alto para la capacidad de procesamiento de los dispositivos móviles utilizados, a fin de que la simulación pueda ser visualizada con la fluidez necesaria. Bajo ambas calidades gráficas la degradación es prácticamente igual.
- El impacto de aplicación de texturas no parece afectar considerablemente la performance general, al menos para la textura utilizada (la cual incluye transparencia), aplicada mediante un tipo de sombreado estándar, más conocido como *standard shader*, y un tamaño de textura moderado.
- Los sistemas de partículas, al ser 2D no se ven considerablemente afectados por los cambios de la calidad; pero ya un volumen alto de los mismos impide

la correcta ejecución de la simulación dada la gran cantidad de cálculos individuales que requiere cada una de las partículas que componen los sistemas.

- La aplicación de detección de colisiones/leyes de física tiene una incidencia relativamente menor sobre la performance cuando se trata de un número de objetos en escena "razonable" (por ejemplo, menos de 2000 cubos).

La figura 3 muestra la relación de FPS entre las calidades *Fastest* y *Fantastic*. Una relación igual a 1 denota que la prueba en cuestión incide de la misma forma en ambas calidades; mientras que un valor positivo denota mayor incidencia sobre *Fantastic* y un valor negativo mayor incidencia sobre *Fastest*. En la prueba 3 (luces y sombras) puede apreciarse notoriamente la mayor incidencia sobre la performance utilizando la configuración *Fantastic*. Se observa una fluidez general promedio del 50% con respecto a la configuración *Fastest*, que obviamente relega calidad gráfica en cuanto a estas características.

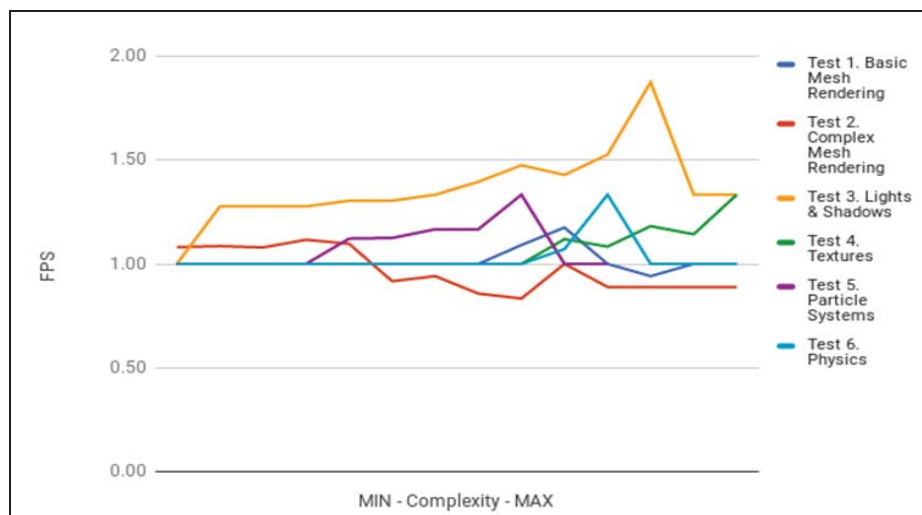


Fig. 3. FPS. *Fastest* vs. *Fantastic*.

Por otra parte, tal como se detalló previamente, la complejidad del objeto a ser renderizado (prueba 2) impacta equitativamente tanto en calidad *Fastest* como en *Fantastic*, y superado cierto umbral la distinción entre una renderización y otra es prácticamente nula o incluso negativa.

Gracias a este análisis, fue posible lograr un equilibrio aceptable entre calidad gráfica y fluidez de visualización en las aplicaciones desarrolladas en Unity. La figura 4 presenta una escena de la aplicación InfoUNLP3D con el mayor nivel posible de detalle en características tales como texturas, luces y sombras, pero con una performance extremadamente pobre en lo que respecta a FPS. La figura 5 muestra la escena una vez modificada a fin de lograr una fluidez óptima, aunque sacrificando en exceso las características previamente mencionadas. La figura 6 presenta los

resultados logrados mediante la calibración óptima seleccionada, basada en los resultados obtenidos en las pruebas. En dicha calibración se prioriza tanto la calidad de imagen como la fluidez de visualización. Se omite, por ejemplo, el uso de luces focalizadas (*spot lights*) para evitar la excesiva generación de sombras, dado que la finalidad de la aplicación no es presentar un render fotorealístico sino servir como guía de referencia para los alumnos.



Fig. 4. InfoUNLP3D. Máximo nivel de detalle.



Fig. 5. InfoUNLP3D. Mínimo nivel de detalle.

Bajo este contexto, se agregó además una nueva funcionalidad a la aplicación InfoUNLP3D que permite al usuario configurar la distancia de renderizado (y por consiguiente el número de polígonos en escena). Al evitar la renderización de los elementos más distantes, se gana aún mayor fluidez sin perder la funcionalidad de la

herramienta, que se basa en recorrer la facultad recibiendo información de las aulas cercanas. Esta nueva opción permite utilizar la aplicación en dispositivos móviles con menor capacidad de procesamiento.



Fig. 6. InfoUNLP3D. Calibración óptima seleccionada.

5 Conclusiones y trabajo a futuro

Este trabajo propone un conjunto de heurísticas para determinar los principales factores de incidencia en la degradación de la fluidez y visualización de las aplicaciones móviles 3D.

La evaluación propuesta permite individualizar estos factores a fin de rediseñar una aplicación y reducir considerablemente sus problemas de performance.

Como prueba de concepto y para validación de su efectividad, se aplicó la evaluación sobre el engine Unity. Se llevaron a cabo las pruebas de simulación, obteniendo como resultado final la información necesaria para realizar las optimizaciones correspondientes en las aplicaciones desarrolladas con el engine mencionado.

A partir de este trabajo, un ingeniero de software de aplicaciones móviles 3D, dispone de un conjunto de pautas a considerar en la optimización de la performance visual de las aplicaciones desarrolladas.

A futuro se prevé aplicar esta misma evaluación sobre otros engines de desarrollo de aplicaciones móviles 3D, tales como Unreal Engine y CryEngine; e idealmente realizar una comparación de performance entre los frameworks para cada una de las características evaluadas.

Referencias

1. Cristina, F.; Dapoto, S.; Thomas, P.; Pesado, P. "Prototipo móvil 3D para el aprendizaje de algoritmos básicos". XXI Congreso Argentino de Ciencias de la Computación CACIC. 2015. ISBN: 978-987-3724-37-4.
2. Cristina, F.; Dapoto, S.; Thomas, P.; Pesado, P. "InfoUNLP3D: An interactive experience for freshman students". XXII Congreso Argentino de Ciencias de la Computación CACIC. 2016. ISBN: 978-987-733-072-4.
3. Linowes J. "Unity Virtual Reality Projects". 2015. ISBN-13: 978-1783988556
4. Unity. <https://unity3d.com>
5. Unreal Engine. <https://www.unrealengine.com/>
6. CryEngine. <https://www.cryengine.com/>
7. Unity online manual. <http://docs.unity3d.com/Manual/index.html>
8. Unity Blogs: The Unity Community. <https://blogs.unity3d.com/2014/03/11/we-got-karma/>
9. Messaoudi F., Simon G., Ksentini A. "Dissecting Games Engines: the Case of Unity3D". International Workshop on Network and Systems Support for Games (NetGames). 2015. Electronic ISSN: 2156-8146.
10. Akekarat Patrasitidecha. "Comparison and evaluation of 3D mobile game engines". Chalmers University of Technology. University of Gothenburg. 2014.
11. Petridis P., Dunwell I., Panzoli D., Arnab S., Protopsaltis A., Hendrix M., de Freitas S. "Game Engines Selection Framework for High-Fidelity Serious Applications". "International Journal of Interactive Worlds". 2012. Article ID 418638. DOI: 10.5171/2012.418638.
12. Optimizing Graphics Performance. <https://docs.unity3d.com/Manual/OptimizingGraphicsPerformance.html>
13. Unity Manual: Quality Settings. <https://docs.unity3d.com/Manual/class-QualitySettings.html>

Evaluating Quality in Agile Developments. A first validation experience with NEA Software SMEs.

Noelia Pinto; César Acuña; Nicolás Tortosa; Blas Cabas Geat

Research Group in Software Engineering and Quality (GICS).
National Technological University, Regional Faculty
{ns.pinto; csr.acn;blasc147;nicotortosa}@gmail.com

Abstract. To obtain products of high quality software it is necessary to carry out good processes management in which measurement is a key factor. Therefore, companies should focus on continuous improvement cycles that integrate both the development process and the product obtained, to increase quality in both aspects. This cycle of improvement involves the adoption of a quality model appropriate to the characteristics of the company and a methodology that guides the software development cycle. In this sense, the agile philosophy proves to be the most suitable approach for the current development environments, and they are positioned as an alternative to the development processes with high cost in documentation and excessively long processes. This paper presents the results obtained by automation of QuAM Model [1] for the quality evaluation of agile projects on actual production environments and the subsequent analysis based on these projects.

Keywords: Software Engineering; Quality Software; Agile Projects

1 Introduction

In Argentina, the Software Industry is mainly made up of SMEs (Small and Medium Enterprises) where the quality of the work done, the low costs and the timely deliveries are essential elements for the increase of internal sales and projection at the international level [2].

There are numerous methodological proposals that guide the software development cycle and that impact in different dimensions in the process. Among the most used proposals we can mention the traditional methodologies, which are especially focused on a rigorous definition of roles, activities that are involved, artifacts that must be produced, tools and notations that will be used [3]. However, these approaches do not prove to be the most appropriate for many of today's projects, where the system environment is very changeable, and where it is required to reduce drastically development times, but maintaining a high quality both at the process level and the product.

In contrast to these approaches, agile methodologies emerge which pursue principles such as the incremental delivery of new functionality to the customer, which is prioritized per business value it adds (in this way the software product evolves in the different deliveries), continuous improvement and focused on close collaboration between the team of programmers and business experts [4].

To develop their products, the agile methodologies offer these SMEs the possibility of having lightweight and simple processes, which can be adapted to their structure. Quality (both in software developed and in its corporate image) and reputation begin to be competitive and differentiation's factors in this segment. Since the quality of the software product developed is closely related to the quality of the process used, SMEs need to implement strategies to improve their processes that allow them to increase the quality of their products.

In this way, there is a need to provide a framework that allows evaluating the quality when they choose to work with agile methodologies, which arises from the analysis of the situation of the Software Industry in the NEA region (North-East Region of Argentina) regarding the adoption of the life cycle that guides processes of software development [5].

In previous works [1][6][7], QuAM Model (Quality Agile Model) has been presented, which is a first approximation to the definition of a model that allows the evaluation of the quality of projects in agile environments. Also, the results obtained have been exposed, after analyzing the validation experience of this proposal with some of the SMEs software development companies in NEA region.

However, to evaluate the quality of agile projects, it is not enough to define a model, for that reason the development of QuAGI has been approached: A web application to provide support the management of the QuAM model, integrating the follow-up agile projects and quality evaluation throughout the process [8].

The present article that continues with this line of research Aims to expose the results obtained when evaluating the quality of agile projects in real production environments using QuAGI together with the analysis of the validation experience. The article is structured as follows: Section 2 presents a brief review respect to the models and tools that could be used for quality evaluation in agile environments. In Section 3, it is briefly described the architecture of QuAGI and it includes the study case that allows defining the QuAGI implementation platform. Then, in Section 4, the results of the experience of linking between the technical team and some companies dedicated to software development in the region are presented. Finally, Section 5 presents conclusions and future works.

2 Related Works

In the literature related to the research topic that we address in this work we have found studies that are focused on the relationship between agile processes and quality assessment from various perspectives. We present below, the topics that we consider most relevant, grouping them in studies where models for the quality evaluation and studies that allow to realize the management and monitoring of the projects are presented.

One of these cases is presented in [9], where the authors present AGIS, a tool capable of measuring the degree of agility of a process per the values of the agile manifesto, based on the principles of improvement and audit of ISO 9001: 2008 [10]. The purpose of this is to obtain an objective measure of the productive process that avoids false positives on both sides. AGIS aims to satisfy two needs: on the one hand,

it is focused on the companies, since this model allows achieving a differentiation with respect to other companies that have only certified quality through ISO 9001: 2008. On the other hand, AGIS to offers a report of improvement suggestions based on the assessment of the dimensions it proposes to evaluate. The model also offers an objective definition of the degree of agility of a project, which can be used to compare different projects.

Another similar model is AGIT (AGile software developmenT) [11] which suggests that the best performance is achieved when the goals of all stakeholders are met. This requires an approach that considers the views of different stakeholders, for which the appropriate indicators are defined for each one. AGIT considers four different points of view for stakeholders: the IT Administrator is the actor concerned with the traditional aspects regarding the performance of SW development considering time, cost and quality; the second actor is represented with team members whose goal is "job satisfaction"; The Scrum Master whose main goal is the "efficient resolution of impediments". Finally, the fourth stakeholder is customer satisfaction. This model suggests evaluating the quality of the development processes considering the points of view of the different stakeholders involved, describing the indicators that are appropriate to each of these profiles. Considering both AGIS and AGIT models, it is observed that there is no proposal that allows the quality evaluation of the agile processes themselves.

Regarding software tools, there are several alternatives that allow the monitoring and knowledge management of software projects based on agile methodologies. One of these applications is Trello [12], the web project management tool based on the agile Kanban method [13], with cards (representation of an activity described in a sentence) that cross different lists considering their status (pending, in process, finished). The lists are inside boards, and within the lists are the cards. The cards support any kind of documentation, images, videos, lists, comments, etc. So, they also serve as an instrument of communication in working groups, about a certain topic.

Another application used for knowledge management in software projects for agile methodologies is Jira [14]. This tool allows the handling of version control, notifications to the team members of new tasks to carry out or modifications to existing ones. Like other similar tools, it records operations to maintain traceability between artifacts. Each of them supports a life cycle in which you can add, remove or change transitions if required.

A third application studied is Taiga.io [15], open source web platform for the control and planning of projects using SCRUM. It is offered as software to be downloaded and installed on your own infrastructure or you can also use a SaaS version. Within the latter alternative, whenever public software projects are created, the tool will be free.

In this way, it is observed that there is no alternative to integrate both monitoring of agile projects and quality evaluation, based in a model defined for this purpose. Therefore, from this line of research arises the proposal of a framework, AQF (Agile Quality Framework), which consists of a new quality model, QuAM, and for the moment, an only web application, QuAGI, which manages the components of that model.

3 Case Study: Real implementation of QuAGI

3.1 QuAM: Model for assessing the quality of agile projects.

In [6] we present QuAM, a model that, in its first approximation, aims to provide a method that allows to evaluate the quality of both software development processes based on agile practices and final products. QuAM, as indicated in Table 1, defines a scheme of components, including metrics and attributes, to configure a quality evaluation model that provides an objective measurement of the quality of the process implemented in each project.

Table 1. Tree of Metrics defined in QuAM

<i>Metrics 1: Life Cycle Selection</i>	
Positive Attributes	Negative Attributes
A1.1 Give value to Iterative and Incremental Cycle.	A1.2 Give value the Waterfall Cycle.
<i>Metrics 2: Evaluation of the Work Team.</i>	
Positive Attributes	Negative Attributes
A2.1 Give value to team meetings.	A2.2 Give value to schedule compliance.
A2.3 Give value to the roles definition	A2.4 Give more value to the process than to the team
<i>Metric 3: Production capacity of deliverables.</i>	
Positive Attributes	Negative Attributes
A3.1 Give value to use of change management tools.	A3.2 Give value to management of requirements.
A3.3 Give value to functional product.	A3.4 Give value to documentation.
<i>Metric 4: Communication with Product Owner</i>	
A4.1 Give value to collaboration with the product owner.	A4.2 – Give value to contractual negotiation

3.2 QuAGI: Web application for the monitoring and evaluation of agile projects.

It is not enough to have a quality model that allows to measure the level of quality of an agile process if a tool is not available that makes it possible to manage the elements of the model and to analyze the results obtained from various evaluated cases. Therefore, to give support QuAM, has been developed QuAGI [8]. This is a web application that allows the monitoring of projects based on agile practices together with the possibility of conducting continuous evaluations regarding the level of quality that is being achieved in the process.

First, QuAGI allows the administration of the projects through the visualization of the plan, provides reports regarding the states of the plan, provides comprehensive

information of the activities, serves as an internal communication tool, among others functions. Secondly, it supports the decision-making processes by assisting those responsible through reports about the quality evaluation of the project in question and recommendations for continuous improvement, all through the incorporation of algorithms of Artificial Intelligence integrated with QuAGI.

The proposed architecture of QuAGI, as shown in Figure 1, consists of a platform based on reusable components, allowing the applications that are integrated to the platform make use of such components. In the presentation layer, the interface design and user interaction was done using the framework Materialize along with jQuery functions. As a development framework Django is used, based on an MVC (Model-View-Controller) pattern that separates business logic from presentation logic by using the template system. The choice of Django is fundamentally due to the philosophy of reuse and fast development of web applications. Related to this, the decision to use PostgreSQL is due to the robustness it provides by being integrated into this framework.

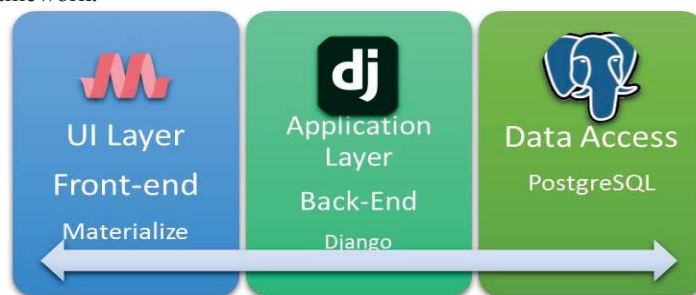


Fig. 1. Architecture

So far, QuAGI allows access to three types of users: Role of Project Manager (maximum responsible for the Project), Team Members (those who are assigned tasks for each project) and Product Owner (who defines the requirements of the project). For example in Figure 2 shows the interface that allow manage projects, track tasks, view reports, and so on, based on the permissions granted.

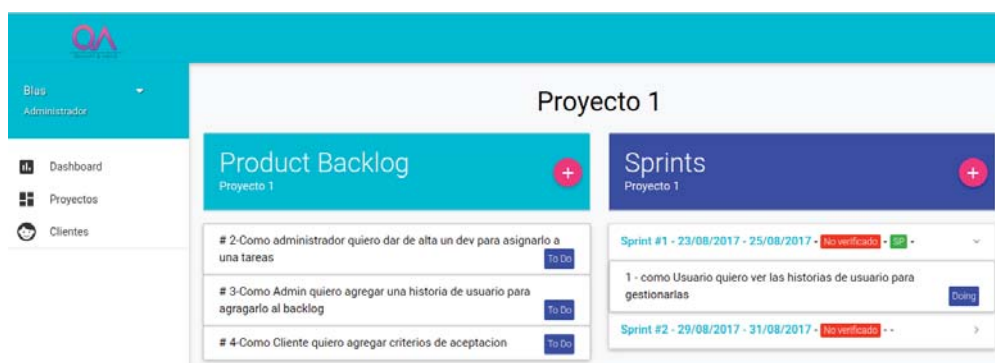


Fig. 2. UI "Details of each Project"

3.3 QuAGI Validation Process

In order to analyze the impact of AQF on SMEs, a stepwise validation process was implemented, as shown in figure 3, on real production environments based on the use of QuAGI as an agile project management tool.



Fig. 3. Validation Process

The stages of the Plan are described below:

- Selection of companies: From the population of SMEs dedicated to the development of software in Chaco and Corrientes, only a few companies were selected for the validation experience, considering those that had agile projects in production. The sample size is equal to 5 companies, which represents 25% of the SMEs population in total.
- Implementation and Monitoring: Because the tool is in the process of adjustment and considering the availability of the companies to collaborate with the implementation process, it was proposed to start with the evaluation of one of the QuAM metrics.

The chosen metric was "Communication with the Product Owner" whose attributes and criteria are shown in table 2. This choice is based on the fact that the agile projects that participated in the experiment have not yet completed; so it is not feasible for QuAGI, in an early stage of the process, to obtain quality values for the rest of the metrics. This does not happen with metric 4, because information about client is requested at start and at the same time it is necessary to indicate how the project requirements will be managed.

Table 2. Metric 4 “Communication with Product Owner”

Positive Attributes	Negative Attributes
A4.1 - Assess the collaboration with the client. - Product Owner is part of the team, responds to queries, plans iterations, and collaborates in writing requirements (3) - Product Owner is part of the team, responds to queries and plans iterations	A4.2 - Assess the contract negotiation. - There is detailed contracting at the beginning and no changes are accepted (-3) - Recruitment requires contemplating changes during the project (-1) - The contract exists but does not

(1) - Product Owner cooperates with team demands (0)	affect the project at the level of the development process (0)
---	--

Each of the teams had to load information from their projects into the enabled dashboards in the tool per each account requested to obtain preliminary results. Then, QuAGI's technical team carried out the monitoring of the project management with the quality evaluation reports that were made at different times.

In case of metric 4, only the interaction of company's Product Owner with the project was analyzed. QuAGI considered the following factors in order to obtain the measured values:

- To give value to the positive attribute, an evaluation is performed when the customer is registered:
 - If your relation to the project was indicated as "Per contract only", the corresponding value is 0.
 - If it was included as a member of the project team, but it was detected that it does not write user stories and it only validated iterations, it corresponds to a value of 1.
 - If it collaborates as a member of the team, in the writing of user stories and validates the iterations corresponds to the maximum value of 3.
- To give value to the negative attribute, the contract is evaluated:
 - If the contract already includes the requirements without possibility of generating new user stories corresponds a value of -3.
 - If the project is associated with a contract, with immovable requirements but it is possible to incorporate new functionalities corresponds a value of -1.
 - If the contract exists but does not directly affect the development process (e.g. only defines economic factors), the value is 0.
- Validation: Based on the measurement process that QuAGI was carrying out, control points were established to validate partial results. Thus, working together with each company, the feedback was obtained taking into account the experience of an agile project.
- Proposals for Improvements: Finally, the responsible team established improvements to the platform that will be included in the next iteration together with the need to modify the model to fit the reality of agile projects.

4 Results Obtained

After completing the first implementation cycle, which lasted approximately 45 days, and considering the checkpoint agreed with the companies, the analysis of the

information generated by QuAGI was performed, obtaining the results that can be observed in figure 4.

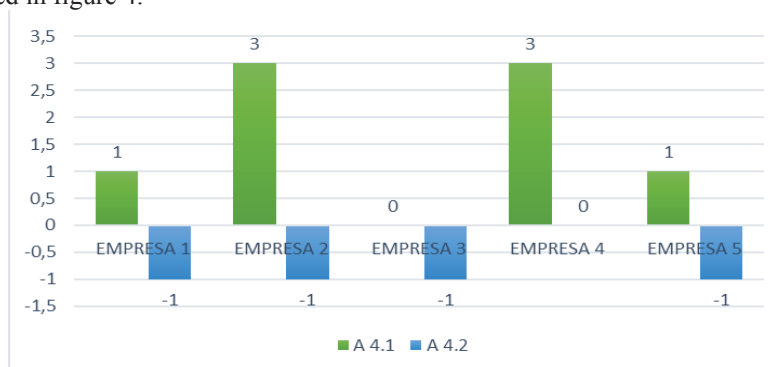


Fig. 4. Evaluation Results: Atributo 4 of Metric 4

In this graph, it can be observed that 40% of the companies studied achieve the maximum value for the positive attribute "Evaluate the collaboration with the product owner". However, it can be distinguished that 4 of the 5 companies obtain a value -1 for the negative attribute "Evaluate the contractual negotiation" which means that there is a contract with rigid requirements but with the possibility of incorporating new functionalities. All of this, is summarized in the quality values associated with metric 4 that are included in figure 5.

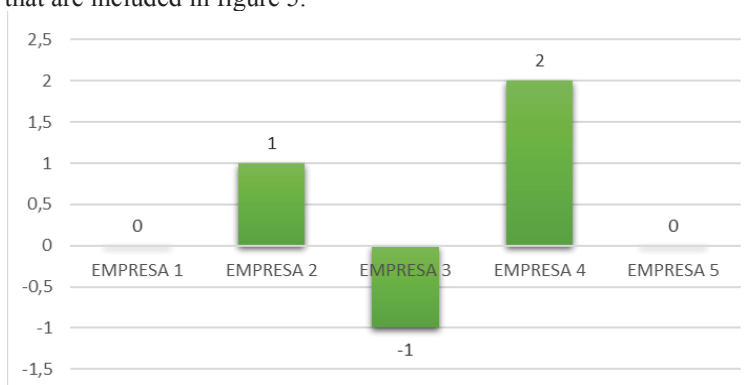


Fig. 5. Quality Values: Metric 4

Only 2 of the 5 companies that participated in the validation experience are close to the optimum value for the metric 4, taking into account that all companies worked with agile projects, using QuAGI as a tracking tool.

5 Conclusions

This work presented the results obtained from the quality evaluation of agile projects in real production environments in SMEs of Chaco and Corrientes, using

QuAGI, as a tool for monitoring agile processes and management of components defined based on a new model called QuAM.

Thus, on the one hand, model QuAM is proposed as an approximation to a new model that allows the initiation of quality evaluation in real software projects guided through agile practices. And, on the other hand, to offer a web platform called QuAGI that allows companies to optimize the quality in the process of their agile software projects, providing information not only referring to the monitoring itself, but also to the associated quality.

After this first stage of implementation, and through different cases that allowed the validation of the framework in real environments, it has been observed that companies easily incorporated QuAGI as a support tool for the development process. This is due, fundamentally, to a user-friendly interface and clearly defined workflows. Likewise, with respect to the results obtained through the quality evaluation of the only implemented metric, the SMEs remarked that they reflect what is expected per the characteristics of each one but that are not significant attributes to the quality of the project in question.

Therefore, and as part of future work, the need to adjust the proposed model arises from the definition of new attributes in each metric according to real agile practices. In addition, it will be necessary adjust QuAGI and to carry out more case studies to achieve a more approached version to the reality of the SMEs and yours agile projects. Finally, continuing with the development of the platform will mean extending the functionality of the tool to the evaluation of the remaining metrics as defined in QuAM.

Acknowledgments. The work presented here is framed in the project "Evaluation of Quality in Agile Processes of Software Development", which is funded by the UTN and executed in GICS of UTN FRRe, with code IAI4445TC. Also, the article is part of the activities planned in the Technological and Social Development Project (PDTS) presented as "Contribution to the competitiveness of the software development companies of the NEA", IP253, evaluated and approved by the National Council of Scientific and Technical Research of Argentina (CONICET). And it is related to activities from Project "Tools and Methods to support Software Engineering: requirements, agile strategies and quality of processes and products", residing in the Regional Faculty of Santa Fe and approved as a Project of the Secretariat of Science, Technology and Postgraduate of the UTN, with code IPN4409. In addition, the results contribute to the activities of the project "Quality Assessment Model in Agile Processes of Software Development" funded by Gastón Dachary University under the code A07010. Finally is necessary to acknowledge to PhD Verónica Bollati for her contributions by your experience in agile environments and the translator Miryam González for her collaboration in the writing of this article in English language.

References

1. Noelia Pinto, Gabriela Tomaselli, Liliana Cuenca Pletsch, Nicolás Tortosa, César J. Acuña “Validación del diseño de componentes de QuAM: un Modelo de Calidad para procesos Ágiles”. Publicado en Libro de Actas del IV Seminario Argentina-Brasil de Tecnologías de la Información y la Comunicación (SABTIC 2016).
2. “Reporte anual del sector de software y servicios informáticos de la República Argentina”, Abril 2016, CESSI, Argentina, <http://www.cessi.org.ar/descarga-institucionales-2007/documento2-130347cd83ae771a9f3db3da5407269a>
3. Letelier, P., Penadés, P. “Metodologías ágiles para el desarrollo de software: eXtreme Programming (XP)” Técnica Administrativa, Buenos Aires. ISSN 1666-1680, 2006.
4. Rujana M., Romero Franco N., Tortosa N., Tomaselli G., Pinto N. (2016). Análisis sobre adopción de metodologías ágiles en los equipos de desarrollo en pymes del NEA. GICS, UTN, FRRe. XVIII Workshop de Investigadores en Ciencias de la Computación (WICC 2016, Entre Ríos, Argentina).
5. Acuña, C., Cuenca Pletsch, L., Tomaselli, G., Pinto, N., Tortosa, N. “Calidad de Software y Metodologías Ágiles en las PYMES de la Industria del Software”. Publicado en Memorias de 3er Congreso Nacional de Ingeniería Informática / Sistemas de Información (CONAIISI 2015). Noviembre, 2015. ISBN 978-987-1896-47-9.
6. Pinto, N., Tomaselli, G. et al. “Hacia un modelo de evaluación de calidad de Procesos Ágiles”. Publicado en Memorias del 4to Congreso Nacional de Ingeniería Informática / Sistemas de Información (CONAIISI 2016). Noviembre, 2016.
7. Pinto, N., Acuña, C., Cuenca Pletsch, L. “Quality Evaluation in Agile Process: A First Approach”. XIII Workshop Ingeniería de Software (WIS). XXII Congreso Argentino de Ciencias de la Computación (CACIC 2016).
8. Pinto, N., Tomaselli, G., et al. “QuAGI: Una propuesta para el seguimiento y evaluación de proyectos de Software Ágiles”. Publicado en los Anales del V SABTIC, VIII STIN y XVIII Foro, Três de Maio, Brasil. DOI: 10.5281/zenodo.583174
9. Matalonga, S., & Rivedieu, G. AGIS: hacia una herramienta basada en ISO9001 para la medición de procesos ágiles. *Computación y Sistemas*, 19(1), 163-175, <http://www.agilemanifesto.org/iso/es/>
10. Beck, K., Beedle, M., Van Bennekum, A., Cockburn, A., Cunningham, W., Fowler, M., Grenning, J., Highsmith, J., Hunt, A., Jeffries, R. and Kern, J., 2001. *Manifiesto for agile software development*.
11. Cohen, D., Lindvall, M. and Costa, P., 2003. *Agile software development*. DACS SOAR Report, 11.
12. Trello, <https://trello.com/>
13. Garzás, J. “¿Qué es el método Kanban para la gestión de proyectos?”, <http://www.javiergarzas.com/2011/11/kanban.html> (2011)
14. Jira, <https://www.atlassian.com/software/jira>
15. Taiga.io, <https://taiga.io>

Exploring Architectural Model Checking with Declarative Specifications ^{*}

Fernando Asteasuain^{1,3} and Francisco Tarulla²

¹ Universidad Nacional de Avellaneda, Dpto Tecnología y Administración, Avellaneda, Argentina. fasteasuain@undav.edu.ar

² Universidad de Buenos Aires, Dpto de Computación, Buenos Aires, Argentina. ftarulla@dc.uba.ar

³ UAI-CAETI, Buenos Aires, Argentina.

Abstract. In this work we explore the FVS language in the context of architectural behavior model checking. FVS holds desirable characteristics for this particular domain. Its flexible notation enables the possibility of performing behavioral exploration when denoting the properties to be satisfied. In addition, FVS expressive power capable of denoting ω -regular properties is useful to denote behavior in a higher level of abstraction. These are two key activities when specifying and validating a system architecture. Given that FVS specifications can be translated into Büchi automata they can be used as input in a validation tool like model checkers. In this sense, we conducted industrial relevant case studies to apply our approach in concrete examples.

Keywords: Software Architecture, Model checking, Declarative Specifications

1 Introduction

One of the most crucial concepts inherent to any Software Engineering activity is the usage of models and abstraction to be able to reason, explore and specify the expected behavior of the system to be built. In this sense, the design and analysis of Software Architectures constitutes a challenging and significant corner stone to achieve these objectives [8].

When specifying architectural behavior it is essential being able to explore and reason about different alternatives since most of the requirements are still yet to be discovered and defined. This is particularly true when trying to formally validate the expected behavior by performing architectural model checking. A model checker is given as input a model abstracting the system to be developed and a set of properties that the model should satisfy. The topic of formal languages to describe architectural behavior has been tackled by a plethora of approaches such as [13, 22, 17, 19]. However, it has been pinpointed by the community that writing the expected behavior in the form of properties is still one of

^{*} This work was partially funded by UNDAVCYT 2014, PAE-PICT-2007-02278:(PAE 37279), PIP 112-200801-00955 and UBACyT X021, UAI-CAETI

the main challenges to be addressed [2, 20, 1, 3]. Some of the most relevant issues included in this challenge are: the usage of formal languages such as temporal logics which may be hard to adopt, the usage of operational notations such as Automata-based notations or ADL's (Architecture Description Languages) whose constructors and structure resembles source code that might lead to premature implementations decisions and a certain lack of flexibility and expressive power in the notations used.

Given this context we explore in this work the Feather Weight Visual Scenarios (FVS) language as a declarative language to denote and validate architectural behavior. FVS is a declarative language based on graphical scenarios and features a flexible and expressive notation with clear and solid language semantics. FVS expressivity is a distinguished characteristic among declarative approaches since it is able to denote ω -regular properties. This makes FVS, for example, more expressive than LTL (Linear Temporal Logic). This fact enables the user to predicate and express behavior in a higher lever of abstraction which is a key factor in the architectural domain. FVS's specifications can be translated into Büchi automata enabling the possibility of realizing architectural model checking. We explore this alternative by analyzing industrial relevant case studies.

1.1 Previous work and new contributions

In [5, 6] FVS was used in the software architectures domain, denoting variability among the specification of product family architectures and expressing the behavior of architectural connectors. We now build on the top of those works introducing the following new aspects:

- * Architectural behavior specifications include not only connectors, but also components and other relevant architectural interactions.
- * We take FVS's specifications one step further by using them as input in a known model checker: LTSA (Labeled Transition System Analyzer [14]). In this sense, we present a tool denominated GTxFVS implementing FVS features and enabling the interaction with the LTSA model checker.
- * We exploit FVS expressive power to denote ω -regular properties to reason in a higher level of abstraction. In particular, we use this FVS's power to denote behavior imposed by the layered architectural style.
- * We introduce relevant case studies exhibiting FVS performance within the context of architectural model checking.

The rest of this work is organized as follows. Section 2 introduces FVS's main features. Section 3 describes the analyzed case studies. Section 4 presents some lessons learned when developing the case studies and summarizes conclusions of the work. Finally, Section 5 briefly discuss related work and raises some points regarding future work.

2 Feather weight Visual Scenarios

In this section we will informally describe the standing features of FVS [3, 4]. The reader is referred to [4] for a formal characterization of the language. FVS is a

graphical language based on scenarios. Scenarios are partial order of events, consisting of points, which are labeled with a logic formula expressing the possible events occurring at that point, and arrows connecting them. An arrow between two points indicates precedence of the source with respect to the destination: for instance, in figure 1-(a) A -event precedes B -event. We use an abbreviation for a frequent sub-pattern: a certain point represents the next occurrence of an event after another. The abbreviation is a second (open) arrow near the destination point. For example, in figure 1-b the scenario captures the very next B -event following an A -event, and not any other B -event. Conversely, to represent the previous occurrence of a (source) event, there is a symmetrical notation: an open arrow near the source extreme. Events labeling an arrow are interpreted as forbidden events between both points. In figure 1-c A -event precedes B -event such that C -event does not occur between them. Finally, FVS features aliasing between points. Scenario in 1-d indicates that a point labeled with A is also labeled with $A \wedge B$. It is worth noticing that A -event is repeated on the labeling of the second point just because of FVS formal syntaxis.

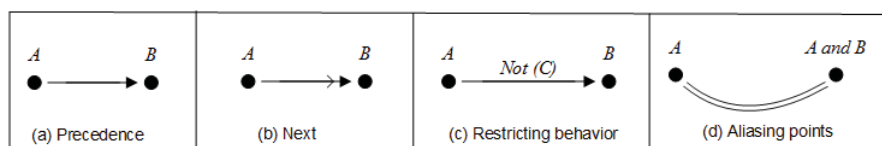


Fig. 1. Basic Elements in FVS

We now introduce the concept of FVS rules, a core concept in the language. Roughly speaking, a rule is divided into two parts: a scenario playing the role of an antecedent and at least one scenario playing the role of a consequent. The intuition is that whenever a trace “matches” a given antecedent scenario, then it must also match at least one of the consequents. In other words, rules take the form of an implication: an antecedent scenario and one or more consequent scenarios. Graphically, the antecedent is shown in black, and consequents in grey. Since a rule can feature more than one consequent, elements which do not belong to the antecedent scenario are numbered to identify the consequent they belong to. An example is shown in figure 2. The rule describes the circumstances under which writing in a pipe is valid. For every occurrence of a write event, then it must be the case that either the pipe did not reach its maximum capacity since it was ready to perform (Consequent 1) or the pipe did reach its capacity, but another component performed a read over the pipe (making the pipe available again) afterwards and the pipe capacity did not reach again its maximum (Consequent 2).

Ghosts Events and Translation into Büchi automata FVS is expressive enough to denote ω -regular properties [4]. This is due to the introduction of

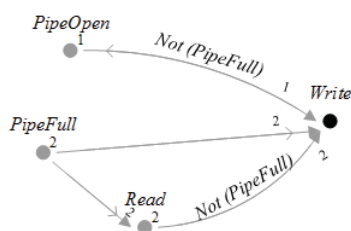


Fig. 2. An FVS rule example

abstraction, which is incorporated in our notation by introducing a new type of events. By using these events, the specifier can abstract behavior and reason about events that are not present in the system traces, but actually represent a higher level of abstraction. We call these special events as “ghost” events, in contrast with “actual” events, the set of events present in the system’s specification. In order to verify that a certain trace of the system (which only contains actual events) satisfies a rule containing ghosts events there is a internal procedure based on morphisms that discards ghost events based on a classic process of existential elimination [4]. Finally, there is a tableau procedure that translates FVS rules into Büchi automata [4]. In this way FVS specifications can play the role of input properties to be analyzed by a model checker.

FVS tool implementation: GTxFVS The tool GTxFVS basically implements the FVS language. It is based on the Meteor platform⁴, an open source platform for web, mobile, and desktop applications. GTxFVS also implements the tableau algorithm which translates FVS rules into Büchi automata and it allows the interaction with the LTSA model checker so that architectural model checking can be performed.

3 Case Study

In this section we describe the case studies illustrating our approach. Section 3.1 shows the formal validation of a known Publish/Subscribe based application whereas Section 3.2 exhibits FVS expressive power verifying properties in a layered system.

3.1 Verifying SIENA’s architectural behavior

We applied our approach within the context of a known Publish/Subscribe event notification service called SIENA [11] developed at University of Colorado. This system was analyzed in [9, 10] were a model representing a minimal schema of SIENA is given. In what follows we describe and verify the behavior of the

⁴ (<https://www.meteor.com/>)

SIENA system based on this latter work. Two main components highlight in the Siena's architecture specification: *clients* and *event-service*. The event-service is composed by a number of servers which offers clients the publish/subscribe interface. Clients are both publishers and subscribers. Subscribers express their interest in events by supplying a *filter*. We described, specified and verified three main properties of the SIENA model. It is worth to point out that these properties are presented in [10, 9] as essential to the behavior of the system. The mentioned properties are:

- Property 1: If a component *C0* subscribes a filter expressing interest in component *C1* publications' and *C1* publishes an event then *C0* receives the corresponding notifications unless *C0* unsubscribes its filter.
- Property 2: Servers must process events in the same order they were received.
- Property 3: Events received by a component which are generated by the same source maintain the publication order.

The FVS rule in Figure 3 specifies the behavior required by *Property 1*. The following events are involved: Event *ACT-FI* (component *C0*'s filter is active in the event-service structure), *PUB-C1* (component *C1* publishes an event), *NOT-C0* (component *C0* receives an event notification) and *UNS-C0* (component *C0* unsubscribes the filter). The rule simply states that whenever *C1* publishes an event given that *C0*'s filter is active, then *C0* receives its notification. The condition in the rule scenarios (*Not UNS-C0*) checks that component *C0* did not unsubscribe the filter.

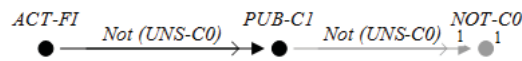


Fig. 3. All the publications must be received by the subscribers

Rule in Figure 4 shows the behavior denoted in *Property 2*. For this particular case we exhibit the rule considering only messages for subscribing and activating filters: the filters must be activated accordingly to the time their subscription was received. Similar rules are defined for other kind of messages. The rule says that when two filter subscriptions are received in a server (*SUB-Filter0* first and *SUB-Filter1* afterwards) and the second one was activated (*ACT-Filter1*) then it must be the case that the first filter was activated before (*ACT-Filter1*).

Finally, rule in Figure 5 illustrates the behavior imposed by *Property 3*. This is a particular case of the previous property (*Property 2* in Figure 4). In this rule the following events are involved: *SUB-C0* (component *C0* subscribes a filter), *ACT-FC0* (the filter for component *C0* is activated), *PUB-C1-E1* (component *C1* publishes event *E1*), *PUB-C1-E2* (component *C1* publishes event *E2*), *NOT-C0E1* (component *C0* receives the notification of event *E1*), *NOT-C0E2* (component *C0* receives the notification of event *E1*) and *UNS-C0* (component *C0* unsubscribes its filter). The rule demands that whenever two events are

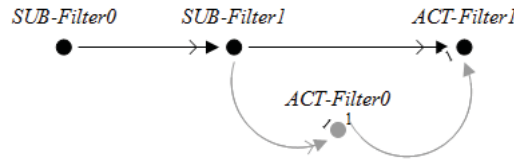


Fig. 4. Messages orders must be preserved

published by the same source then if the interested subscriber received the notification of the latter event and did not realize a unsubscription, then the subscriber received the notification of the former event before the occurrence of the notification of the latter.

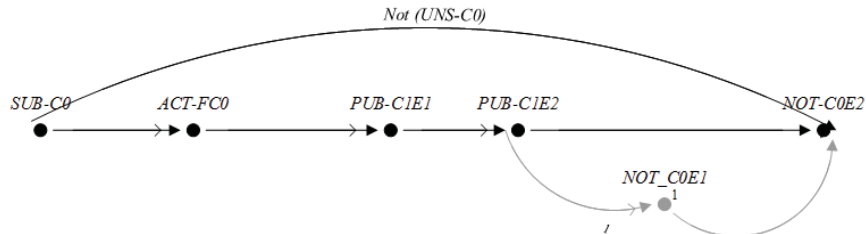


Fig. 5. Events generated by the same source respect the publication order

It is worth mentioning that all the rules were satisfied when employing the LTSA model checker to validate their behavior.

3.2 Verifying architectural conformance within Layered-based Systems

Inspired by the case of study introduced in [15] we show next how using *FVS's ghosts events* we can specify rules that verify that a system is satisfying the restrictions imposed by the layered pattern [8]. While services invocations are regular events present in the system's traces, the notion of layers lives in a higher level of abstraction. Given this context we introduce *ghost events* to capture the notion of layers. Events named *Layer-1, Layer-2, ..., Layer-N* represents the layers of the system. Once these events are defined, the user can predicate about architectural behavior based on these events. For example, the user can introduce rules that verify that services are only invoked from the immediate lower layer. The FVS rules in Figure 6 tackle the ghost layered events definition. For simplicity reasons, we show the definitions for a general schema with three services: *S-1, S-2* and *S-3* and three layers: *Layer-1, Layer-2* and *Layer-3*, where

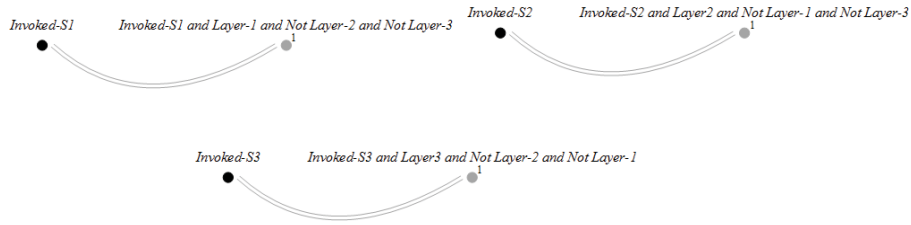


Fig. 6. Layered Ghost Events Definition

service $S-i$ belongs to the layer $Layer-i$. In few words, each service invocation is associated with the corresponding ghost event.

The FVS rules in Figure 7 reflect the behavior restrictions imposed by the layered pattern: each service can only be invoked by services in the immediate lower level. These rules are based only in the ghost layered events.



Fig. 7. Rules reflecting the behavior of a layered pattern

We applied this approach to a layered system based on the example introduced in [15]. Roughly speaking, the system contains four layers that behave in the following way: the Web client and the application client call the session beans, the session beans invoke the entity beans, and the entity beans access the database tables on the back end. We defined the corresponding ghosts events and specified layered-restrictions with FVS rules (similar to the ones in Figures 6 and 7 instantiating the general schema to the concrete system), we obtained a model of the system and checked whether its satisfied the given properties. Similar to the results exhibited in [15] we found violations within the layered structure, since sessions beans interacted directly with the database layer.

4 Lessons Learned and Conclusions

Based of the results obtained in the case studies we believe that the FVS language is suitable for specifying and validating architectural behavior. The operational flavour given by the translation of FVS rules into automata made possible the integration with model checkers, which in turn, enabled the possibility of realizing architectural model checking.

The flexility and expressiveness of the language do have an important impact when modeling and exploring architectural behavior. The properties specified for the *SIENA* system were expressed in [10,9] using temporal logic formulas although their approach uses another notation (a graphical language called

Property Sequence Charts). In addition, when specifying the formulas some auxiliary predicates were defined in order to simplify the specification. This might indicate that the comprehension of the formula may be a challenging task. This could become even more difficult if the property needs to be modified to adapt to a different context or compared against other possible formula describing an alternative solution. These activities are crucial since exploration and reasoning about behavior is fundamental when defining architectural behavior and interactions between the components of the system.

Consider, for example, that a more relaxed version of the *SIENA* system is considered, where messages can be processed by servers in any order, and not necessarily respecting the time they were received. The architect would then need to modify the requirements described in Section 3.1. For the FVS specification described by rules in Figure 4 only one modification is needed: the elimination of the precedence relationship between events *ACT-FILTER0* and *ACT-FILTER1*, which states that *ACT-FILTER0* must precede *ACT-FILTER1*. If this relationship is removed, then the rule simply state that both filters should be activated, but it does not impose any order of occurrence. The new version of the rule is shown in Figure 8. In the case of a temporal logic specification, the architect of the system might be faced against a complex formula whose modification is far from being trivial. The same analysis can be obtained when employing another notations such as structured ADL's or automata-based notations.

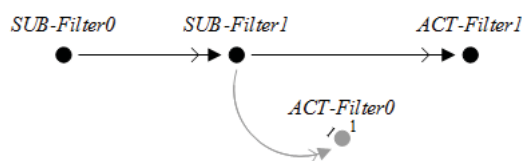


Fig. 8. A new rule modeling an architectural change in the system

The case study shown in Section 3.2 exhibits the benefits given by the expressive power of FVS. The use of ghost events makes a FVS a powerful language since ω -regular properties can be expressed, which is a distinguishable feature among declarative notations. In this case, ghost events are introduced so that the architect of the system can predicate in a higher level of abstraction and indicate interactions restrictions imposed by the layered architectural style.

Although the positive observations described in the previous paragraphs some threats to validity should be mentioned. First of all, a well designed experiment is needed to conclude with more precision about the flexibility and expressive power of our notation compared with other notations. FVS was compared against other notations regarding this issue in [4, 3] in the context of expressing properties describing early behavior, showing that is more suitable for expressing behavior and its validation. However, a new experiment is needed to validate this aspects in the architectural domain. This experiment is beyond the scope of this paper

and is addressed as future work. The development of the case studies is another threat to be considered. In all the cases we replicated experiments shown in the literature, perhaps easing the specification and analyses process. Therefore, new case studies from scratch are needed to further validate the obtained results. This clearly represent a challenging issue regarding feature work.

5 Related and Future Work

The first step regarding future work is to validate the results presented in this work with case studies with industrial relevance from scratch. This is a logical continuation of the experiments shown in this paper. We also would like to interact with architectural conformance between a system's implementation with respect to its architecture specification. This would involve to combine our approach with others techniques focused on extracting, either statically or dynamically, the architecture from the source code [21, 2]. Finally, we are interested in performing an experiment to measure and compare flexibility and expressiveness of FVS with other notations. A good starting point could be considering the comparison of formal architectural approaches made in [23].

Related work can be divided into approaches denoting graphical languages and approaches employing ADLs (Architectural Description Languages) based on operational or source-code flavoured notations. Among the first group, probably the most representative approach is Property Sequence Chart (PSC) [7]. PSC is a graphical language inspired in UML 2.0 Interaction Sequence Diagrams that has been applied in the software architecture domain [18, 10, 9]. In PSC, denoting complex constraints between events may require textual annotations. In addition, properties in PCS are described as anti-scenarios and not as conditional or triggered scenarios. PSC is less expressive than FVS since it can only describe a subset of LTL, whereas FVS is more expressive than LTL. Work in [17] employs a graphical notation based on UML together with a textual notation focusing on a model driven approach. Our approach is focused on expressing behavioral properties to be model checked. FVS features only graphical scenarios and a more rich and flexible triggering mechanism, since for example, the antecedent need not to precede the consequents in time. Work in [22] also employs an hybrid notation combining UML like scenarios with additional textual description based on OCL constrains. This approach is mostly focused on modeling behavior rather than its specification and validation.

Other approaches rely on different ADLs (Architectural Description Languages) [16, 19, 13] based on operational, textual or source-code flavoured notations. We believe the graphical and declarative notation of FVS might make it more suitable for early behavior exploration [20]. Since FVS scenarios can be translated into Büchi automata [4] it would be interesting to investigate if FVS can be combined with others ADLs such as the mentioned ones or others.

References

1. S. T. Albin. *The art of software architecture: design methods and techniques*, volume 9. John Wiley & Sons, 2003.
2. J. Aldrich, C. Chambers, and D. Notkin. Archjava: connecting software architecture to implementation. In *ICSE 2002*, pages 187–197. IEEE, 2002.
3. F. Asteasuain and V. Braberman. Specification patterns: formal and easy. *IJSEKE*, 25(04):669–700, 2015.
4. F. Asteasuain and V. Braberman. Declaratively building behavior by means of scenario clauses. *Requirements Engineering*, pages Vol 22,239–274, 2017.
5. F. Asteasuain and F. Tarulla. Modelado de comportamiento de conectores de software a travs de lenguajes declarativos. In *CONAIISI*, 2016.
6. F. Asteasuain and L. P. Vultaggio. Declarative and flexible modeling of software product line architectures. *IEEE Latin America Transactions*, 14(2):885–892, 2016.
7. M. Autili, P. Inverardi, and P. Pelliccione. A scenario based notation for specifying temporal properties. In *SCESM '06*, pages 21–28. ACM, 2006.
8. L. Bass, P. Clements, and R. Kazman. *Software Architecture in Practice*. Addison-Wesley Professional, 3rd edition, 2012.
9. M. Caporuscio, P. Inverardi, and P. Pelliccione. Compositional verification of middleware-based software architecture descriptions. In *ICSE 2004*.
10. M. Caporuscio, P. Inverardi, and P. Pelliccione. Formal analysis of architectural patterns. In *EWSA*, pages 10–24. Springer, 2004.
11. A. Carzaniga, D. S. Rosenblum, and A. L. Wolf. Design and evaluation of a wide-area event notification service. *TOCS*, 19(3):332–383, 2001.
12. P. Clements, D. Garlan, L. Bass, J. Stafford, R. Nord, J. Ivers, and R. Little. *Documenting software architectures: views and beyond*. Pearson Education, 2002.
13. J. L. Fiadeiro, A. Lopes, and L. Bocchi. A formal approach to service component architecture. In *WS-FM*, volume 4184, pages 193–213. Springer, 2006.
14. J. Magee and J. Kramer. *State models and java programs*. Wiley, 1999.
15. P. Merson. Using aspect-oriented programming to enforce architecture, software engineering institute. Technical report, CMU/SEI-2007-TN-019, 2007.
16. F. Oquendo, J. Leite, and T. Batista. Specifying architecture behavior with sysadl. In *WICSA 2016*, pages 140–145. IEEE, 2016.
17. F. Oquendo, B. Warboys, R. Morrison, R. Dindeleux, F. Gallo, H. Garavel, and C. Occhipinti. Archware: Architecting evolvable software. In *EWSA 2004*.
18. P. Pelliccione, P. Inverardi, and H. Muccini. Charmy: A framework for designing and verifying architectural specifications. *IEEE TSE*, 35(3):325–346, 2009.
19. T. K. Satyananda, D. Lee, and S. Kang. Formal verification of consistency between feature model and software architecture in software product line. In *ICSEA 2007*.
20. A. Van Lamsweerde. From system goals to software architecture. *Formal Methods for Software Architectures*, pages 25–43, 2003.
21. H. Yan, D. Garlan, B. Schmerl, J. Aldrich, and R. Kazman. Discotect: A system for discovering architectures from running systems. In *ICSE*, aosfpp 470-479, 2004.
22. U. Zdun and P. Avgeriou. Modeling architectural patterns using architectural primitives. In *ACM SIGPLAN Notices*, volume 40, pages 133–146. ACM, 2005.
23. P. Zhang, H. Muccini, and B. Li. A classification and comparison of model checking software architecture techniques. *Journal of Systems and Software*, 83(5):723–744, 2010.

Hacia una Meta-Herramienta de Análisis Automático de Modelos de Variabilidad

Esteban Ruiz de Galarreta *1, Matias Pol'la*12

Agustina Buccella*12 , Alejandra Cechich*1

1- *GIISCo, Facultad de Informática - Uncoma*

2- *CONICET, Consejo Nacional de Investigaciones Científicas y Técnicas*

Universidad Nacional del Comahue, Neuquén, Argentina

esteban.ruizdegalarreta@fi.uncoma.edu.ar, matias.polla@fi.uncoma.edu.ar,
agustina.buccella@fi.uncoma.edu.ar, alejandra.cechich@fi.uncoma.edu.ar

Abstract. El análisis automático de modelos de variabilidad es una actividad clave dentro de la gestión de la variabilidad. Existe un variado número de investigaciones y de enfoques que se han centrado en esta temática, fundamentalmente debido a que el hecho de determinar la validez de los modelos e identificar los problemas que contiene tanto a nivel de la definición de una línea de productos de software así como la instanciación y generación de productos en etapas tempranas, agiliza el desarrollo de la misma. Estos numerosos enfoques presentan diferentes tipos de modelos con distintas reglas, elementos y dependencias; diferentes procesos de validación y diversas reglas lógicas que soportan el análisis, así como diferentes solver (resolvidores lógicos) que determinan la validez de los modelos. Teniendo en cuenta este contexto, en este trabajo presentamos una Meta-herramienta de análisis automático basada en un proceso general, con el objetivo de proveer la suficiente flexibilidad que permita ser adaptada a diferentes modelos, reglas y resolvidores lógicos.

1 Introducción

La gestión de la variabilidad es una actividad dedicada a proporcionar flexibilidad y un alto nivel de reutilización durante el desarrollo del software. Dentro del enfoque de Líneas de Productos de Software (LPS), las actividades relacionadas con la gestión de variabilidad son las encargadas de proveer la flexibilidad necesaria para desarrollar un conjunto de aplicaciones similares basadas en un rango manejable de funcionalidades variables según las necesidades de los usuarios expertos. Dentro de la gestión de variabilidad, ha surgido un nuevo campo de investigación y estudio denominado análisis (automatizado) de variabilidad, centrado específicamente en la validación de modelos de variabilidad de acuerdo con un conjunto de problemas o errores predefinidos [2, 6, 10, 13]. Aunque la comunidad de investigación ha llegado a algún consenso sobre los escenarios base que deben evaluarse, el gran número de enfoques de modelado hace que la forma de evaluar estos escenarios sea investigada extensamente en la actualidad. La comprensión de los enfoques de modelado es fundamental para analizar las actividades de validación aplicadas a estos modelos. Por lo tanto, las actividades

de validación o el análisis de variabilidad deben considerarse desde estas perspectivas de modelado. Por ejemplo, el proceso de análisis debe incluir un conjunto diferente de tareas cuando la especificación se basa en un solo modelo [5] o cuando se representan diferentes modelos enlazados [3, 11]. Además, es importante considerar el lenguaje formal utilizado para traducir estos modelos de variabilidad porque éste es el punto de partida para realizar un análisis automatizado [2]. En la búsqueda de la automatización del proceso de verificación y validación de modelos de variabilidad, se han desarrollado diferentes herramientas que abordan la temática teniendo en cuenta características específicas de modelos de variabilidad particulares, como por ejemplo S.P.L.O.T. [9] y Variamos [8] entre otras. Teniendo en cuenta este contexto, en este trabajo se propone una meta-herramienta de análisis automatizado, basado en un proceso general de verificación presentado en [2]. Esta herramienta fue desarrollada con la flexibilidad necesaria para adaptarse a modelos de variabilidad, lógicas asociadas a los modelos y reglas particulares a las necesidades de cada usuario potencial.

Este trabajo se organiza de la siguiente manera: en la siguiente sección se detalla el marco teórico y trabajos relacionados. En la sección 4 se describe la meta-herramienta junto con su arquitectura, y el modelo de variabilidad asociado que permite su adaptación. En la sección 5 se detalla un caso de estudio en donde se instancia la meta-herramienta con nuestro propio modelo de variabilidad, reglas lógicas y escenarios de validación. Por último se detallan las conclusiones y trabajos futuros.

2 Marco Teórico y Trabajos Relacionados

El objetivo principal de una LPS es proveer una plataforma común lo suficientemente flexible de manera que permita adaptarse a las diferentes necesidades de los diversos productos dentro del rango de requerimientos establecido. Dicha flexibilidad se logra mediante la identificación, definición y posterior configuración de la variabilidad. La misma se define en [1] como la *capacidad de un artefacto de software de ser configurado, adaptado, extendido o cambiado en un contexto específico*.

La gestión de variabilidad [11] involucra todas las actividades relacionadas con la variabilidad a lo largo de todo el ciclo de vida de una LPS, desde la especificación de requerimientos, modelado e implementación, hasta el testing de los productos derivados. Entre las mismas podemos incluir a la definición de la variabilidad junto con los puntos de variación, la gestión de los componentes variables y la resolución de la misma al momento de realizar el proceso de derivación.

Si nos centramos en análisis automático de la variabilidad, existen una gran variedad de trabajos y propuestas que abordan dicho tema, centrándose en actividades de validación específicas para la fase de ingeniería de dominio del desarrollo de LPS. Algunos enfoques que presentan surveys o revisiones se pueden encontrar en [2, 15]. Por ejemplo, en [15] los autores presentan una comparación de análisis automatizado usando Alloy [4] para modelos de características (Feature Model - FM). En [2] existe una revisión bibliográfica más amplia del análisis automático de FM, basado en un proceso general que define un conjunto de tareas para evaluar FMs. En la Figura 1 podemos ver una adaptación de este proceso, en la cual identificamos cinco componentes principales,

definiendo un proceso general que sienta las bases de la actividad de la verificación automática de modelos.

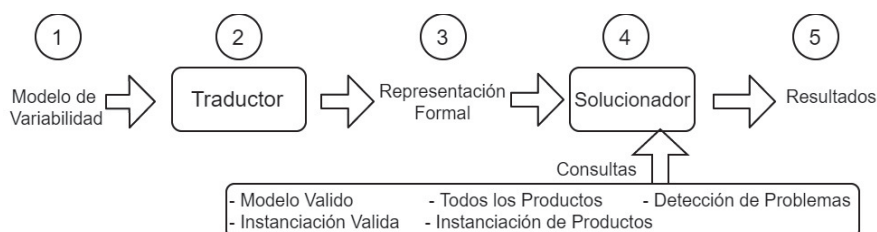


Figura 1. Proceso general para análisis automático de modelos de variabilidad

El primer componente es el *modelo de variabilidad* que puede definirse siguiendo cualquiera de los diferentes enfoques de modelado. El segundo componente es el *traductor*, que realiza el proceso de transformación entre el modelo y la representación formal teniendo en cuenta el enfoque seleccionado. El tercer componente es el *modelo formal* o representación formal, que incluye la variabilidad en términos lógicos. El cuarto componente es un *solver*, responsable de validar el modelo formal. Este componente también recibe el conjunto de escenarios o consultas de validación, que determinan los *resultados* (quinto componente) del proceso de análisis.

Centrándonos en el modelo de variabilidad, podemos encontrar un gran número de trabajos proponiendo diferentes enfoques para modelar la variabilidad. Entre ellos se destacan tres propuestas altamente referenciadas y utilizadas en el desarrollo de LPS: Modelo de características (FM)[5], Modelos de Variabilidad Ortogonal (Orthogonal Variability Model - OVM) [11] y Lenguaje de Variabilidad Común (Common Variability Language - CVL) [3]¹. Si bien la expresividad de cada modelo propuesto podría ser similar, existen diferencias sustanciales entre ellos. Las propuestas basadas en FM presentan un modelo jerárquico que permite describir los puntos variantes en común y las variabilidades dentro de una misma estructura. OVM permite especificar variabilidades en modelos separados pero vinculados (por medio de enlaces); y CVL es un lenguaje independiente del dominio para especificar la variabilidad, en el que los modelos se expresan en DSL (Domain Specific Language) o UML.

En cuanto a los enfoques que presentan soluciones para un proceso de análisis automatizado, durante los últimos años han surgido nuevos enfoques en este campo proponiendo novedosas ideas [7, 9, 13, 14]. Un resumen de algunas de estas propuestas se muestra en la Tabla 1 de acuerdo con cuatro aspectos: la herramienta soporte (S.P), el modelo de variabilidad (V.M.), el modelo formal (M.F.) y el solver. La columna de herramienta de soporte se establece como P cuando la herramienta es un prototipo, T cuando se trata de un complemento o herramienta de escritorio o WT cuando se trata de una herramienta web.

¹ Cada categoría contiene un gran número de propuestas, que abarcan diferentes aspectos

Enfoque	S.P.	V.M.	F.M.	Solver
<i>Lauenroth et al.</i> [7]	P	OVM	CTL	SAT-VM
		I-O/automata		
FAMA [14]	T	OVM -FM	CSP	BDD SAT
				Choco
<i>Metzger et al.</i> [10]	P	OVM - FD	CNF	SAT
VariaMos [8]	T	Independiente	CSP	SWI-Prolog
S.P.L.O.T. [9]	WT	FM	3-CNF	SAT

Tabla 1. Resumen de Enfoques de Analisis Automatico de Modelos de Variabilidad

Finalmente, otro aspecto importante a analizar es el conjunto de consultas que los solvers son capaces de responder. En este sentido, en la literatura existen varias definiciones del conjunto de anomalías o desajustes que pueden encontrarse en un modelo de variabilidad [2, 6, 16]. En estos trabajos se destacan consultas como modelo válido, instanciación valida y características muertas (Dead Feature).

3 Antecedentes

En trabajos anteriores [12] hemos definido un modelo de variabilidad propio, llamado SeVaTax, que toma como entrada modelos de variabilidad basados en primitivas OVM. En este trabajo, presentamos una meta-herramienta cuyo objetivo es poder ser adaptada a las diferentes características y particularidades que pueden presentar los diversos modelos de variabilidad. Como caso de estudio, tomaremos nuestro modelo SeVaTax para analizar la adaptación de la misma.

En términos generales el modelo SeVaTax está basado en OVM, que permite ser diseñado por medio de un conjunto de modelos interrelacionados llamados Datasheets; y está compuesto por un conjunto de puntos variables, dependencias y operadores de alcance. Dentro de los puntos de variabilidad utilizamos ObligatorioVP, OpcionalVP, AlternativoVP (una opción de un conjunto) y VarianteVP (al menos una opción de un conjunto). En cuando a las dependencias desarrollamos tres dependencias: usa (requiere mutuo), requiere y excluye. Por último presentamos dos operadores de alcance, que definen el alcance de cada uno de los puntos variantes (tipos de variabilidad) presentes en el modelo. Punto variante global determina que el punto variante sólo tendrá una sola configuración a en todo el modelo; y punto variante específico permite que el punto variante contenga diferentes configuraciones en diferentes casos. Además para nuestro modelo hemos definido una lógica que rige la validación automática basada en la lógica proposicional representada en CNF (Forma normal conjuntiva).

4 Meta-Herramienta de validación automática de Modelos de Variabilidad

La meta-herramienta, está basada en el metamodelo de la Figura 2 (A) que especifica los diferentes aspectos y características de cada uno de los elementos que conformarán

la instancia de la herramienta y que permitirán la posterior evaluación. Entre estos aspectos y características es necesario definir principalmente el modelo de variabilidad que será analizado junto con sus reglas de traducción.

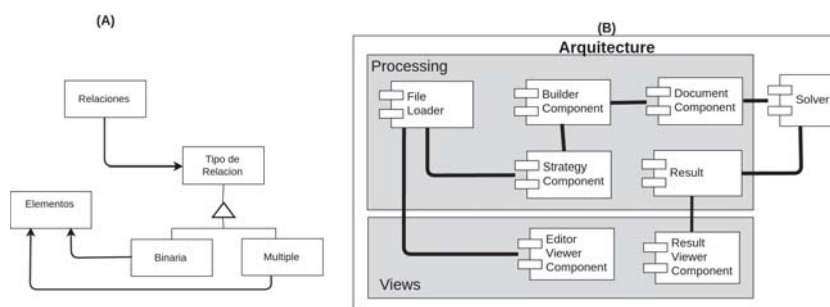


Figura 2. (A) Modelo estructural de la Meta-herramienta - (B) Arquitectura de la Meta-herramienta

En la Figura 2 (A) podemos visualizar los conceptos más básicos del meta-modelo utilizado para la definición de nuestra herramienta que fueron mencionados anteriormente. Estos conceptos son los de relaciones, su tipo y los elementos que componen a las mismas. Dichos tipos de relaciones, nos permiten que las mismas sean de dos tipos diferentes, uno es el tipo binario, que como bien dice su nombre, nos limita a una relación entre dos elementos de nuestra estructura. Por otro lado, existe la múltiple, que es la utilizada cuando se necesita relacionar más de dos elementos. Una vez entendidos estos conceptos básicos, se debería ser capaz de lograr una definición de un modelo requerido para el funcionamiento de la herramienta. Dicho diagrama, puede ser extendido a futuro para ampliar las características disponibles en el meta-modelo de nuestra herramienta,

En cuanto al modelo de variabilidad es necesario definir los elementos que lo conforman, las relaciones o dependencias pueden tener dichos elementos entre ellos (obligatorio, opcional, alternativa, variante, entre otras); y cuáles son sus posibles restricciones (uso, de requerimiento y de exclusión). De esta manera estaremos definiendo aspectos sintácticos y semánticos del modelo. Además es necesario definir cómo se debe comportar nuestra herramienta ante cada una de estas situaciones al momento del análisis de correctitud. Por lo tanto, se debe seleccionar y definir una traducción lógica de cada elemento, relación y restricción.

4.1 Proceso general

En esta sección explicaremos en detalle la arquitectura de la herramienta presentada en este trabajo. Como se mencionó anteriormente el objetivo de esta herramienta es presentar una estructura que se adapte a las particularidades de diversos modelos de variabilidad con diferentes elementos, relaciones y reglas para determinar la validez o no de cada modelo.

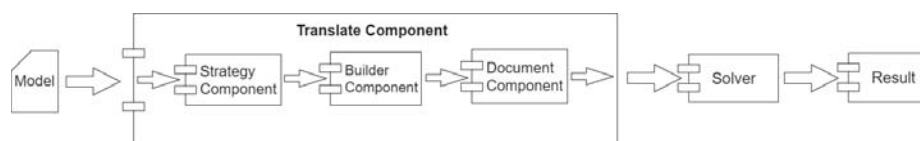


Figura 3. Proceso de análisis automático de la Meta-herramienta

La Figura 3 muestra un proceso general implementado en este trabajo. Este proceso debe ser lo suficientemente flexible como para adaptarse a diferentes configuraciones de distintos modelos de variabilidad. Para ello hemos tomado el proceso general de verificación automático propuesto en [2] como base para armar la meta-herramienta de validación. Como puede observarse en la Figura, este proceso está compuesto por cinco componentes y siete flujos de entrada/salida. Como se puede observar, la entrada principal es un modelo de variabilidad (conjunto de hojas de datos funcionales llamados datasheets particularmente para nuestro caso de estudio), que inician el proceso. Este modelo o conjunto de modelos es tomado por el “Traductor”, que está formado por 3 sub-componentes de software, que detallaremos a continuación:

- El primero de ellos es el componente ‘Strategy’ que se encarga de implementar el procedimiento que marcará las reglas utilizadas para traducir las entradas que posee un datasheet. Este, para cada punto variante de cada servicio, estipula el formato que se añadirá al archivo de resultado final.
- El componente ‘Builder’ recibe como entrada la estructura de las reglas dispuestas en el componente ‘Strategy’ y se encarga de representarlas en un documento.
- El componente ‘Document’ se encarga de diagramar el formato del resultado final de todo el proceso, estipulando la disposición final de cada regla generada con anterioridad en el componente ‘builder’. Esta disposición, está sumamente relacionada con lo que un componente ‘solver’ pueda necesitar.
- El componente ‘Solver’ es el encargado de que luego de resolver lógicamente un documento que contiene todas las traducciones de los modelos de variabilidad requeridos provisto por las etapas anteriores del proceso. Este componente, es el que nos marcara si dicha traducción es soluble o no, y si no lo es, qué reglas (y por lo tanto, sus servicios) están causando tal error.
- Por último, el componente ‘Result’ es el encargado de interpretar y darle un formato más amigable a la salida provista por el componente ‘solver’, para así simplificar datos redundantes o que normalmente no son necesarios en ciertos casos.

Para visualizar la flexibilidad buscada con esta meta-herramienta, hemos desarrollado un diagrama de variabilidad, diseñado a través de nuestras reglas de modelado (sección 3). En este sentido el diagrama de variabilidad para la arquitectura de la meta-herramienta, que determina cómo se pueden variar los componentes utilizados en cada parte del proceso. Dicho diagrama se puede observar en la Figura 4, donde se puede apreciar la variabilidad de cada componente explicado anteriormente. El diagrama de variabilidad está compuesto por 6 puntos variantes y 3 tipos de variabilidad (Obligatorio, Alternativo y Variante). Por ejemplo, como podemos observar, el punto variante

especificado para el componente del Solver, especifica que existe un punto variante entre un servicio llamado DL-Solver y otro denominado SAT-Solver. Así mismo, el servicio SAT-Solver tiene su propio punto variante, en este caso, es un punto variante alternativo, que nos indica que debemos instanciar dicho servicio con "Sat4J" o bien con "Choco".

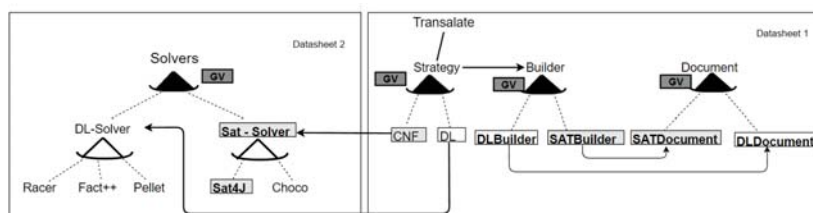


Figura 4. Modelo de Variabilidad de la Meta-Herramienta

Por medio de los servicios disponibles en el modelo de variabilidad podemos apreciar la existencia de implementaciones de distintas características para cada servicio, como por ejemplo, el componente/servicio de strategy, posee un punto variante global con los servicios de CNF (Forma normal conjuntiva) y DL (Lógica Descriptiva), que nos indica que estas son algunas de las dos posibles estrategias disponibles para aplicar a la hora de traducir una hoja de datos funcional. Además, dicho servicio, tiene una dependencia de uso con el servicio denominado "Builder". Esta relación lo que determina es que, cualquier servicio elegido del punto global de "Strategy", necesitará para su correcto funcionamiento, poder comunicarse con una de las tantas posibles instancias del componente "Builder". Así mismo, el modelo de variabilidad también nos indica que el servicio denominado "SATBuilder" que pertenece a la categoría de los "Builders", también tiene una relación de uso con su homónimo de la categoría "Document".

En nuestro caso, para la realización de la implementación propuesta, se optó solamente por implementar aquellos servicios que se encuentran coloreados con un gris de fondo. El resto de las implementaciones posibles, quedan como futuros trabajos a realizar.

4.2 Arquitectura de la Meta-herramienta

La Figura 2 (B) detalla la arquitectura de la meta-herramienta de validación, la cual está basada en una arquitectura genérica de cliente-servidor. Como podemos ver, la arquitectura se divide en 3 capas, la primera es determinar los componentes utilizados en la etapa de procesamiento, es decir, los componentes que se encontrarán del lado del servidor. Por otra parte, se visualizan los componentes que se encargará de mostrarle al usuario los resultados obtenidos en todo el proceso, estos, fueron demarcados en la zona de Views, tales como el componente encargado de ofrecernos una visual gráfica para la herramienta (Editor Viewer Component) y el ya mencionado componente que

nos ayuda a visualizar los resultados obtenidos del componente 'solver' (Result Viewer Component), donde los mismos, en nuestra implementación se alojarán del lado del cliente. Por último, en la arquitectura también utiliza un componente externo a nuestra implementación como es el caso del solucionador (solver).

5 Caso de estudio: Instanciación herramienta de validación

En esta sección explicaremos el proceso de instanciación realizado para adaptar nuestro proceso SeVaTax (introducido en la Sección 3) a la meta-herramienta. En la siguiente subsecciones explicaremos en detalle cada uno de los elementos instanciados.

5.1 Modelo SeVaTax

Para adaptar el modelo SeVaTax hemos instanciado el meta-modelo. Éste quedó compuesto por servicios de software como los elementos del modelo, relacionados por medio de los tipos de variabilidad de SeVaTax 3. Cabe destacar que la instanciación del meta-modelo puede ser definida en cualquier lenguaje de texto estructurado que brinde la capacidad para definir jerarquías de elementos. Nuestra instanciación fue definida utilizando la estructura provista por el formato de texto JSON (JavaScript Object Notation) para evitar futuros conflictos con la herramienta implementada. Una vez que se definió el modelo, definimos un conjunto de elementos de prueba o casos de test, que fueron la guía utilizada para el desarrollo de la instanciación de la herramienta.

5.2 Herramienta de Análisis Automatizado

Para el desarrollo de la instanciación de la herramienta, se utilizó el framework de Typescript "Angular 2". Teniendo en cuenta la tecnología seleccionada, hemos implementado cada uno de los componentes de arquitectura detallados en la figura 2 (B). Para esto, primero se implementó el componente "Editor Viewer Component" para que el usuario logre editar y brindar al servidor los datos necesarios para la traducción. El componente de "File Loader", fue implementado de manera de descomponer los documentos JSON mediante la búsqueda de los diferentes puntos variantes de los diferentes servicios. Para cada uno de estos puntos variantes, el componente identifica la información de los mismos y la prepara para el componente "Strategy". Este último realiza el análisis de cada elemento perteneciente al modelo definido, reconociendo entre otras cosas el tipo de variabilidad (Mandatario, opcional, alternativo o variante). Luego, procede a traducirlo aplicando las reglas de traducción especificados en la estrategia requerida. Para nuestro experimento, se utilizó la estrategia CNF propias del modelo SeVaTax. Finalmente, se implementó el "Builder", que se encarga de recibir los resultados y transformarlos de manera tal que se cumpla lo estipulado por el componente "Document". Este último componente fue implementado de manera de estructurar un documento con el formato específico de CNF para SAT4J.

5.3 Validación y Experimentación

Una vez implementada la instanciación de la herramienta, se desarrollaron diferentes casos de prueba para evaluar individualmente cada una de las dependencias del modelo utilizado, junto con un caso de prueba que combina todas las dependencias y relaciones del modelo. Cada uno de estos casos de test fueron ejecutados en la instanciación de la herramienta, y otorgaron resultados positivos en cuanto a la correctitud de la solución, que fue el único criterio evaluado en esta instancia. Por lo tanto, podemos remarcar que la Meta-Herramienta implementada se adaptó correctamente al modelo SeVaTax sin mayores complicaciones. Esto, nos brinda un panorama alentador con respecto a posibles mejoras en la misma, ya que la totalidad de los experimentos efectuados con la herramienta, entregaron los resultados correctos, lo que nos permitirá adicionar mayor complejidad y robustez al validador, a través de la utilización de diferentes modelos de variabilidad, validadores y reglas de análisis.

6 Conclusiones

En este trabajo hemos presentado una Meta-herramienta de análisis automático de variabilidad la cual nos permite adaptar diferentes modelos y reglas, generando diversas instancias acorde a diferentes enfoques y particularidades. En el mismo se demuestra la capacidad de adaptación de la misma, por medio de un caso de estudio basado en el modelo de variabilidad SeVaTax desarrollado anteriormente.

Como trabajo futuro, se buscará ampliar la gama de modelos de variabilidad instanciados por la herramienta, de forma de darle una mayor flexibilidad y adaptación a los diferentes enfoques. Además estamos trabajando en la traducción a lógicas DL de nuestro modelo de variabilidad SeVaTax. Esto nos abrirá la posibilidad de configurar diferentes solvers y evaluar la adaptabilidad de la meta-herramienta a este conjunto de características.

Bibliografía

- [1] Felix Bachmann and Paul Clements. Variability in software product lines. Technical Report CMU/SEI-2005-TR-012, Software Engineering Institute, Carnegie Mellon University, Pittsburgh, PA, 2005.
- [2] David Benavides, Sergio Segura, and Antonio Ruiz-Cortés. Automated analysis of feature models 20 years later: A literature review. *Inf. Syst.*, 35(6):615–636, September 2010.
- [3] Ø. Haugen, B. Møller-Pedersen, J. Oldevik, G. K. Olsen, and A. Svendsen. Adding standardized variability to domain specific languages. In *2008 12th International Software Product Line Conference*, pages 139–148, Sept 2008.
- [4] Daniel Jackson. *Software Abstractions: Logic, Language, and Analysis*. The MIT Press, 2006.
- [5] K. Kang, S. Cohen, J. Hess, W. Nowak, and S. Peterson. Feature-Oriented Domain Analysis (FODA) Feasibility Study. Technical Report CMU/SEI-90-TR-21, Software Engineering Institute, Carnegie Mellon University Pittsburgh, PA., 1990.

- [6] Matthias Kowal, Sofia Ananieva, and Thomas Thüm. Explaining anomalies in feature models. In *Proceedings of the 2016 ACM SIGPLAN International Conference on Generative Programming: Concepts and Experiences*, GPCE 2016, pages 132–143, New York, NY, USA, 2016. ACM.
- [7] K. Lauenroth, K. Pohl, and S. Toehning. Model checking of domain artifacts in product line engineering. In *2009 IEEE/ACM International Conference on Automated Software Engineering*, pages 269–280, Nov 2009.
- [8] Raul Mazo, Juan C. Munoz-Fernandez, Luisa Rincon, Camille Salinesi, and Gabriel Tamura. VariaMos: an extensible tool for engineering (dynamic) product lines. In *Proceedings of the 19th International Software Product Line Conference*, pages 374–379. ACM, 2015.
- [9] Marcilio Mendonca, Moises Branco, and Donald Cowan. S.p.l.o.t.: Software product lines online tools. In *Proceedings of the 24th ACM SIGPLAN Conference Companion on Object Oriented Programming Systems Languages and Applications*, OOPSLA '09, pages 761–762, New York, NY, USA, 2009. ACM.
- [10] A. Metzger, K. Pohl, P. Heymans, P. Y. Schobbens, and G. Saval. Disambiguating the documentation of variability in software product lines: A separation of concerns, formalization and automated analysis. In *15th IEEE International Requirements Engineering Conference (RE 2007)*, pages 243–253, Oct 2007.
- [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] M. Pol'la, A. Buccella, M. Arias, and A. Cechich. Sevatax: service taxonomy selection validation process for spl development. In *2015 34th International Conference of the Chilean Computer Science Society (SCCC)*, pages 1–6, Nov 2015.
- [13] Fabricia Roos-Frantz, José A Galindo, David Benavides, Antonio Ruiz Cortés, and J Garcia-Galán. Automated analysis of diverse variability models with tool support. *Jornadas de Ingeniería del Software y de Bases de Datos (JISBD 2014)*, Cádiz, Spain, page 160, 2014.
- [14] Fabricia Roos-Frantz, José A Galindo, David Benavides, and Antonio Ruiz-Cortés. Fama-ovm: a tool for the automated analysis of ovms. In *Proceedings of the 16th International Software Product Line Conference-Volume 2*, pages 250–254. ACM, 2012.
- [15] Anjali Sree-Kumar, Elena Planas, and Robert Clariso. Analysis of feature models using alloy: A survey. In *Proceedings 7th International Workshop on Formal Methods and Analysis in Software Product Line Engineering, FMSPLE@ETAPS 2016, Eindhoven, The Netherlands, April 3, 2016.*, pages 46–60, 2016.
- [16] T. von der Massen and H. H. Lichter. Deficiencies in feature models. In Tomi Mannisto and Jan Bosch, editors, *Workshop on Software Variability Management for Product Derivation - Towards Tool Support*, 2004.

Ingeniería de Requisitos centrada en la Estrategia CRM

Gladys Kaplan^{1,3}, Miriam Taboada¹, Gabriel Pousada², Claudia Litvak¹, Andrea Vera¹, Gabriel Blanco¹

¹Departamento de Ingeniería e Innovación Tecnológica, Universidad Nacional de La Matanza

²Departamento de Ciencias Económicas. Universidad Nacional de La Matanza
Florencio Varela 1903, B1754JEC San Justo, Buenos Aires

³Departamento de Ciencias Básicas. Universidad Nacional de Luján
Ruta 5 y Avenida Constitución, 6700 Luján, Buenos Aires
{gkaplan, mtaboada, gpousada, clitvak, avera, gblanco}@unlam.edu.ar

Resumen. Existen en el mercado diversos sistemas de software denominados CRM (Customer Relationship Management) que permiten gestionar todo tipo de relación organización-clientes. A pesar de esta gran variedad, las implementaciones de estos sistemas fallan con mucha frecuencia. En algunos casos, esto se debe a la creencia de que la mera incorporación de tecnología es suficiente para resolver los problemas internos de la organización. En el presente artículo se propone un mecanismo para resolver la desvinculación entre la estrategia CRM, las tareas realizadas y el software CRM. Se desarrolla una Ingeniería de Requisitos centrada en la estrategia CRM, que tiene por objetivos: i) modelar la estrategia CRM, ii) validarla y iii) determinar los servicios del software CRM particulares para cada organización. Para ello se generan los escenarios estratégicos que permiten modelar la estrategia CRM de una manera sencilla y fácil de comprender por todos los involucrados. También se construye un glosario del dominio para controlar la ambigüedad. Con estos modelos construidos se cuenta con el conocimiento necesario para identificar los servicios que son el nexo entre la estrategia CRM y el sistema de software más adecuado para la organización.

Palabras Clave: Estrategia de CRM, sistemas de software CRM, ingeniería de requisitos, escenarios estratégicos.

1 Introducción

Entre las primeras actividades que se realizan en las organizaciones se encuentra la determinación de la visión, misión y los objetivos [1], [2], [3], [4]. Estas definiciones son el eje para determinar la estrategia CRM [5], [6]. En muchos casos, para gestionar con éxito esta estrategia es necesario el uso de tecnología que permita satisfacer las demandas internas. Algunas organizaciones creen que la sola incorporación de la tecnología asegura el éxito. Sin embargo, las implementaciones de los sistemas de software CRM tienen un alto porcentaje de fracaso. Según el CRM Forum más del 30% de los casos de fracasos se deben a una pobre comprensión de la estrategia CRM o una deficiente planificación. Otro 30%, a un incremento en la brecha entre la misión y la estrategia CRM, siendo algunos ejemplos el crecimiento abrupto en las ventas que no fue acompañado por una estructura

organizacional adecuada; la falta de capacitación de los actores que intervienen en la gestión; no ser permeables a los cambios lógicos que impone el mercado, etc. Estos factores suelen ser la consecuencia de la vorágine de lo cotidiano que los aleja del ideal o del camino planificado. Si bien esto aplica a cualquier tipo de empresa, en general se ve con mayor frecuencia en las PYMES, que intentan resolver estos problemas implementando un sistema de software CRM, obligando a toda la organización a un tipo de comportamiento que no es el propio, sin alcanzar el objetivo propuesto y con un alto grado de frustración. En general, el éxito depende de retomar la estrategia CRM olvidada, volverla a analizar y recién después buscar un software que los ayude a gestionarla. Esta secuencia de trabajo es la propuesta central del presente artículo.

La existencia en el mercado actual de diversos sistemas de software CRM con un alto grado de parametrización, hacen encuadrar a estos sistemas en la definición de productos COTS [7], [8], los cuales restringen las posibilidades de cambios una vez adquiridos. Por lo tanto, una mala selección del sistema de software CRM hará prácticamente irreversible el fracaso de su implementación.

A los motivos de fracaso descriptos, se les suma un problema léxico con el acrónimo CRM. El mismo es utilizado, tanto por las organizaciones como por los proveedores de los sistemas informáticos, como sinónimo para referirse a la estrategia CRM y al sistema de software CRM. Este mal uso o uso parcial del término CRM, trae consecuencias no deseadas, como pensar que un sistema de software solucionará los problemas no resueltos en la organización. En este camino, se implementan sistemas de software CRM sin un previo análisis, induciendo a una incorrecta selección del mismo, una mala parametrización o una implementación antagónica a las expectativas y objetivos de la organización. Es claro entonces, que la selección del software no es arbitraria.

En el presente artículo se propone una alternativa para mejorar la selección e incrementar el éxito en las implementaciones de los sistemas de software CRM. La primera actividad es analizar la estrategia CRM para determinar con claridad los servicios que debe prestar el software. Para ello se debe realizar una Ingeniería de Requisitos (IR) [9] en el marco de los sistemas sociotécnicos [10], [11], centrada en la estrategia CRM. La salida de esta IR es un conjunto de servicios funcionales duros (SFD) (aquellos que no son negociables) y otro conjunto de servicios funcionales blandos (SFB) (aquellos que pueden ser revisados). Cabe resaltar que la estrategia CRM, en las empresas medianas y grandes está normalmente definida, mientras que en las PYMES suelen estar diluidas en tareas dispersas. Sea con lectura de documentos o entrevistas, se deben analizar todos los aspectos esenciales de la relación organización-cliente en función a los objetivos comerciales y definir puntos de acción para alcanzarlos. Para esto se generan los escenarios

estratégicos (EE), que se presentan en la sección 4, y son utilizados para modelar y validar la estrategia CRM. Estos escenarios han sido pensados en base al conocimiento de los escenarios futuros [12], adaptando su contenido a las nuevas necesidades. En la misma dirección, se pensó en un modelo que acerque al cliente al proceso de requisitos [13], [14], donde le sea fácil comprender el proceso y se sienta cómodo y dispuesto a aportar todo el conocimiento necesario. Los escenarios estratégicos utilizan el lenguaje natural y una representación coloquial, están semi-estructurados sólo con el objetivo de facilitar su escritura/lectura y su posterior trabajo de análisis. Los EE tienen embebidos todos los servicios de la estrategia CRM, existiendo un salto conceptual entre estas descripciones y los servicios de un software. Por lo tanto, es necesario transformar estas descripciones en servicios de software, para finalmente seleccionar el sistema más adecuado para la organización.

En la sección 2 se analiza el concepto de estrategia CRM, en la sección 3 se describe el estudio realizado para obtener una lista unificada de servicios CRM, en la sección 4 se presentan los escenarios estratégicos, en la sección 5 el mecanismo para modelar la estrategia CRM desde los EE y finalmente en la sección 6 se presentan las conclusiones y los trabajos futuros.

2 Estrategia CRM

Una estrategia CRM [15], [16], [17], [18] es un modelo de negocios cuyo fin es identificar y administrar las relaciones con los clientes, personalizando en cada una de ellas, de forma tal, de mejorar la efectividad sobre los mismos. En definitiva, es conseguir que los clientes sean fieles, eso supone conocerlos en su totalidad. Según Microsoft “una estrategia CRM pone al cliente en el centro de la estrategia empresarial. Una empresa que se vuelca hacia sus clientes es una empresa que utiliza la información para obtener una ventaja competitiva permitiendo su crecimiento y mejora de la rentabilidad maximizando el valor de cada cliente”.

Para emprender una estrategia CRM, se debe tener en cuenta algunos aspectos: Definición de visión y misión, objetivos en ventas, seguimiento de los clientes, medir y tener siempre presente la satisfacción del cliente, implementar una cultura de gestión corporativa a largo plazo, mentalidad de servicio al cliente entre los empleados para que estos estén satisfechos, alinear la filosofía de empresa a los objetivos de fidelización del cliente y generar valor en cada proceso.

Las estrategias CRM deben estar alineadas a la misión de la organización, una buena estrategia es una reflexión directa de la misión, soportándola en términos claros, directos y fáciles de entender. Deben estar orientadas al cliente, hacer foco sobre el posicionamiento e involucramiento que tiene que

existir en las relaciones con los clientes. También requiere del sponsor del máximo ejecutivo y el compromiso de los trabajadores, de forma visible, vocal y activa para que sea exitosa. Los procesos que se deben implementar deben ser iterativos, o sea que a medida que la organización evoluciona, lo hará la estrategia CRM. Su implementación trae un cambio de mentalidad en los integrantes de la organización la cual debe estar preparada para recibirla, ya que cambia la visión del negocio donde el cliente no es más una fuente de ingreso sino de oportunidades, transformando todos los contactos reactivos en contactos proactivos. El punto anterior trae aparejado ofrecer productos y servicios a la medida del cliente, logrando la plena satisfacción y lealtad de estos.

Para desarrollar una estrategia CRM se requiere tener presente los siguientes pasos:

- Fijar objetivos concretos, que sean cuantitativos y cualitativos,
- Determinar el valor agregado que se espera obtener,
- Delimitar el alcance del Proyecto,
- Incluir a todos los Departamentos involucrados,
- Tener un patrocinador,
- Designar un gerente de Proyecto,
- Definir roles y funciones de los trabajadores que participaran en el mismo.

En resumen, una estrategia de CRM, permite que la organización se centre en la atención de sus clientes, construyendo una relación duradera, identificando, comprendiendo y satisfaciendo sus necesidades, de tal modo, que evita que se vayan con la competencia.

3 Servicios CRM

Se ha realizado un estudio comparativo de los sistemas de software CRM existentes en el mercado actual, a fin de evaluar las características de los mismos y generar una tabla que ayude a identificar los servicios que ofrecen estos sistemas. Estas características han sido unificadas para conformar un listado único de categorías y servicios. Para realizar este estudio se han analizado los sistemas de software CRM con mayor presencia en el mercado actual: Salesforce, Oracle CRM, SAP CRM, Microsoft Dynamics, Sugar, Soho y NetSuite.

Cada una de las categorías de la Tabla 1, conforman el universo de tipos de servicios existentes en los sistemas actuales. A partir de cada categoría se han refinado los servicios para obtener una lista servicios potenciales unificados, como se muestra en la Tabla 2. Cuanto más refinada están estas tablas más exacta es la identificación de los servicios buscados.

Categorías Funcionales	Categorías de Integración	Categorías de Configuración/ Parametrización
1. Gestión de contactos y cuentas 2. Gestión de oportunidades 3. Gestión de actividades 4. Gestión de ofertas 5. Gestión de pedidos 6. Gestión de facturas 7. Gestión de contratos 8. Administración de comisiones 9. Campañas 10. Gestión de clientes potenciales/ leads 11. Administración de casos 12. Encuestas de satisfacción 13. Suite Business Intelligence 14. Workflows 15. Base de conocimiento 16. Servicios Mobile	1. Herramientas Colaborativas 2. Correo electrónico 3. Email Marketing (MailChimp) 4. Ecommerce (Magento, Clarity) 5. CMS Gestor de Contenidos (Wordpress) 6. ERPs 7. Contact Center (Zendesk)	1. Customización de campos 2. Customización de formularios 3. Flexibilidad para desarrollos 4. Soporta tecnología Web Services custom

Tabla 1 – Tabla de Categorías

Categoría		Servicios Potenciales
1. Gestión de contactos y cuentas	1.1	Registro unificado de la información del cliente y sus contactos.
	1.2	Visión 360 del cliente desde un sólo lugar (actividades realizadas, preferencias, oportunidades, ventas, reclamos, etc.)
	1.3	Gestión de mejora continua.
	1.4	Atención personalizada conociendo sus preferencias y comportamiento.
	1.5	Importación de datos desde distintas fuentes de datos.
	1.6	Exportación de los datos.
	1.7	Actualización masiva de datos.
	1.8	Creación de datos customizados.
	...	
2. Gestión de oportunidades	2.1	Administración de etapas de oportunidades de negocio
	2.2	Asignación a vendedores
	2.3	Seguimiento de oportunidades de negocio a través de la información registrada (pipeline)
	2.4	Estado Comercial actual de una oportunidad de venta
...		

Tabla 2 – Tabla de Servicios Potenciales Unificados (TSPU)

Por razones de espacio se muestran parcialmente ambas tablas, pero se han desarrollado completas para analizar el presente trabajo. Cabe destacar que estas tablas deben ser actualizadas cuando exista algún cambio en el mercado.

4 Modelado de la estrategia CRM con escenarios estratégicos

Como se mencionó en la Introducción, es necesario contar con la visión, misión y objetivos de la organización. Retornando al área de Marketing, se debe profundizar en la estrategia CRM donde, a partir de los objetivos, se debe generar la Tabla 3, detallando los sub_objetivos y las estrategias necesarias para alcanzarlos. Esta información se elicitó en entrevistas o desde documentos comerciales, siendo necesario, en ambos casos, validarla con los responsables del área de Marketing. A continuación se presenta en la Tabla 3 un ejemplo parcial de la tabla de Sub_objetivos- estrategias, de un caso real realizado a una Agencia de Turismo.

SUB_OBJETIVOS	ESTRATEGIAS
Seguimiento on line de las ventas	Registrar todas las actividades realizadas por los vendedores
	Actualizar el estado de la venta en tiempo real y todos sus datos.
Registro automático de clientes potenciales	Registrar desde su sitio web clientes potenciales en el sistema
	Integrar con otros sitios webs el sistema CRM para registrar clientes potenciales
Centralizar información entre sucursales	Registrar información de los clientes de distintas sucursales en el sistema
	Importar información existente de todas las sucursales

Tabla 3 – Ejemplo de la Tabla Sub_objetivo - estrategia

Puede observarse en la Fig 1 que la relación entre sub_objetivos y estrategias es del tipo (1:M). O sea, que un sub_objetivo se efectiviza con una o más estrategias.



Fig 1 – Relación entre sub_objetivos y estrategias

Avanzando en el análisis, por cada sub_objetivo - estrategia (SE) o fila de la Tabla 3, identifica un único escenario estratégico (EE). Como se observa en la Fig 2, la relación es del tipo (1:1).

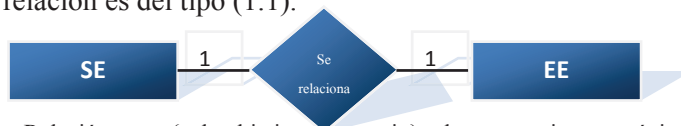


Fig 2 – Relación entre {sub_objetivo, estrategia} y los escenarios estratégicos

Con la Tabla 3 se está en condiciones de describir los EE. Este es el modelo central de todo el proceso. Se completan en entrevistas con el personal del área de marketing o desde documentos. En la Fig 3 se presenta un EE con un ejemplo del caso Agencia de Turismo.

Sub_Objetivo: **Seguimiento on line de las ventas**
 Estrategia: **Registrar todas las actividades realizadas por los vendedores**
 Negociable: **No** Si Nivel de Importancia: Alta Media Baja
 ACCIONES

- **Todos los vendedores deben registrar todas sus actividades relacionadas con el proceso de pre venta, venta y post venta.**
- **Cambiar la etapa de venta de cada oportunidad de negocio.**
- **Actualizar el estado del pago de cada oportunidad de venta.**
- **Acceder a informes o cuadros de mando**

Dirigido a: Personalizado: Grupos: **Vendedores**
 Datos que se desean guardar-conocer: **detalle de actividades (llamadas, reuniones, mails, etc.) realizadas por el vendedor como fecha, hora, nombre de vendedor, detalle de la actividad, próximo paso, etc.**
 Parámetro de medición de la estrategia: **10 días**
 Acciones post estrategia:

- **Reprogramar nuevas actividades (llamadas, mails, etc.)**
- **Mejorar o cambiar la atención al cliente**

Fig. 3 – Modelo de Escenario estratégico para el caso Agencia de Turismo

A continuación se detalla cada componente del EE:

Sub_objetivo y Estrategia: corresponde a cada fila de la Tabla 3.

Negociable: no son negociables aquellos sub_objetivos-estrategia que son servicios funcionales duros (SFD), indispensables para lograr el objetivo. Los negociables, por el contrario, son los servicios funcionales blandos (SFB), posibles de redefinir.

Nivel de Importancia: Solo aplica para los negociables. Indica como repercute en la estrategia CRM.

Acciones: tareas o acciones a realizar para alcanzar cada sub_objetivo-estrategia.

Dirigido a: Puede ser una persona/rol o grupos. En ambos casos se debe explicitar quiénes son.

Datos que se desean guardar-conocer: aquella información relevante para retroalimentar el proceso.

Parámetro de medición de la estrategia: debe contener la unidad cuantificable (cantidad, precio, etc.) y la unidad de medida (días, artículos, etc.) Ejemplo: 10 días o 20%

Acciones post estrategia: son las acciones que se deben realizar para potenciar la estrategia.

El EE es un modelo para representar y analizar la estrategia CRM. Permite definir una estrategia cuando no existe o validarla/repensarla/refinarla en caso

contrario. Cada componente del escenario está pensado para que el personal de Marketing se introduzca en los conceptos claves de la estrategia CRM, antes de adquirir un software. Cuando la estrategia existe, además permite detectar desvíos entre lo planificado y lo real.

5 Mecanismo para identificar los sistema de software CRM

A partir de los objetivos comerciales se desarrolla el presente mecanismo (ver Fig 4), teniendo en cuenta que dichos objetivos actúan como eje central de todo el conocimiento generado.

Las actividades propuestas son del tipo de las utilizadas para los productos COTS [7], [8] pero aplicados al análisis de una estrategia comercial. La primera actividad es generar la Tabla 3 de sub_objetivos-estrategias (ver sección 4). Con esta tabla completa y validada, se comienzan a describir los EE (ver sección 4). Durante esta descripción pueden aparecer cambios que se deben impactar en la Tabla 3. Una vez completos los EE, se construye el glosario con el vocabulario propio del dominio, denominado LEL de Requisitos [19]. Este glosario disminuye la ambigüedad de los EE. Con los escenarios y el glosario se está en condiciones de comenzar la búsqueda de los servicios. Para ello se analizan los componentes *Acciones* y *Acciones post estrategia* de cada EE, donde se encuentran embebidos los mismos. Hay que tener en cuenta que estos servicios no están explícitos. En este punto comienza el apareo entre estrategia CRM y software CRM. Los primeros servicios a determinar son los no negociables que corresponden al núcleo duro de la estrategia. En segunda instancia se busca satisfacer los servicios blandos, o sea, los negociables. En función a su nivel de importancia se negociará su forma de incluirse. Cada acción del EE se debe transcribir como un servicio CRM. Este pasaje es una interpretación semántica que realiza el ingeniero de requisitos para aproximar cada acción a uno o varios servicios de la Tabla de Servicios Potenciales Unificados (TSPU) (ver sección 3). La decisión de qué software adquirir estará dada por la mayor cantidad de coincidencias en estos servicios, la forma de resolución de los problemas, el precio, la integración con otros sistemas, la manera de resolver las características no funcionales, etc. En este último paso, se recomienda armar un pequeño comité con diferentes roles del área de Marketing, un especialista o conocedor de los software CRM existentes en el mercado actual y los ingenieros de requisitos.

6 Conclusión y trabajo futuro

La selección de un sistema de software CRM con un previo análisis de la estrategia CRM, tiene una alta probabilidad de éxito. Esto se debe a que después del análisis de los EE, quedaron claramente expuestas las necesidades de la organización

para con sus clientes, los desvíos producidos y no detectados, las funciones negociables y no negociables de la estrategia, etc. Estos escenarios ofrecen una mayor visibilidad sobre el direccionamiento de la estrategia y como se proyecta la toma de decisiones a partir de la relación con sus clientes. Esto determinó varios aspectos muy importantes a tener en cuenta para la selección de un software CRM. Para el presente artículo se realizaron dos casos reales completos (el de una agencia de turismo y el de una empresa que produce alfajores). En ambos casos se concluyó que el análisis sobre la estrategia CRM mejoraba significativamente la misma estrategia y daba seguridad al personal de Marketing, a la hora de revisar el mercado de sistemas de software existentes, teniendo en cuenta que la incorporación de tecnología debe dar soluciones y mejorar la toma de decisiones.

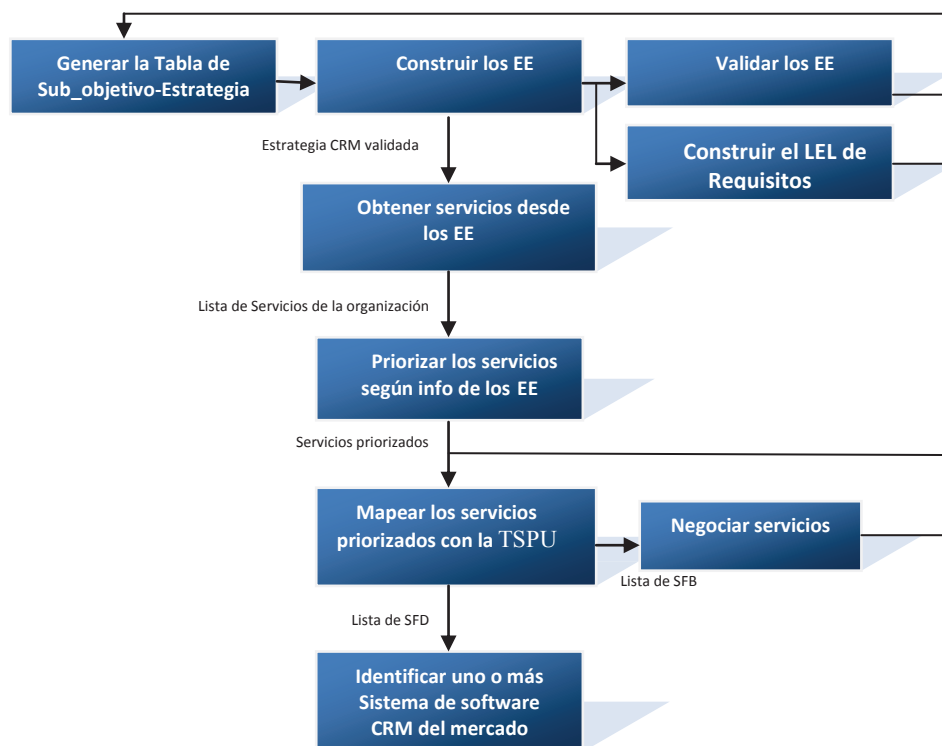


Fig. 4 – Ingeniería de Requisitos centrada en la estrategia CRM

Los trabajos futuros son varios. En este artículo se presentó la primera versión del mecanismo de selección. Si bien el proceso se ha estudiado completo, falta ahondar en algunas características particulares, como ser la incidencia de los requisitos no funcionales, profundizar en el mapeo y analizar las necesidades analíticas requeridas en la estrategia hasta identificarlas en el software. También se espera realizar más casos con el objetivo de probar estadísticamente el cubrimiento de los EE.

Referencias

1. Chang, W., Park, J. E., Chaiky, S.; "How does CRM technology transform into organizational performance? A mediating role of marketing capability", *Journal of Business Research*, 63(8), 849-855, 2010.
2. Rubio, R. "El camino hacia la excelencia en CRM". *Harvard-Deusto Marketing & ventas*, (59), 24-33, 2003.
3. Starkey, M., y Woodcock, N.; "CRM systems: Necessary, but not sufficient. REAP the benefits of customer management". *The Journal of Database Marketing*, 9(3), 267- 275, 2002.
4. Neil Woodcock, Bryan Foss, Merlin Stone; "The Customer Management Scorecard: Managing CRM for Profit"; ISBN-13: 978-0749438951, Kogan Page, 2003.
5. Finnegan, D. J., y Currie, W. L. "A multi-layered approach to CRM implementation: An integration perspective". *European Management Journal*, 28(2), 153-167, 2010.
6. Curry, A., y Kkolou, E. "Evaluating CRM to contribute to TQM improvement-a cross-case comparison". *The TQM Magazine*, 16(5), 314-324, 2004.
7. T. Ellis, "COTS Integration in Software Solutions – A Cost Model", *Systems Engineering in the Global Marketplace*, NCOSE Symposium, St. Louis, MO July 24026, 1995.
8. Lisa Brownsword, Carol A. Sledge, Tricia Oberndorf; "An Activity Framework for COTS-Based Systems", *Software Engineering Institute, CMU/SEI Report Number: CMU/SEI-2000-TR-010*, 2000.
9. 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 Ian Sommerville, "Ingeniería de Software", Capítulo 10, Pearson, 2011
10. Maté J.L., Silva A., "Requirements Engineering for Sociotechnical Systems", *Information Science Publishing*, ISBN: 1-59140-506-8, pp 2, pp 319, 2004.
11. Parviainen, P., Tihinen, M., van Solingen, R., "Requirements Engineering: dealing with the Complexity of Sociotechnical Systems Development", in *Requirements Engineering for Sociotechnical Systems*, J.L. Maté & A. Silva (eds.) , Chapter I, pp1-20, 2005.
12. Leite, J.C.S.P., Rossi, G., Balaguer, F., Maiorana, V., Kaplan, G., Hadad, G., Oliveros, A., "Enhancing a Requirements Baseline with Scenarios", *Requirements Engineering Journal*, Vol. 2, N° 4, 1997, pp. 184-198. 3.
13. Jackson, M., "Software Requirements & Specifications. A lexicon of practice, principles and prejudices", Addison -Wesley, Reading, MA/ACM Press, New York, 1995. 36 4.
14. Loucopoulos, P., Karakostas, V., "System Requirements Engineering", McGraw-Hill, London, 1995.
15. Hugo Brunetta, "Fundamentos para establecer una estrategia CRM", 2012. <http://www.redk.net/fundamentos-para-establecer-una-estrategia-crm-parte-1/>
16. Ignasi Vidal Díez, "Cómo conquistar el mercado con una estrategia CRM", ISBN: 84-96169-38-3, 2017.
17. Cosimo Chiesa de Negri, "CRM: Las 5 pirámides del marketing relacional: Cómo conseguir que los clientes lleguen para quedarse (MARKETING Y VENTAS)", Ed.DEUSTO, 2009
18. Pedro Reinares, "Cien errores del CRM", ESIC, 2016.
19. Gladys Kaplan, Jorge Doorn, Nora Gigante, "Evolución Semántica de Glosarios en los Procesos de Requisitos", CACIC 2013.

Modelo de Proceso de Gestión para Proyectos de Ingeniería del Conocimiento

Hernán Amatriain, Hernán Merlino, Sebastián Martins y Santiago Bianco

Laboratorio de Investigación y Desarrollo en Ingeniería de Software
Grupo de Investigación en Sistemas de Información.
Departamento de Desarrollo Productivo y Tecnológico. Universidad Nacional de Lanús
<http://www.unla.edu.ar/sistemas/gisi/>
hamatriain@gmail.com, hmerlino@gmail.com, smartins089@gmail.com,
santiago.bianco.sb@gmail.com

Resumen. Las reformulaciones de paradigmas metodológicos en Ingeniería de Software de los últimos 25 años relacionadas con los conceptos de: proceso, actividades, técnicas y metodologías; plantean la necesidad de una revisión sobre el uso de ellos en algunos de sus subcampos. En este contexto, se propone un Modelo de Proceso de Gestión para Proyectos de Ingeniería del Conocimiento que pone foco en la gestión de este tipo de proyectos.

Palabras clave. Ingeniería del Conocimiento, Sistemas Basados en Conocimientos, Sistemas Expertos, Modelo de Procesos, Control y Gestión de Proyectos, Metodología IDEAL, Metodología GROVER, Método BGM, Estándar IEEE 1074-1997, ISO/IEC 12.207, PMBOK.

1. Introducción

La Ingeniería de Conocimiento (INCO) tiene por misión adquirir, formalizar, representar y utilizar grandes cantidades de conocimiento de la más alta calidad y específicos de una tarea [; García-Martínez y Britos; 2004; García-Martínez et al, 2003; Gómez et al, 2001; Bermejo et al, 2002; Diez E, 2013] para el desarrollo de los sistemas Basados en Conocimiento (SSBBCC) o Sistemas Expertos (SSEE). Para este propósito, se han desarrollado varias metodologías, técnicas y herramientas para desarrollar SSBBCC [Grover, 1983; Brulé y Blount, 1989; Schreiber, 2000; Staab et al., 2001].

La INCO ha desarrollado una variedad de soluciones a problemas que requieren un abordaje distinto que el ofrecido por la Ingeniería de Software [Britos, 2001; Britos et al, 2004]. En los últimos años ha cobrado auge los Sistemas Basados en Gestión de Reglas de Negocios [Ibarra y Bazán, 2013; Kliegr et al, 2014; Parra et al, 2014], que son un tipo específico de SBC, con varias aplicaciones prácticas, por ejemplo en medicina [Martínez del Busto et al, 2012; Pacheco Cárdenas et al, 2015]; existiendo, incluso, una comunidad dedicada a ello [Club-BPM] que ha celebrado sus últimos congresos en Perú (2014) y España (2015).

Dentro de las metodologías establecidas para el desarrollo de SSBBCC y SSEE, la que se destaca desde su aparición en 1987 es la Metodología IDEAL, que se ha convertido en un estándar de facto [Britos P., 2001; Rossi, 2001; Rizzi, 2001; Bermejo, 2002; Ierache, 2001; Cao, 2003; Diez E., 2003; Gómez, 2003; Hossian, 2003].

La metodología IDEAL [Carrillo Verdum, 1987; Gómez et al., 1997; García-Martínez y Britos; 2004] consta de cinco fases: (i) identificación de la tarea, (ii) desarrollo de los prototipos, (iii) ejecución de la construcción del sistema integrado. (iv) actuación para conseguir el mantenimiento perfecto, y (v) lograr una adecuada transferencia tecnológica. Cada una de éstas fases se “subdivide” en distintas etapas.

La Metodología IDEAL se centra en: adquisición, conceptualización y formalización de conocimientos. Si bien esta metodología tiene etapas para el estudio de viabilidad, integración con otros sistemas y etapas de validación, carece de las etapas de gestión a nivel proyecto, y no prevé actividades específicas para la gestión de proyectos. Las actividades propias de gestión, en ésta metodología están embebidas en las mismas actividades de desarrollo, dejando su responsabilidad al mismo Ingeniero de Conocimiento que realiza el desarrollo. El Modelo de Ciclo de Vida que se utiliza conjuntamente con esta metodología es el modelo en espiral tronco-cónico [Alonso et al., 1996].

Otras metodologías tienen el mismo inconveniente: carecen de actividades propias de gestión y deja el control de las actividades de desarrollo al mismo Ingeniero de Conocimiento.

La metodología de Grover [Grover; 1983] se concentra en el la definición del dominio (conocimiento, referencias, situaciones y procedimientos) en la formulación del conocimiento fundamental (reglas elementales, creencias y expectativas) y en la consolidación del conocimiento de base (revisión y ciclos de corrección).

Otra metodología para el desarrollo de Sistemas Expertos o Sistemas Basados en conocimientos es el Método BGM [García-Martínez y Britos; 2004]. El mismo tiene las etapas de Adquisición de Conocimientos, Enunciación de Conceptos, Parametrización de Conceptos, Planteo de Causalidades, y Verificación.

Por otro lado, los modelos de proceso de software, o metodologías, que tienen en cuentas todas las actividades de control y gestión, como ser el Estándar IEEE 1074 [IEEE; 1998], la ISO/IEC 12.207 [ISO; 2008] o el PMBOK [PMBOK; 2013], carecen de aquellas actividades de gestión específicas para los Proyectos INCO.

En este trabajo presenta: el problema abordado (sección 2), la solución propuesta (sección 3), descripción de etapas y formalismos de salida (sección 4) y se plantean algunas conclusiones y futuras investigaciones (sección 5).

2. Descripción del Problema

Desde la Crisis del Software [Sommerville, 2005], se ha aprendido que la planificación de proyectos y su gestión y control son esenciales para el éxito de cualquier proyecto software. Es decir que la construcción de software requiere una visión ingenieril [Sommerville, 2005; Pressman, 2010], siendo muy factible un fracaso cuando así no se hace.

Los Proyectos INCO tienen un énfasis en las actividades de Adquisición de Conocimientos y Conceptualización, considerablemente mayor al resto de las actividades de desarrollo, ya que la construcción de reglas que van a constituir la Base de Conocimiento (pieza fundamental en la arquitectura de un SBC o SE, y ausente en cualquier otro tipo de sistema basado en software) es clave en el éxito del sistema. Por ello, las actividades de Adquisición de Conocimientos, Conceptualización de Conocimientos, y la verificación de Coherencia (de las reglas en la Base de Conocimientos), deben ser planificadas y monitoreadas de manera rigurosa. Las actividades que deberían desarrollarse para llevar a cabo esta gestión son específicas de estos sistemas (así como lo son las mismas actividades de desarrollo a gestionar) y siguen reglas propias.

Los procesos de y metodologías para gestión y control en Proyectos de Software existentes (IEEE 1074, ISO/IEC 12.207, PMBOK), carecen de estas actividades. Por otro lado, las metodologías para el desarrollo de SBC y SE, tienen en cuenta las actividades de desarrollo específicas, pero no la gestión y control de la misma, dejándolas, en el mejor de los casos, embebidas (solo algunas de ellas) en las mismas actividades de desarrollo sin una figura de gestor para la planificación y monitoreo, recayendo la responsabilidad en el mismo Ingeniero de Conocimiento que realiza el desarrollo.

La importancia de la Gestión y Control de los Proyectos de Software es indiscutible, y la inexistencia de actividades de gestión y control específicos en los procesos y metodologías actuales es un área que debe ser cubierta.

3. Solución Propuesta

Para afrontar el problema identificado, se propone un modelo de proceso para control y gestión de proyectos de Ingeniería del Conocimiento. Se establece para el proceso las fases, etapas, actividades, insumos y el producto obtenido que surge del desarrollo de cada actividad. También se desarrollan heurísticas para la realización de las mismas.

El proceso propuesto está compuesto por tres fases que se muestran en la Figura 1 y detallan en la Tabla 1.

En la Figura 1 se muestran las fases, etapas (en recuadros redondeados) y productos de cada etapa (rectángulos con esquina inferior derecha cortada), del proceso de control y gestión. La Tabla 1 muestra, además, el detalle de actividades a desarrollar por cada etapa del proceso de control y gestión.

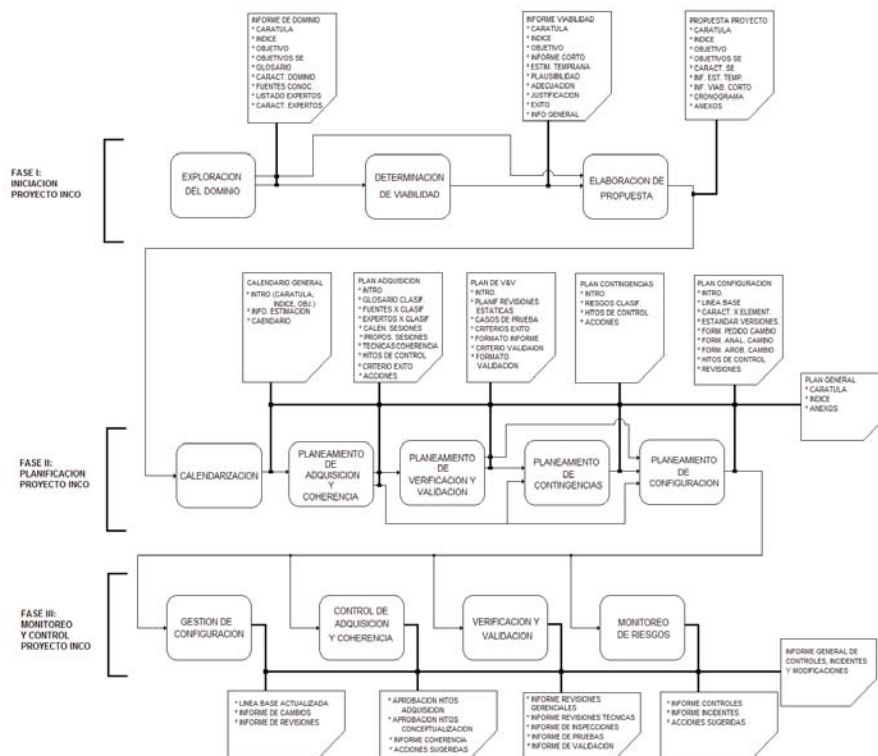


Figura 1. Proceso de Control y Gestión de Proyecto

Tabla 2. Proceso de Control y Gestión de Proyecto

PROCESO DE CONTROL Y GESTION PARA PROYECTOS INCO			
FASE I: INICIACION DEL PROYECTO INCO			
ETAPAS	ACTIVIDADES	INSUMOS	PRODUCTOS
EXPLORACION DEL DOMINIO	1) REALIZAR ENTREVISTA INICIAL 2) PLANIFICAR ENTREVISTAS CON USUARIOS Y EXPERTOS 3) PLANIFICAR OBSERVACION DE TAREAS 4) ANALIZAR ENTREVISTAS 5) PLANIFICAR NUEVO CICLO	•DOCUMENTACION ORGANIZACION	•INFORME DE DOMINIO •INFORME DE EXPERTOS
TESTEO DE VIABILIDAD	6) ESTIMACION TEMPRANA 7) TESTEAR PLAUSIBILIDAD 8) TESTEAR ADECUACION 9) TESTEAR JUSTIFICACION 10) TESTEAR ÉXITO 11) GENERAR INFORME GENERAL	•INFORME DE DOMINIO •INFORME DE EXPERTOS	•INFORME DE VIABILIDAD
ELABORACION DE LA PROPUESTA DE PROYECTO	12) DELIMITAR ALCANCE DEL PROYECTO 13) ESPECIFICAR CARACTERISTICAS DEL SE 14) GENERAR CALENDARIO GENERAL 15) ELABORAR PROPUESTA	•INFORME DE DOMINIO •INFORME DE EXPERTOS •INFORME DE VIABILIDAD	•PROPUESTA PROYECTO
PRODUCTO FINAL DE LA FASE: PROPUESTA DEL PROYECTO INCO			

FASE II: PLANIFICACION DEL PROYECTO INCO			
ETAPAS	ACTIVIDADES	INSUMOS	PRODUCTOS
CALENDARIZACION DEL PROYECTO INCO	16) SLECCIONAR DE METRICAS (HAUGE) 17) SELECCIONAR EL MCV (PROPONER ESPIRAL) 18) ESTIMAR EL ESFUERZO (OBEJERO) 19) DESARROLLAR MAPA DE ACTIVIDADES 20) DEFINIR HITOS 21) ESTABLECER CALENDARIO	•PROPUESTA DEL PROYECTO NCO	•MCV •INFORME DE ESTIMACION •MAPA DE ACTIVIDADES •CALENDARIO GENERAL
PLANEAMIENTO DE LA ADQUISICION Y COHERENCIA	22) FACTORIZAR EDUCCION 23) CALENDARIZAR CADA EDUCCION CON EXPERTO 24) SUGERIR TECNICAS DE EDUCCION 25) ESTABLECER PROPOSITO Y CRITERIO DE ÉXITO POR SESION 26) SELECCIONAR TECNICAS DE CONTROL DE COHERENCIA 27) ESTABLECER HITOS DE CONTROL 28) ESTABECER ROPOSITO Y CRITERIO DE ÉXITO DE CADA CONTROL DE COHERENCIA	•CALENDARIO GENERAL	•PLAN DE ADQUISICION Y COHERENCIA •CALENDARIO ACTUALIZADO
PLANEAMIENTO DE LA VERIFICACION Y VALIDACION	29) PLANIFICAR REVISIONES ESTATICAS 30) DEFINIR CASOS DE PRUEBA 31) ESTABLECER PROPOSITO Y CRITERIO DE ÉXITO 32) ESTABLECER FORMATOS INFORMES	•PLAN DE ADQUISICION Y COHERENCIA •CALENDARIO ACTUALIZADO	•PLAN DE REVISIONES •PLAN DE VERIFICACION Y VALIDACION •CALENDARIO ACTUALIZADO
PLANEAMIENTO DE CONTINGENCIAS	33) IDENTIFICAR RIESGOS 34) CLASIFICAR RIESGOS 35) ESTABLECER CRITICIDAD 36) ESTABLECER PROBABILIDAD DE OCURRENCIA 37) RANQUEAR RIESGOS 38) ESTABLECER ACCIONES PREVENTIVAS 39) ESTABLECER MEDICIONES DETECTIVAS 40) ESTABLECER PUNTOS DE CONTROL 41) ESTABLECER ACCIONES CORRECTIVAS	•PLAN DE ADQUISICION Y COHERENCIA •PLAN DE V & V •CALENDARIO ACTUALIZADO	•PLAN DE CONTINGENCIAS •CALENDARIO ACTUALIZADO
PLANEAMIENTO DE LA CONFIGURACION	42) ESTABLECER LINEA BASE 43) INFORMACION DE CADA ELEMENTO 44) ESTABLECER FORMATO DE PEDIDOS, ANALISIS Y APROBACION 45) PLANEAR REVISIONES	•PLAN DE ADQUISICION Y COHERENCIA •PLAN DE V & V •CALENDARIO ACTUALIZADO •PLAN DE CONTINGENCIAS	•PLAN DE CONFIGURACION •PLAN GNERAL DEL PROYECTO
PRODUCTO FINAL DE LA FASE: PLAN GENERAL DEL PROYECTO INCO			
FASE III: MONITOREO Y CONTROL DEL PROYECTO INCO			
ETAPAS	ACTIVIDADES	INSUMOS	PRODUCTOS
CONTROLES DE AVANCE Y GESTION DE LA CONFIGURACION	46) CONTROLAR ELEMENTO DE CONFIGURACION EN CADA HITO (MEDIR Y COMPARAR CON CRITERIO DE EXITO) 47) ACTUALIZAR INFORMACION DE TRAZABILIDAD 48) EJECUTAR PROTOCOLO DE CAMBIO 49) EJECUTAR REVISIONES	•PLAN GENERAL DEL PROYECTO	•LINEA BASE ACTUALIZADA •PLAN DE CAMBIOS APROBADOS •INFORME DE REVISIONES
CONTROLES DE ADQUISICION Y COHERENCIA	50) MEDIR EN CADA HITO EL ESTADO DE ADQUISICION Y COHERENCIA 51) COMPARAR EN CADA HITO EL RESULTADO DE LA MEDICION CON CRITERIO DE ÉXITO 52) ANALIZAR RESULTADO Y ESTABLECER SIGUIENTE CICLO 53) ESTABLECER REGLAS INCONSISTENTES 54) DETERMINAR SUGERENCI	•PLAN GENERAL DEL PROYECTO	•INFORME DE APROBACION DE ADQUISICION •INFORME DE APROBACION DE CONCEPTUALIZACION •INFORME DE CONSISTENCIAS Y SUGERENCIAS DE CAMBIO
VERIFICACION Y VALIDACION	55) EJECUTAR PRUEBAS DE VALIDACION 56) COMPARAR CON CRITERIOS DE ÉXITO PREVIAMENTE DEFINIDOS 57) DETERMINAR SUGERENCIAS 58) EJECUTAR VALIDACION 59) REALIZAR ANALISIS Y SUGERENCIAS	•PLAN GENERAL DEL PROYECTO	•INFORME DE VERIFICACION Y VALIDACION

MONITOREO DE RIESGOS	60) REALIZAR MEDICIONES EN HITOS 61) ESTABLECER OCURRENCIA DE INCIDENTE 62) DETERMINAR GRADO DE CRITICIDAD 63) SUGERIR ACCIONES A TOMAR	• PLAN GENERAL DEL PROYECTO	• INFORME DE CONTROLES • INFORME DE INCIDENTES • ACCIONES A TOMAR
PRODUCTO FINAL DE LA FASE: INFORME GENERAL DE CONTROLES, INCIDENTES Y MODIFICACIONES			

4. Descripción de Etapas y Formatos de Salida

El modelo de proceso propuesto como solución para la gestión de Proyectos INCO se encuentra dividido en tres fases: Fase de Iniciación del Proyecto, Fase de Planificación y Fase de Control, como queda detallado en la Figura 1 y la Tabla 1.

La Fase de Iniciación, comprende todas aquellas actividades que deben desarrollarse al principio de cualquier Proyecto, incluyendo el Test de Viabilidad [García-Martínez y Britos; 2004], finalizando con la Propuesta de Proyecto. Esta fase se compone de tres etapas: Exploración del Dominio (se realiza la primera aproximación al SBC a construir introduciendo el léxico del dominio y sus características), Test de Viabilidad (fundamental para seguir adelante con el Proyecto), y Elaboración de la Propuesta (donde se realiza una primera calendarización de las actividades a desarrollar).

Durante la Fase de Planificación se realizan todos los planes que deberán ejecutarse durante el proyecto, y que están relacionados con la gestión del mismo. Las etapas de esta fase son cinco: Calendarización del Proyecto (debe realizarse una estimación del proyecto utilizando [Obejero; 2006], seleccionar un MCV para lo que será necesario escoger las métricas adecuadas [Hauge et al; 2006], y elaboración del calendario), Planeamiento de la Adquisición, Conceptualización y Coherencia (estas actividades son específicas del desarrollo de SE y no figuran en ningún modelo de proceso de gestión de Ingeniería de Software y en las metodologías de construcción de SBC están, a lo sumo, embebidas en las mismas actividades de desarrollo y a cargo del mismo Ingeniero de Conocimiento; las actividades de gestión incluyen la factorización de los conocimientos a través de una categorización del glosario del dominio, programar las sesiones estableciendo los objetivos de cada una y criterios de éxito; otra actividad específica de los SBC y SE es el control de consistencias de las reglas, ya que la Base de Conocimiento es una pieza fundamental de la arquitectura de estos sistemas, por lo que es necesario mantener la consistencias de las nuevas reglas agregadas), Planeamiento de Verificación y Validación (incluye las actividades de seleccionar cuidadosamente con los expertos los casos de prueba y criterios de éxito; también deben definirse los formatos de los informes), Planeamiento de Contingencias (este plan es común a todos los proyectos de Ingeniería de Software, aquí se agregan heurísticas a tener en cuenta para la categorización de riesgos específicos de los sistemas INCO), y Planeamiento de la Configuración (otro plan común en los proyectos de Ingeniería de Software, pero no expresado en las metodologías de construcción de SE y SBC, agregando heurísticas de revisiones y versionados). Esta fase tiene como producto final el Plan General del Proyecto INCO.

La Fase final es la de Monitoreo y Control, donde se realizan los controles de las actividades de desarrollo según lo establecido en la fase de planificación (hitos de control), realizando las mediciones correspondientes según las métricas definidas (también en la fase previa), llevando a cabo las acciones correctivas necesarias y

preestablecidas. Los formatos de los informes están definidos en el Plan General del Proyecto, y los productos de cada una de las etapas que conforman esta fase serán los informes correspondientes. Al final de esta fase se realiza un Informe General de Controles, Incidentes y Modificaciones conformado por un resumen de cada uno de los informes producidos durante cada etapa.

Las actividades de cada etapa se encuentran detalladas en la Tabla 1, y los productos de cada una de ellas son los que se detallan a continuación.

Informe de Dominio: (a) Carátula (Título, fecha realización, última revisión, responsable y revisor), (b) Índice, (c) Objetivo del Informe, (d) Objetivos del SBC/SE, (e) Glosario del dominio (léxico del dominio, con una categorización prematura identificando conceptos, atributos, valores y reglas), (f) Características del dominio (habilidades necesarias para manejarlo, tipo de tareas desarrolladas, sensibilidad de la información manejada, necesidad de procesamiento de información por expertos, profundidad y variabilidad de los conocimientos), (g) Fuentes de conocimientos (toda fuente que provea información del dominio, explicitando disponibilidad y accesibilidad a la misma), (h) Listado de expertos (nombres, contacto y disponibilidad), (i) Características de los expertos (sub área específica de experticia y facilidades de articular su experiencia).

Informe de Viabilidad: (a) Introducción (Carátula: título, fecha realización, última revisión, responsable y revisor; Índice; Objetivo del Informe), (b) Estimación temprana (basada en el informe de dominio y aplicación prematura de [Hauge et al; 2006]), (g) Resultado de la viabilidad (informe corto), (h) Informe detallado de Plausibilidad (según Test de Viabilidad en [García-Martínez y Britos; 2004]), (i) Informe detallado de Adecuación (según Test de Viabilidad), (j) Informe detallado de Justificación (según Test de Viabilidad), (k) Informe detallado de Éxito (según Test de Viabilidad), (l) Informe General.

Propuesta de Proyecto: (a) Introducción, (b) Objetivos del SE, (c) Características del SE (necesidades de disponibilidad, características del dominio, tipo de problemas que resuelve), (d) Informe de estimación temprana, (e) Informe de viabilidad corto, (f) Cronograma de actividades principales, (g) Anexos (Informe de Dominio; Informe de Viabilidad).

Calendario General: (a) Introducción, (b) Informe de estimación (según [Obejero; 2006]), (c) Mapa de Actividades a desarrollar, (d) Dependencia de tareas (técnicas sugeridas: Grafo PER y Matriz de Dependencias), (e) Calendario del Proyecto (técnica sugerida: Diagrama de Gantt).

Plan de Adquisición y Coherencia: (a) Introducción, (b) Glosario Clasificado y factorizado (agrupar por conceptos), (c) Identificación de fuentes por agrupación de conceptos, (d) Asignación de expertos por agrupación, (e) Calendario de sesiones por experto, (f) Propósito de cada sesión, (g) Técnicas de control sugeridas para coherencia (de acuerdo a cada inconsistencia posible de la Base de Conocimiento), (h) Listado de acciones a tomar según cada tipo de inconsistencia, (i) Hitos de control y criterios de éxito, (j) Formato de los informes de control.

Plan de Verificación y Validación: (a) Introducción, (b) Plan de revisiones estáticas (gerenciales, técnicas e inspecciones), (c) Casos de pruebas, (d) Criterios de éxito, (e) Formato de los informes de control, (f) Criterios de validación, (g) Formato de los informes de validación.

Plan de Contingencia: (a) Introducción, (b) Listado de riesgos clasificados (heurística sugerida: enumerar un riesgo por cada tipo de inconsistencia no detectada y asignar grado de seriedad dependiendo de la afectación a la BC; por ejemplo, las reglas conflictivas crean incoherencia y son críticas, las reglas incluidas en otras y las reglas redundantes aumentan innecesariamente el tamaño de la BC afectando la eficiencia por lo que son de criticidad moderada, y las reglas cíclicas pueden hacer entrar al SE en un bucle sin llegar a una solución si no se siguen otros criterios para evitarlo por lo que tienen criticidad entre alta y moderada), (c) Hitos de control y criterios de éxito, (d) Listado de acciones a tomar por cada incidente, (e) Formato de los informes de control.

Plan de Configuración: (a) Introducción, (b) Línea base (heurística sugerida: tomar como elementos de configuración cada documento de gestión, cada informe de educación y los informes de coherencia; tomar como elemento de configuración especial la BC), (c) Estándares de versionados y anotaciones (heurística sugerida: identificar separadamente los elementos de configuración consistentes en los informes de educación separados por la clasificación y agrupación de conceptos identificados del dominio), (d) Formato de pedido de cambios, (f) Formato de análisis de cambios, (g) Formato de aprobación de cambios, (h) Hitos de control y revisiones, (i) Formatos de revisiones.

Las etapas de la fase final (monitoreo y control) tienen como salida los informes de los controles realizados en los hitos establecidos en la fase anterior, al igual que los formatos de los mismos. Esta etapa tiene además un producto final que consiste en un resumen de todos los informes de control, que es el Informe General de Controles, Incidentes y Cambios, cuyo fin es documentar el proceso completo de desarrollo para posteriores análisis y mejoras eventuales del proceso aplicado.

5. Conclusiones y futuras líneas de investigación

La Metodología IDEAL, GROVER y BGM se centran en la construcción (análisis y desarrollo) de los SSBCC y SSEE, pero carecen de las actividades necesarias para la gestión y control del proyecto, teniendo a lo sumo alguna planificación embebida en las mismas actividades de desarrollo y a cargo del mismo desarrollador. La solución propuesta, que pertenece a una investigación en desarrollo, se adecúa más a los proyectos INCO, sobre todo para aquellos de mayor tamaño donde la falta de gestión permite predecir un seguro fracaso. La metodología IDEAL, además de carecer de las actividades de planificación, control y gestión, establece un modelo de ciclo de vida prototipado, al cual queda atada, de manera innecesaria.

La solución propuesta pone foco en la definición de procesos de control y gestión de Proyectos de Ingeniería de Conocimiento, partiendo de las Metodologías IDEAL y del Estándar IEEE 1074-1997, definiendo claramente cuáles son las actividades a realizar en cada fase, así como sus entradas (insumos) y salidas (productos).

Como futuras líneas de trabajo se prevé: [a] definir las distintas técnicas, y los procedimientos asociados, necesarias para el desarrollo de las actividades propuestas en el modelo de proceso, [b] mejorar las heurísticas de clasificación e identificación de riesgos y configuración, [c] desarrollar un modelo de ciclo de vida que tenga en

cuenta la cuantificación de los conocimientos y la verificación de la base de conocimientos, [c] establecer una etapa de planificación y una de control de personal en las fases de planificación y control del proceso de gestión, [e] definir un modelo de desarrollo libre de actividades de gestión y que articule el modelo de proceso para Proyectos INCO, y [f] completar la validación de la propuesta a través de la aplicación práctica del modelo propuesto en el presente trabajo en más casos testigos.

6. Financiamiento

Las investigaciones que se reportan en este artículo han sido financiadas parcialmente por el Programa PROMINF-UNLa-2015-2017 del Ministerio de Educación Argentina y por el Proyecto 33B170 de la Secretaría de Ciencia y Tecnología de la Universidad Nacional de Lanús.

7. Bibliografía

- Alonso, F., Juristo, N., Maté, J., Pazos, J. 1996. Software engineering and knowledge engineering: Towards a common life cycle. *Journal of Systems and Software*, 33(1): 65-79.
- Bermejo F., Britos P., Rossi B, García-Martínez R. 2002 Sistemas de Asistencia para la Configuración de Ventiladores OAF en Neonatología. *Revista del Instituto Tecnológico de Buenos Aires*. Volumen 28, Pag. 24-68
- Borrajo, D. 1993. *Inteligencia artificial: Métodos y técnicas*. Ed. Centro de Estudios Ramón Areces.
- Britos, P., 2001. Sistema de ayuda sobre Legislación Argentina en Riesgos de Trabajo. Tesis Maestría. Universidad Politécnica de Madrid e Instituto Tecnológico de Buenos Aires.
- Britos, P., Sierra, A., García-Martínez, R., 2004. Sistema de Asistencia Legal en Riesgos de Trabajo. *Journal Informática y Sociedad*. Vol. 1 N° 1 (2004), pág. 34-49. ISSN: 1668-3137.
- Brulé, J., Blount, A. 1989. *Knowledge acquisition*. McGraw-Hill, Inc.
- Carrillo Verdun, J. 1987. *Metodología para el Desarrollo de Sistemas Expertos*. Tesis Doctoral. Universidad Politécnica de Madrid.
- Club-BPM. *Business Process Management*. Centro Oficial del BPM. España y Latinoamérica. (<http://www.club-bpm.com/>, último acceso 30-05-2016).
- Diez, E.. 2013. Aseguramiento de la Calidad en la Construcción de Sistemas Basados en el Conocimiento Un Enfoque Práctico. *Revista Latinoamericana de Ingeniería de Software*, 1(5) 167-206, ISSN 2314-2642.
- Dramis, L., Britos, P., Rossi, B., García-Martínez, R. 2000 Verificación de Base de Conocimiento basada en Algebra de Grafos. *Proceedings del VI Congreso Internacional de Ingeniería Informática*. Pag. 729-738. Editado
- García-Martínez, R., 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. ISBN 987-1104-05-7.
- Grover, M. 1983. A Pragmatic Knowledge Acquisition Methodology. *IJCAI Vol. 83*: 436-438.
- Gómez, A., Juristo, N., Montes, C., Pazos, J. 1997. *Ingeniería del conocimiento*. Editorial Centro de Estudios Ramón Areces.

- Gómez, S. 2003. Sistema Informático Jurídico para la Individualización y Acuerdos sobre la Pena. Tesis de Magister en Ingeniería del Software. Centro de Ingeniería del Software e Ingeniería del Conocimiento. ITBA.
- Hauge, O., Britos, P., García-Martínez, R. 2006. Conceptualization Maturity Metrics for Expert Systems. IFIP International Federation for Information Processing, Volume 217, ed. M. Bramer, (BostonSpringer), Pag. 435-444.
- Hossian, A. 2003. Sistema de Asistencia para la Selección de Estrategias y Actividades Instruccionales. Tesis de Magister en Ingeniería del Software. ITBA.
- Ibarra, G., Bazán, P. 2013. Análisis y comparación de plataformas BRMS a través de una prueba de concepto. XV Workshop de Investigadores en Ciencias de la Computación. IEEE, 1998. IEEE Std 1074-1997. IEEE Standard 1074 for Developing Software Life Cycle Processes.
- Ierache, J. 2002. Sistema Experto para la Asistencia a la Toma de Decisiones en Centro de Información y Control Aéreo. Tesis de Magister en Ingeniería del Software. Centro de Ingeniería del Software e Ingeniería del Conocimiento. ITBA.
- ISO; 2008. ISO/IEC 12.207, Information Technology / Software Life Cycle Processes. International Organization for Standardization
- Kliegr, T., Kuchar, J., Sottara, D., Y Vojr, S. 2014. Learning business rules with association rule classifiers. In RuleML14 The 8th International Web Rule Symposium. Springer.
- Martínez del Busto, M., Núñez Pérez, L., Caballero Martínez, A., Moreno Montes de Oca, I., Boggiano Castillo, M., González González, L., Hernández Hernández, P., 2012. Aplicación médica para trasplante renal usando reglas de negocio. Revista Habanera de Ciencias Médicas 2012:11(1)176-184
- Nguyen, T., Perkins, W., Laffey, T., Pecora, D. 1987. Knowledge-base verification. AI Magazine, 8(2) 69-75.
- Ovejero, J. 2006 Estimación de Proyectos para Sistemas Basados en Conocimientos, Tesis Magister Ingeniería de Software, ITBA-Universidad Politécnica de Madrid.
- Pacheco Cárdenas, Y., Estévez Abrahantes, S., Martínez del Busto, M., 2015. Integración de un Sistema de Gestión de Reglas de negocio al flujo de trabajo “control de historias clínicas para el trasplante renal”. Revista Cubana de Informática Médica, vol. 7, no. 1. ISBN 1684-1859.
- Parra, J., Bazán, P., Garro, J., 2014. Una propuesta arquitectónica para integrar una herramienta BPMS y un sistema de gestión de reglas de negocio. XV Workshop de Investigadores en Ciencias de la Computación.
- PMBOK, 2013. A Guide to the Project Management Body of Knowledge. Project Management Institute.
- Pressman, R., 2010. Ingeniería del Software. Un Enfoque Práctico. Séptima edición. Capítulo 1. McGraw-Hill, Inc. ISBN: 978-607-15-0314-5.
- Rizzi, M. 2001. Sistema Experto Asistente de Requerimientos. Tesis de Magister en Ingeniería del Software. Centro de Ingeniería del Software e Ingeniería del Conocimiento. ITBA.
- Rossi, B. 2001. Sistema Experto de Ayuda para la Selección del Modelo de Ciclo de Vida. Tesis de Master en Ingeniería del Conocimiento. Universidad Politécnica de Madrid.
- Schreiber, G. 2000. Knowledge engineering and management: the CommonKADS methodology. MIT press.
- Sommerville, I., 2005. Ingeniería del Software. Séptima edición. Capítulo 1. Pearson Addison. ISBN: 84-7829-074-5.
- Staab, S., Studer, R., Schnurr, H. P., Sure, Y. 2001. Knowledge processes and ontologies. IEEE Intelligent systems, (1): 26-34.

Process Variability: concepts, approaches and its application on a model of Cloud BPM

Jose Martinez Garro¹, Patricia Bazan¹, Javier Diaz¹

¹ LINTI, Facultad de Informática, UNLP,
La Plata, Argentina
{josemartinezarro, pbaz, jdiaz}@info.unlp.edu.ar

Abstract. Business Process Management as a discipline has suffered several changes during the implementation of the execution and monitoring phases in the cloud model. Different strategies have been seen in terms of the implementation needed in order to gather information from the different nodes during process execution, and finally show the results seamlessly without the notion of a partitioned business process. Another aspect to introduce in this context is Process Variability, in terms of the changes suffered by a process model during its lifecycle, and how these changes affect the actual instances in execution. In terms of a cloud BPM implementation, Process Variability adds even more complexity during execution considering the different process portions, as well as during the gathering and monitoring phases. The main purpose of this work is to establish how the different aspects of a cloud BPM implementation with decomposed processes are affected by introducing concepts of Process Variability, both in execution as well as in the monitoring phase. To achieve this goal an analysis of some current bibliography and the main aspects of process variability management is accomplished.

Keywords: BPM, Cloud, Execution, Monitoring, Process Variability.

1 Introduction

BPM (*Business Process Management*) has presented a pronounced growth during the last years, provoking an exploration over different related technologies, such as the cloud model in terms of execution, or CEP (*Complex Event Processing*) for process monitoring. Another aspect that must be considered during process execution and monitoring is Process Variability. Most of the non-open source products have features to support process variability, but in terms of a single BPM node. Since the cloud orientation is nowadays a trend in mostly every software paradigm, including BPM, it is necessary to consider how the management of process variations could affect the process decomposition and execution in the cloud. Process variability refers essentially to the different variations that a process model could suffer during its lifecycle, and how these changes could affect the instances in execution, and in consequence the monitoring of them. In chapter 2 an analysis of the current status of process variability and formal verification is accomplished. After that, in chapter 3 a brief revision of the mechanisms needed for process variant handling are evaluated. In chapter 4 the requirements for a formal specification are presented, and the different

policies and actions in declarative and imperative variability are detailed. Finally in chapter 5 it is addressed the way on how all these previous concepts affect a concrete implementation of Cloud BPM presented in some previous works by the same authors of this work [22] [23][29].

2 State of the art: variability and formal verification

There are several related works ([1], [2], [3]) addressing these two concepts, especially when it comes to process models and version management. In terms of definition, a formal verification mechanism implies proving or disproving the correctness of a system model according to a formal specification using some formal methods of mathematics. When employing formal verification, a system model – often represented by a labeled transition system – is verified against a formal specification using logic enunciations. One approach towards formal verification is model checking. When this task is performed, a system model is automatically, systematically, and exhaustively explored while each state is verified for compliance with the formal specification. In this way, Business Process Verification is the act of determining whether a business process model complies with a set of formal correctness properties [1] [2] [3] [29].

2.1 Soundness

Business process correctness verification entails the verification of a set of basic properties such as reachability and termination. Reachability applied to a business activity requires an execution path from the starting activity to every other activity in the model. The termination property requires that all possible execution traces reach a final state. Business process soundness, a property originally proposed in the area of Petri Net verification, is known as the combination of these two properties adding a third one: the absence of related running activities at process termination (i.e., proper completion) [2] [4] [5] [6].

2.2 Compliance

Business process compliance aims to confirm that a business process adheres to a set of rules imposed on that process. Rules can, for example, be imposed upon a process by international regulations, national law, or internal business rules. Whereas soundness verification aims at the verification of a limited set of requirements like reachability, termination, and possibly proper completion – compliance verification requires the verification of a broad set of specifications [1] [3] [7] [8] [24].

2.3 Variability

BPM is evolving rapidly due to emerging mass customization and personalization trends, the need for adaptation to varying business and execution contexts, and a wider availability of service-based infrastructures. Variability is an abstraction and management method that addresses a number of related issues, especially in a cloud environment where different models are deployed constantly over different nodes of the architecture. Variability can be introduced to the BPM area by using imperative or declarative approaches. Whereas imperative approaches exactly specify possible changes, declarative approaches constrain the process behavior, allowing any change within those constraints. In addition, since imperative approaches exactly specify all possible changes, they require all of them to be known in advance. Conversely, declarative approaches do not require such knowledge [9] [10] [21] [22].

3 Dealing with process variants

There are different ways to think solutions for managing variants in existing BPM tools according to the current bibliography, and they can be divided into two approaches: the multi-model and the single-model approach. In this section some concepts evaluated in a few related works are presented according to how they address all this terms in traditional BPM.

3.1 Multi-model approach

In existing BPM tools, process variants often have to be defined and kept in separate process models. Typically, this results in highly redundant model data as the variant models are identical or similar for most parts. Furthermore, the variants cannot be strongly related to each other; i.e., their models are only loosely coupled (e.g., based on naming conventions). As a conclusion, generally, modeling all process variants in separate models does not constitute an adequate solution for variant management [2] [3] [11].

3.2 Single model approach

Another approach, frequently applied in practice, is to capture multiple variants in one single model using conditional branchings (i.e., XOR-/OR-Splits). Each execution path in the model represents a particular variant. Therefore, branching conditions indicate which path belongs to which variant. Generally, specifying all variants in one process model could result in a large model, which is difficult to comprehend and expensive to maintain. Neither the use of separate models for capturing process variants nor their definitions in a model based on conditional branching constitute an adequate method. Both approaches do not treat variants as first class objects; i.e., the variant-specific parts of a process are maintained and hidden either in separate models

(multi-model approach) or in control flow logic modules (single-model approach) [1] [3] [12] [13].

3.3 Lifecycle

In terms of Process Variants, the standard process lifecycle consists of three phases (Fig. 1), namely the design and modeling of the process, the creation of a particular process variant, and the deployment of this variant in a runtime environment (according to the selected approach).

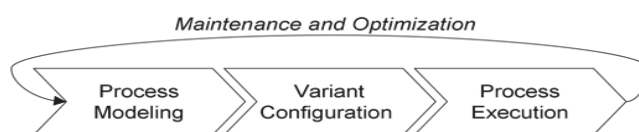


Fig 1: Process Variant lifecycle [2]

Modeling: the efforts for modeling process variants should be kept as minimal as possible, so in this direction, reusing variant models (or parts of them) has to be supported. In particular, it should be possible to create new variants by taking over properties from existing ones, but without creating redundant or inconsistent model data.

Variant configuration: the configuration of a process variant (what essentially means its derivation from a given master or base process) should be done automatically if possible.

Execution: to execute a process variant, its model has to be interpreted by a workflow engine, so in this context, it is important to keep information about the configured process variant and its relation to a master or base process (and to other variants) in the runtime system.

Maintenance and optimization: in order to reduce maintenance efforts and cost of change, fundamental updates affecting multiple process variants should be conducted only once [2] [3] [14] [15].

4. Applying verification in Process Variability. The challenge in Requirements Specification

In terms of process verification, as we have seen previously in Section 2, there are three main aspects to be considered as well: soundness, compliance and variability. This last one builds upon the concept of compliance. In the context of BPM, variability indicates that parts of a business process remain variable, or not fully defined, in order to support different versions of the same process depending on the

intended use or execution context. Imperative variability employs the use of variation points to provide different options at certain locations inside the process. Declarative variability uses specifications like those of compliance to specify how each version of a process should behave and absorb the new requirements [1] [21] [22] [29].

4.1 Imperative Variability

Imperative structural adaptation consists of atomic operations which, when executed in a specific predefined sequence, rearrange a business process to form a specific variant. Table 1 shows some mechanisms commonly known as Atomic Structural Adaptations.

Table 1: Imperative variability

Object	Action
Process Fragment	Insert/Delete/Move/Replace/Swap/Copy/Embed in loop/Parallelize/Embed in traditional branching
Subprocess	Extract/Inline
Control dependency	Add/Remove
Other	Update condition

Continuing with this idea, it also should be able to express the following self-explanatory atomic resource adaptations to the business process. These features are commonly named as Atomic Resource Adaptation. A variability management framework should at the same time allow the process designer to express the above imperative structural adaptation requirements, which are commonly named as Variation Relations [1] [3] [16] [17].

4.2 Declarative Variability

Declarative specifications consist of a set of rules expressing variations by acknowledging the borders which limit the possible process modifications. They are useful in order to set boundaries at the time of modifying a process model. Unlike atomic structural changes which indicate imperatively what can vary, a declarative specification limits the borders of changes explicitly [2] [18] [19] [20].

5 Applying Process Variability policies to Cloud BPM

As it was presented previously, there are different approaches and mechanisms to handle process variants in BPM. In our case, it is important to perceive that managing variability in a cloud environment amplifies every issue presented for single-tenant scenarios. In case of applying this approach to process monitoring, during the gathering phase, it is important for each instance to feed the monitoring server (i.e. using complex events in CEP or BAM), no matter the version of the process they belong to [1] [2] [22] [24].

In practice, process variants are often created by cloning and adjusting an existing process model of a particular type according to the given context. Generally, every process model can be derived out from another one by adjusting it accordingly, i.e., by applying a set of change operations and change patterns, respectively, to it. There are different implementations of process variability, and each framework applies policies according to its own particularities. One of these frameworks present in current bibliography is Provop ([2], implemented for single tenant BPM servers). Starting from this observation, Provop provides an operational approach for managing process variants based on a single process model. In particular, process variants can be configured by applying a set of high-level change operations to a given process model. The latter one is denoted as the base process [25].

5.1 Modeling

In the modeling phase, first of all, a base process, from which the different process variants can be derived through configuration, has to be defined. Following this, some high-level change operations, which can be applied to this base process, are specified. There is a thing that results fundamental for configuring a process variant: the base process. This serves as reference for the high-level change operations. Basically, the approach (e.g. Provop) should support policies that consider the standard process, the most frequently used process, the minimal average distance between a model and its variants, the superset of all process variants and the intersection of all process variants. This framework should also consider change operations, grouping change operations into options, constraint-based use of options and the context model [1] [3].

5.1.1 Policies application in a Cloud environment

As it was previously presented in [29], in a cloud based decomposed solution there are several process partitions that together conform the original process model. In this scenario, it is necessary to coordinate different models that in combination compose the original base process. In case of adding to this some process variability features, it could be necessary to apply a particular policy during defining the base process (or even more than one) in a particular portion of it, and the other ones could be not affected by the change [2] [21] [22] [23] [29].

5.2 Variant Configuration

In the configuration phase, the base process, the options defined for it and the context model are used to configure the models of the different variants. More precisely, a particular variant is configured by applying a sequence of options and their corresponding change operations to the base process. The sequence of steps is given as (1) select the relevant options, (2) evaluate relations between selected options, (3) determine the order in which options shall be applied, (4) apply options and their change operations and checking consistency [1] [2] [26].

5.2.1 Configuration in a cloud based environment

Once again following the same line as in [29], in a cloud based model the complexity of each step application gets amplified because of the existence of several decomposed models forming the original base model. In summary, option constraints are considered to ensure semantic correctness and consistency of the selected set of options at configuration time. This follows from the above described policies for defining the base process, assuming, for example, a base decomposed process being defined as an intersection of its variants [2] [22] [23] [29].

5.3 Deployment and Execution

After the configuration phase, the resulting variant model needs to be translated into an executable workflow model. Common tasks emerging in this context are to assign graphical user interfaces, to subdivide workflow activities into human and automated tasks, or to choose the right level of granularity for the workflow model [2] [26].

5.3.1 Deployment in a cloud based environment

According to the scenario previously presented in [22] and [29], and using the same architecture that integrates several nodes with Bonita in a private cloud, to apply all these concepts in the execution and deployment of decomposed process could result a very intricate task. As it was seen in previous works, each part of the process is in charge of invoking the next one with the goal of maintaining the original process flow, and for this invocation some vital information is needed: server direction, version of the process model, user and password to connect with the API [2] [29].

5.4 Maintenance and Optimization

When evolving base processes (e.g., due to organizational optimization efforts or changes in the business rules), all related process variants (i.e., their models) must be reconfigured automatically. Thus, maintenance efforts can be significantly reduced [27] [28].

5.4.1 Maintenance in a cloud based environment

Evolving and optimizing the base process may affect existing options, for example, when the referred adjustment points are moved to a new position or are even deleted. These actions cause that processes lose reference points and then certain actions are more complicated, for example checking whether the definitions of existing options are affected by the adaptations of the base process model.

As we have seen previously in [29], there are methods like BAM (*Business Activity Monitoring*) or even CEP (*Complex Event Processing*) used in a distributed environment in order to obtain relevant information about the process in execution. Table 2 shows how the different policies and actions could be added to a cloud decomposed model in order to handle process variant management during the whole lifecycle [1] [2] [29].

Lifecycle Phase	Policy/Action
Modeling	Definition: standard process - most frequently used process - minimal average distance - superset of all process variants - intersection of all process variants. change operations, grouping change operations, constraint-based options and context model
Variant configuration	Select relevant options - Evaluate relations between selected options - Determine the order in which options shall be applied - Applying options and their change operations - Checking consistency
Deployment and Execution	Assign graphical user interfaces - Subdivide workflow activities into human and automated tasks - Choose the right level of granularity for the workflow model
Maintenance and optimization	Automated reconfiguration

Table 2: Policies and actions to apply in cloud BPM

6. Conclusions

BPM since the beginning was conceived as a methodology capable of reducing the gap between the market and the final implementation of the business processes that interact with it. The conditions affecting a process may change any time, so the rules within the organization, causing in fact that every process definition should be adapted according to some new specifications.

When a BPMS is already inserted in the organization and each process has instances in execution, to generate a process variant is not a simple task: the process analyst should decide how to apply the changes in the base model, how to promulgate them in the existing instances, if every instance is going to be affected by the changes or not, and finally being capable of manage the different versions of the process in parallel. If this does not seem simple even in a single tenant environment, neither it is in a cloud

decomposed one where each process runs separately, in different servers united by an execution chain provided by an API in the BPMS. There are several frameworks in the current bibliography for process variant handling (like Provop as it was named repeatedly in previous sections), tending to implement different policies in order to apply changes in process instances and automate different tasks that enhance process variant management and maintenance. A possible line for future works is to implement a concrete version of a framework (such as Provop) that implements the automation of process variants promulgation in a cloud environment considering decomposed processes that need to be chained during execution and monitoring.

References

1. M Reichert, A Hallerbach, T Bauer. "Lifecycle Management for Business Process Variants". University of Ulm, Daimler TSS GmbH, Neu-Ulm University of Applied Science, Neu-Ulm, Germany. March 2015
2. G, Heerko. "Business Process Variability: A Study into Process Management and Verification". Rijksuniversiteit Groningen. 2016
3. B Estrada-Torres, A del Río-Ortega, M Resinas and A Ruiz-Cortes. "Identifying Variability in Process Performance Indicators". Departamento de Lenguajes y Sistemas Informáticos, Universidad de Sevilla, Seville, Spain. March 2017
4. SM Reza Beheshti, B Sherif Sakr, D Grigori, H Nezhad, M Ahmed Gater, S Hwan Ryu. "Process Analytics: Concepts and Techniques for Querying and Analyzing Process Data". Springer International Publishing. Switzerland 2016.
5. G Castro Barbosa Costa, C M.L. Werner and R Braga. "Software Process Performance Improvement Using Data Provenance and Ontology". Systems Engineering and Computer Science Department, Federal University of Rio de Janeiro COPPE, Rio de Janeiro, RJ, Brazil. August 2016.
6. V Ferme, A Ivanchikj and C Pautasso. "Estimating the Cost for Executing Business Processes in the Cloud". Faculty of Informatics, USI Lugano, Lugano, Switzerland. August 2016
7. M Hewelt and M Weske. "A Hybrid Approach for Flexible Case Modeling and Execution". Hasso Plattner Institute Potsdam, Potsdam, Germany. August 2016
8. D Sanchez-Charles, V Munes-Mulero, J Carmona, and M Sole. "Process Model Comparison Based on Cophenetic Distance". CA Strategic Research Labs, CA Technologies, Barcelona, Spain. August 2016
9. T Lehto, M Hinkka and J Hollmen. "Focusing Business Improvements Using Process Mining Based Influence Analysis". QPR Software Plc, Helsinki, Finland. August 2016
10. A Mos and M Cortes-Cornax. "Business Matter Experts do Matter: A Model-Driven Approach for Domain Specific Process Design and Monitoring". Xerox Research Center, 6 Chemin de Maupertuis, Meylan, France. August 2016
11. B Karim, Q Tan, I El Emary, B A. Alyoubi, R Soler Costa. "A proposed novel enterprise cloud development application model". Springer-Verlag Berlin Heidelberg 2016
12. O Skarlat, M Borkowski and S Schulte. "Towards a Methodology and Instrumentation Toolset for Cloud Manufacturing". Distributed Systems Group, TU Wien. April 2016

13. E Hachicha, N Assy, W Gaaloul and J Mendling. "A Configurable Resource Allocation for Multi-Tenant Process Development in the Cloud". Telecom SudParis, UMR 5157 Samovar, Universite Paris-Saclay, France. December 2015
14. G Rosinosky, S Youcef, F Charoy. "An Efficient Approach for Multi-tenant Elastic Business Processes Management in Cloud Computing environment". 2016 hal-01300188
15. A Gunka, H Kuehn and S Seycek. "BPM in the Cloud: The BOC Case". BOC Information Technologies Consulting GmbH, Vienna, Austria. 2017
16. G Rosinosky, S Youcef, F Charoy. "A Framework for BPMS Performance and Cost Evaluation on the Cloud". Workshop "Business Process Monitoring and Performance Analysis in the Cloud", Dec 2016, Luxembourg, Luxembourg.
17. N. Herzberg, A. Meyer, M. Weske. "An Event Processing Platform for Business Process Management". Business Process Technology Group, Hasso Plattner Institute at the University of Potsdam. Potsdam, Germany. June 2013.
18. S. Bulow, M. Backmann, N. Herzberg, T. Hille, A. Meyer, B. Ulm, T. Y. Wong, M. Weske. "Monitoring of Business Processes with Complex Event Processing". Business Process Technology Group, Hasso Plattner Institute at the University of Potsdam. Potsdam, Germany. July 2013
19. C. Zeginis, K. Kritikos, P. Garefalakis, K. Konsolaki, K. Magoutis and D Plexousakis. "Towards Cross-Layer Monitoring of Multi-Cloud Service-Based Applications". Institute of Computer Science Foundation for Research & Technology – Hellas. Grece. August 2013.
20. M. Goetz. "Integration of Business Process Management and Complex Event Processing". iTransparent GmbH, IT Consulting, Bergstraße 5, 90403 Nuremberg, Germany. November 2010.
21. J Martínez Garro, P Bazán. "Monitoreo de procesos en el cloud. Una propuesta arquitectónica". JCC 2013. Universidad de Temuco. Chile. November 2013.
22. J Martínez Garro, P Bazán. "Constructing and monitoring processes in BPM using hybrid architectures". IJACSA Journal. Londres. Febrero 2014.
23. J Martínez Garro, P Bazán, J Diaz. "Decomposed processes in Cloud BPM: techniques for monitoring and the use of OLC". WORLD COMP 2014. Las Vegas, USA, July 2014.
24. J Martínez Garro, P Bazán, J Diaz. "OLC y Monitoreo de procesos en el cloud: un caso de estudio". JCC 2014. Chile. November 2014.
25. R Confortia, M La Rosaa, G Fortinoc, A H. M. ter Hofstede, J Reckera, M Adamsa. "Real-Time Risk Monitoring in Business Processes: A Sensor-based Approach". Queensland University of Technology, Brisbane, Australia. May 2013.
26. E Mulo, U Zdun, S Dustdar. "Domain-Specific Language for Eventbased Compliance Monitoring in Process-driven SOAs". Distributed Systems Group Institute of Information Systems Vienna University of Technology, Vienna, Austria. April 2013.
27. P Szwed, W Chmiel, S Jedruzik, P Kadluczka, "Business Process in a Distributed Surveillance System integrated through workflow". Automatika. Vol 17. No 1. November 2013.
28. V Stavrou, M Kandias, G Karoulas, D Gritzalis. "Business Process Modeling for Insider Threat Monitoring and Handling". Information Security & Critical Infrastructure Protection Laboratory Dept. of Informatics, Athens University of Economics & Business. Greece. May 2014.
29. J Martínez Garro, P Bazan, J Diaz. "Using BAM and CEP for Process Monitoring in Cloud BPM". JCST Cloud Journal. April 2016

Propuesta de Protocolo de Formación de Pares Experimentales de Programadores

Mauricio Dávila¹, Marisa Panizzi¹, Darío Rodríguez^{2,3}

¹ Programa de Maestría en Ingeniería de Sistemas de Información. Universidad Tecnológica Nacional. Facultad Regional Buenos Aires, Castro Barros 91, (C1178AAA), C.A.B.A., Argentina.

² Grupo de Ingeniería de Espacios Virtuales de Trabajo y Grupo de Investigación en Sistemas de Información. Departamento de Desarrollo Productivo Tecnológico. Universidad Nacional de Lanús, 29 de Septiembre 3901, (B1826GLC), Lanús, Buenos Aires, Argentina.

³ Comisión de Investigación Científicas - CIC. Calle 526 e/10 y 11, (B1906APP), La Plata, Buenos Aires, Argentina.
davilamr.80@gmail.com, marisapanizzi@outlook.com, dariorodriguez1977@gmail.com

Resumen. En este trabajo, se propone un protocolo que permite formar pares experimentales homogéneos de programadores con el propósito de asegurar que dos sujetos de iguales características son indistinguibles en lo que refiere a sus aptitudes como programadores. Este protocolo permite evaluar las características de los lenguajes de programación sin que estas mediciones se vean afectadas por las habilidades o ausencias de las mismas por parte de los programadores. Se presenta un caso testigo que permita validar el protocolo, a tal efecto se aplicará dicho protocolo a un grupo de programadores del lenguaje C con el fin de demostrar que los pares experimentales formados no presentan diferencias significativas, en lo que refiere a calidad y tiempo, en la codificación de una especificación.

Palabras Clave: experimentación en ingeniería de software, protocolo, pares experimentales de programadores, programación

1 Introducción

En la actualidad existe una gran diversidad de lenguajes de programación y a menudo esto dificulta la tarea de seleccionar el lenguaje que mejor se adapta a las necesidades del desarrollo. La decisión del lenguaje de programación para realizar una solución implica múltiples factores de análisis, muchos de los cuales están sujetos al código fuente producido y al tiempo empleado para poder producirlo.

Utilizando el método de revisiones sistemáticas [1] se ha realizado una investigación documental de estudios en los cuales se hace referencia a las métricas que se pueden establecer sobre el código producido con un lenguaje de programación determinado [2][3][4] y [5]. Cuando se utilizan métricas para determinar si un lenguaje es una mejor alternativa frente a otro, no se debe perder de vista que dichas métricas solo se enfocan en el código resultante y no tienen en consideración las características

del programador que construyó el código. Esto puede ocasionar la sub calificación o sobre calificación de un lenguaje por habilidades o ausencias de las mismas por parte de los programadores que los utilizan. Esto último, plantea un problema a la hora de diseñar un experimento cuyo propósito sea determinar el lenguaje a utilizar, ya que el grupo de programadores que utilizan el lenguaje A puede poseer un nivel de conocimientos superior que el grupo de programadores que emplean el lenguaje B o viceversa y esto impactaría en el resultado del experimento. Una posible solución a esta problemática sería contar con un conjunto de programadores que sean capaces de resolver la misma tarea en los lenguajes que se requirieron evaluar.

Más allá de la dificultad que sería reunir a este conjunto de sujetos con los conocimientos necesarios en cada uno de los lenguajes que se pretenden evaluar, esta solución presenta algunas dificultades adicionales que surgen a la hora de diseñar experimentos.

Tanto en el trabajo de Juristo y Moreno [6] como en el de Wohlin [7] se considera que los experimentos realizados en el campo de la ingeniería de software son fuertemente influenciados por las características de los sujetos al igual que ocurre en otras ciencias, generalmente conocidas como ciencias sociales. Juristo y Moreno [6] enumeran algunos puntos relacionados con los factores sociales y las características específicas de desarrollo de software que deben tenerse en cuenta al diseñar experimentos.

- Efecto aprendizaje: Si un sujeto tiene que resolver el mismo problema aplicando diferentes lenguajes de programación es muy probable que este aprenda cada vez más sobre el problema y que el último resultado sea mejor que el primero, simplemente porque el sujeto sabe más sobre el problema y no porque el lenguaje de programación sea mejor.
- Efecto de aburrimiento: los sujetos se aburren o se cansan del experimento y ponen menos esfuerzo e interés a medida que pasa el tiempo.
- Efecto de entusiasmo: Puede suceder que los sujetos que utilizan un lenguaje de programación antiguo no están motivados para hacer un buen trabajo, mientras que los que utilizan un nuevo lenguaje de programación sí lo están.
- Efecto de la experiencia: A la hora de realizar un experimento que involucra programadores es de esperar que existan distintos niveles tanto de conocimiento como de habilidad sobre el lenguaje de programación empleado.
- La formalización inconsciente: Surge cuando un mismo sujeto utiliza dos o más lenguajes de programación con diferentes grados de definición o formalidad.
- Efecto de ajuste: El estado emocional de los sujetos participantes se encuentra íntimamente relacionado con el rendimiento que estos tendrán.

A continuación se describen un conjunto de acciones a considerar para controlar los efectos antes enumerados:

- Efecto Aprendizaje: No utilizar al mismo conjunto de sujetos para desarrollar en más de un lenguaje de programación.
- Efecto de aburrimiento: Motivar a los sujetos que ejecutan el experimento de igual manera independientemente del grupo que integren.

- Efecto de entusiasmo: No poner al tanto a los sujetos sobre las hipótesis formuladas o los objetivos del experimento.
- Efecto de la experiencia: Para controlar este efecto se formarán pares experimentales de programadores que sean indistinguibles en lo que refiere a sus conocimientos y habilidades con el lenguaje de programación.
- La formalización inconsciente: Deberán ser considerados aspectos referidos al nivel de conocimiento que cada sujeto tiene en lo que refiere al lenguaje de programación empleado en el momento de formar el par experimental
- Efecto de ajuste: Se debe tener en cuenta que todas las instancias del experimento se realicen bajo las mismas condiciones.

Dentro del área de ingeniería de software, es habitual enfrentarse a la necesidad de realizar experimentos con un conjunto reducido de personas las cuales aplican diferentes tratamientos a los objetos de estudio. Teniendo en cuenta que realizar comparaciones dentro de pares homogéneos de unidades experimentales no solo aumenta la precisión del análisis sino que también permite controlar gran parte de los efectos no deseados [6] a la hora de realizar un experimento. Este trabajo se desarrolla bajo la hipótesis de que es posible formar pares experimentales homogéneos de programadores y que por lo tanto dos sujetos de iguales características son indistinguibles en lo que refiere a sus aptitudes como programadores. De esta hipótesis se deriva las siguientes preguntas de investigación: ¿es posible formar pares experimentales indistinguibles en lo que refiere a sus aptitudes como programadores? De ser posible formarlos, ¿requieren el mismo tiempo para resolver con el mismo lenguaje la misma tarea?

En la (Sección 2) se propone el protocolo de formación de pares experimentales de programadores, en la (Sección 3) se lleva adelante la validación del protocolo mediante la realización de una prueba piloto y en la (Sección 4) se plantean las conclusiones y las futuras líneas de investigación.

2 Propuesta del protocolo de formación de pares experimentales.

A la hora de definir cómo formar pares homogéneos de programadores, los autores adherimos a la propuesta de Campbell [8], que sostiene que muchos son los factores afectan el rendimiento de un individuo de manera indirecta, pero sólo hay tres determinantes directos del rendimiento: *conocimiento, habilidad y motivación*, por lo cual, se elaborará un protocolo que tenga por propósito la formación de pares experimentales de programadores que resulten homogéneos en lo que refiere tanto a nivel de conocimiento como a sus habilidades y dará por sentado que todos los participantes a caracterizar no presentan diferencias significativas en lo que respecta a la motivación.

El primero de los pasos consiste en identificar las formas utilizadas para categorizar a los programadores en otros estudios de investigación experimental (Sección 2.1), luego se describen los lineamientos considerados para la construcción de los instrumentos de caracterización utilizados en el experimento (Sección 2.2) y por último

en la (Sección 2.3) se establece un mecanismo que permite formar pares experimentales homogéneos de programadores en base a la caracterización realizada.

2.1 Experiencia de los programadores

Al igual que en la mayoría de las actividades humanas, el rendimiento individual en el desarrollo de software varía considerablemente de una persona a otra y se deben articular mecanismos para que dichas variaciones no afecten los resultados del estudio. Feigenspan y otros [9] realizaron un trabajo documental en el cual se analizaron 161 publicaciones y se identificaron nueve formas utilizadas por los investigadores para determinar la experiencia que un programador posee. Estos autores definen la experiencia como la cantidad de conocimientos adquiridos respecto al desarrollo de programas.

- Años: En cuarenta y siete de los trabajos, los años que un participante ha estado programando en general o en una empresa o en cierto lenguaje se usó para determinar la experiencia de programación.
- Educación: La educación de los participantes se utilizó para indicar su experiencia en diecinueve de los artículos revisados, en los cuales se incluyó información sobre el nivel de educación alcanzado (pregrado, grado, posgrado, etc.) o las calificaciones de los cursos.
- Auto Estimación: En doce trabajos, se pidió a los participantes que estimaran su propia experiencia.
- Cuestionario específico: En nueve trabajos los autores aplicaron un cuestionario para evaluar la experiencia de programación.
- Tamaño: El tamaño de los programas que los participantes habían escrito fué utilizado como indicador en seis artículos.
- Examen: En tres trabajos se realizó un examen de programación para evaluar la experiencia de los participantes.
- Supervisor: En dos trabajos, en los que los programadores profesionales fueron utilizados como participantes, un supervisor fue el encargado de estimar su experiencia.
- No especificado: A menudo, los autores afirman que midieron la experiencia de programación, pero no especificaron cómo. Este fue el caso en treinta y nueve trabajos.
- No controlada: La experiencia de programación no se mencionó en absoluto en cuarenta y cinco trabajos, lo que amenaza la validez de los experimentos correspondientes.

2.2 Caracterización de los programadores

Se requiere desarrollar un método de caracterización que no se base en la percepción que tiene cada individuo acerca de sus habilidades como programador ya que las personas menos competentes tienden a sobrestimar su habilidad al no tener conocimiento suficiente para reconocer sus propias limitaciones y que también es común que personas más preparadas tiendan a subestimar sus logros y sus competencias

[12]. Esta caracterización persigue como único objetivo calificar un conjunto de aptitudes del programador a solo efecto de poder encontrar parejas de programadores que se puedan considerar homogéneas. El propósito que tiene la caracterización es asegurar que dos sujetos de iguales características son indistinguibles en lo que refiere a sus aptitudes como programadores, para lo cual, se optó por analizar un amplio conjunto de aptitudes del programador, los autores adhieren a la idea que la realización de una caracterización basada tan solo en unos pocos criterios puede incurrir en graves errores. Algunas características de los programadores son independientes del lenguaje de programación y otras son dependientes de él, por lo cual se fijarán los lineamientos que permitan desarrollar dos instrumentos de caracterización, uno independiente del lenguaje de programación (CILP) y el otro que tenga en cuenta un lenguaje de programación (CDLP).

Existe un gran número de medidas para capturar atributos de los procesos y productos software, que tradicionalmente se han realizado confiando en la sabiduría de los expertos y esta situación ha conducido frecuentemente a cierto grado de imprecisión en las definiciones, propiedades y suposiciones de las medidas, haciendo que el uso de las medidas sea difícil, la interpretación peligrosa y los resultados de muchos estudios de validación contradictorios [10]. Para la construcción de las herramientas de caracterización se ha tenido en cuenta el procedimiento general de construcción de un instrumento de medición propuesto por Sampiere [11] y el método para la definición de medidas válidas propuesto por Genero, Cruz-Lemus y Piattini [10], adaptando dichos procedimientos a la necesidad de este trabajo.

Por cuestiones relacionadas a la síntesis que esta publicación demanda, no es posible detallar cada uno de los pasos que han permitido definir los instrumentos ni el contenido de cada dimensión de los mismos.

En la Tabla 1, se presentan los dominios del contenido de la variable (dimensiones), los indicadores de cada dimensión, y la nomenclatura propuesta de las dimensiones para el instrumento que se empleará para la caracterización independiente del lenguaje de programación.

En la Tabla 2, se presentan los dominios del contenido de la variable (dimensiones), los indicadores de cada dimensión, y la nomenclatura propuesta de las dimensiones para el instrumento que se empleará para la caracterización dependiente del lenguaje de programación.

En lo que respecta a la decisión del tipo y formato del instrumento y su contexto de administración se utilizará el procedimiento de recolección de datos mixto conformado por dos cuestionarios (CILP y CDLP) y luego por una entrevista.

Con el propósito de minimizar los errores de caracterización, una vez que el participante ha contestado el cuestionario se desarrollara una entrevista individual donde el entrevistador realiza preguntas a efectos de que el participante justifique sus respuestas y ante la correcta justificación del participante el entrevistador dará la misma como válida.

El contexto de administración será una sala provista de una computadora por programador ya que la primer instancia, la que refiere al cuestionario, es auto administrada permitiendo llevar esta acción de forma individual o grupal, para luego si pasar a la instancia de entrevista de forma individual.

Tabla 1. Variable, dimensiones, nomenclatura de cada dimensión y sus indicadores.

Variable a medir	Dimensión	Nomenclatura	Indicador
Características independientes del lenguaje de programación	<i>Nivel de formación.</i>	<i>CILP1</i>	Refiere al nivel alcanzado por el participante en su educación ya sea formal o informal.
	<i>Experiencia</i>	<i>CILP2</i>	Refiere a los años de experiencia como programador y a la cantidad de lenguajes que manifiesta conocer.
	<i>Comprensión de una Especificación</i>	<i>CILP3</i>	Capacidad que tiene el programador para comprender una especificación simple y el tiempo insumido.
	<i>Comprensión de Pseudocódigo</i>	<i>CILP4</i>	Capacidad que tiene el programador de interpretar el funcionamiento de bloques de pseudocódigo y el tiempo insumido.
	<i>Capacidad Algorítmica</i>	<i>CILP5</i>	Capacidad que tiene el programador en desarrollar una solución con pseudocódigo y el tiempo que emplea en realizar dicha solución.

Tabla 2. Variable, dimensiones, nomenclatura de cada dimensión y sus indicadores.

Variable a medir	Dimensión	Nomenclatura de la Dimensión	Indicador
Características dependientes del lenguaje de programación	<i>Lenguaje Elegido</i>	<i>CDLP1</i>	Refiere al lenguaje de programación que el participante expresa dominar con mayor fluidez y a los años de experiencia que tiene con este.
	<i>Conocimiento Teórico</i>	<i>CDLP2</i>	Refiere al nivel de conocimientos sobre aspectos teóricos del lenguaje de programación que el participante posee.
	<i>Comprensión de Código Fuente</i>	<i>CDLP3</i>	Capacidad que tiene el programador de interpretar el funcionamiento de bloques de código y el tiempo insumido en ello.

2.3 Mecanismo de formación de pares experimentales

Es pertinente contar con un mecanismo que asegure que las parejas experimentales se encuentran integradas por sujetos homogéneos, esto implica que según su caracterización son indistinguibles o presentan diferencias mínimas.

2.3.1 Criterio de Utilización de las Variables

Del procedimiento de caracterización surgen múltiples variables, algunas relacionadas a características independientes del lenguaje de programación y otras a características que dependen del desempeño del programador en un lenguaje de programación en particular.

2.3.2 Normalización

El proceso de normalización implica transformar los valores de las variables independientes a fin de que estas queden representadas en el rango [0-10] independientemente de cual fuese su escala original. Este paso, permitirá asegurar que ninguna de las variables intervinientes en el cálculo de distancia se encuentran ponderadas por sobre otras.

2.3.3 Penalidades por tiempo

Cada una de las variables medida es acompañada por el tiempo insumido por el participante en completar los ejercicios relacionados a la misma. A efectos de formar pares experimentales que no solo sean indistinguibles en lo que refiere a conocimiento sino también al tiempo insumido para dar dicha respuesta, se aplicará una penalidad en el puntaje obtenido según el tiempo insumido, la puntuación obtenida se irá decrementando en un 10% por cada cinco minutos insumidos.

2.3.4 Cálculo de Distancia

En un escenario donde se requiere evaluar múltiples variables, todas ellas cuantitativas y cuyos valores pertenecen luego del proceso de normalización al intervalo [0,10] se debe definir el criterio a ser empleado para determinar la distancia existente entre dos sujetos, por tratarse de un espacio n-dimensional se aplicará el cálculo de distancia euclídea [15] (Figura 1).

$$d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_i - q_i)^2 + \dots + (p_n - q_n)^2}.$$

Figura 1. Distancia euclídea [15]

Se realizará el cálculo de distancia entre participantes donde el valor del módulo de la diferencia entre variables del mismo tipo no supere un umbral, esta restricción permitirá establecer una distancia máxima tolerada la cual no podrá ser producto de una diferencia significativa en una sola variable.

2.3.4 Algoritmo de Selección de Parejas Experimentales

El algoritmo sigue un proceso secuencial en el que a cada paso toma una decisión, seleccionar la mínima distancia de entre las disponibles y luego en el siguiente paso el algoritmo se encuentra con un problema idéntico, pero estrictamente menor, al que tenía en el paso anterior y vuelve a aplicar la misma función de selección para tomar la siguiente decisión [14].

3 Caso de validación del protocolo.

Se realizaron una serie de acciones para demostrar el nivel de confiabilidad y validez inicial de los instrumentos de medición. Las características que intervienen en la formación de los pares experimentales son: Comprensión de una Especificación (CILP3), Comprensión de Pseudocódigo (CILP4), Capacidad Algorítmica (CILP5), Conocimiento Teórico (CDLP2) y Comprensión de Código Fuente (CDLP3).

Para la realización de la prueba piloto inicial, se convocó a participar de la investigación a personas mayores de edad que han manifestado conocer el lenguaje de programación C, logrando conformar un grupo de 14 programadores. Luego de ser caracterizados los participantes, se les solicitó resolver una tarea de mediana complejidad, por último se aplicó el protocolo de formación de pares experimentales de programadores a efecto de poder determinar si cada pareja experimental presenta o no diferencias significativas en la resolución de la tarea. Si los pares experimentales conformados aplicando el protocolo no presentan diferencias significativas al resolver con el mismo lenguaje la misma tarea, se podrá considerar que tanto los instrumentos como el protocolo de formación de pares experimentales se encuentran en una versión estable.

Tabla 3. Resultados normalizados de la caracterización.

Sujeto	CDLP2	CDLP3	CILP3	CILP4	CILP5
1	2.40	0.00	10.00	0.00	1.64
2	8.10	6.75	9.00	5.00	6.36
3	6.40	4.00	7.50	5.00	2.73
4	9.00	9.00	4.50	5.00	6.36
5	6.30	2.25	6.75	7.00	2.18
6	7.20	6.00	4.50	9.00	0.67
7	8.10	0.00	7.50	4.50	1.48
8	5.60	2.25	10.00	4.50	4.43
9	8.00	4.50	6.75	9.00	2.96
10	6.30	4.50	4.50	4.50	7.64
11	7.20	2.50	6.75	5.00	1.86
12	9.00	6.75	9.00	9.00	5.45
13	5.60	0.00	7.50	5.00	7.00
14	6.40	0.00	9.00	4.50	3.87

Los resultados que surgen del proceso de caracterización figuran en la (Tabla 3). En la (Tabla 4) se presenta la matriz de distancia correspondiente, las intersecciones pintadas de color negro representan pares de sujetos que no deben ser considerados ya que en por lo menos una de sus dimensiones han presentado una distancia superior al cincuenta por ciento. Los pares experimentales que surgen de aplicar el algoritmo se encuentran resaltados en color gris. Por último se presentan las diferencias de tiempo que existen entre los programadores que integran cada par experimental (Tabla 5).

Tabla 4. Matriz de distancia.

ID	2	4	5	7	9	11	12	14
2	-	5.11	7.07		6.14	6.65	4.20	
4	5.11	-			7.34		6.49	
5	7.07		-	3.95	3.54	2.23	6.88	4.38
7			3.95	-	6.58	2.83		9.20
9	6.14	7.34	3.54	6.58	-	4.67	4.17	7.00
11	6.65		2.23	2.83	4.67	-	7.44	4.03
12	4.20	6.49	6.88		4.17	7.44	-	
14			4.38	9.20	7.00	4.03		-

Tabla 5. Diferencias de tiempo empleado entre integrantes del par experimental

Sujetos		Diferencia	
		Tiempo	%
2	4	6	7.06%
5	11	3	2.48%
7	14	5	4.76%
9	12	5	4.35%

En lo referente al tiempo empleado para realizar la caracterización, el tiempo promedio utilizado para llevar adelante la caracterización independiente del lenguaje de programación fue de 28 minutos, para la caracterización dependiente del lenguaje de programación el tiempo promedio empleado fue de 19 minutos y para llevar a cabo las entrevistas en promedio se utilizaron 6 minutos por participante. Puede observarse en la (Tabla 5) que los pares experimentales de programadores no presentan diferencias significativas en el tiempo empleado para resolver la misma tarea.

4 Conclusiones y Futuros Trabajos.

Con el fin de proponer un protocolo de formación de pares experimentales de programadores se ha llevado a cabo un análisis documental en lo que refiere a los beneficios que esta forma de experimentación aporta en el área de ingeniería de software. Se ha logrado la construcción de dos instrumentos que permiten caracterizar a los programadores. Se ha definido el procedimiento que permite formar los pares experimentales a partir de los datos obtenidos con los instrumentos de caracterización.

Por último se ha verificado mediante un caso de validación si los pares experimentales obtenidos presentaban diferencias en lo que refiere al tiempo empleado en resolver una misma tarea de programación.

Se puede concluir que en lo que refiere a la formación de los pares experimentales de programadores, el protocolo ha funcionado satisfactoriamente y por lo tanto se considera que tiene un nivel de confiabilidad, validez y objetividad aceptable ya que demostró consistencia entre los resultados obtenidos.

Como trabajo futuro se identifica la necesidad de: (1) poder aplicar el protocolo en otros lenguajes de programación, (2) poder utilizar el protocolo para formar pares experimentales entre sujetos que utilizan distintos lenguajes de programación a efectos de poder medir si un lenguaje de programación en particular afecta o no a la productividad informática.

Referencias

1. Argimón, J. Métodos de Investigación Clínica y Epidemiológica. Elsevier España. 84-8174-709-2. (2004).
2. Halstead, M.. Elements of Software Science. Elsevier Science Inc., New York, NY, USA.(1977).
3. McCabe, T. J. A complexity measure. IEEE Transactions on software Engineering , 4, 308-320. 1976.
4. Riaz, M., Mendes, E., & Tempero, E. A systematic review of software maintainability prediction and metrics. In Proceedings of the 2009 3rd International Symposium on Empirical Software Engineering and Measurement (pp. 367-377). IEEE Computer Society. 2009.
5. Rilling, J., y Klemola, T. Identifying Comprehension Bottlenecks Using Program Slicing and Cognitive Complexity Metrics. 10th IEEE Working Conference on Reverse Engineering. 10, pp. 115-125. Oregon, USA: IEEE. 2003.
6. Juristo, N., & Moreno, A. M. Basics of software engineering experimentation. Springer Science & Business Media. 2013
7. Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., & Wesslén, A. Experimentation in software engineering. Springer Science & Business Media. 2012.
8. Campbell, J. P., McCloy, R. A., Oppler, S. H., & Sager, C. E. A theory of performance. In Schmitt, N., Bormann, W.C. et al. (Eds.), Personnel selection in organizations, 35-70, San Francisco, Jossey-Bass. 1993.
9. Feigenspan, J., Kästner, C., Liebig, J., Apel, S., & Hanenberg, S. Measuring programming experience. In Program Comprehension (ICPC), IEEE 20th International Conference on (pp. 73-82). IEEE. 2012.
10. Genero, M., Cruz-Lemus, J. A., Piattini, M.: Métodos de Investigación en Ingeniería del Software. RaMa 2014.
11. Sampieri, R. H., Collado, C. F., & Lucio, P. B. Metodología de la investigación. México: McGraw-Hill. 2010.
12. Kruger, J., & Dunning, D. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. Journal of personality and social psychology, 77(6), 1121. 1999.
13. Wilking, D., Schilli, D., & Kowalewski, S. Measuring the human factor with the rasch model. In Balancing agility and formalism in software engineering (pp. 157-168). Springer Berlin Heidelberg. 2008.
14. Soriano, M. A. Algoritmos Voraces. Facultat d'Informàtica, U.P.C. Consultado el 3 de Enero de 2017, disponible en <http://www.cs.upc.edu/~mabad/ADA/curs0708/GREEDY.pdf>. 2007.
15. Deza, Elena; Deza, Michel Marie. Encyclopedia of Distances. Springer. p. 94. 2009.

Strategy for Comparing and Adopting Strengths

Guido Tebes, Denis Peppino, Pablo Becker, M.F. Papa, M.B. Rivera, and Luis Olsina

GIDIS_Web, Facultad de Ingeniería, UNLPam, General Pico, LP, Argentina
guido_tebes@hotmail.com, denispeppino92@gmail.com,
[beckerp, pmfer, riveramb, olsinal]@ing.unlpam.edu.ar

Abstract. To perform evaluation and improvement activities, the definition of a goal is usually the starting point of a project. Typically, a goal contains the statement of a purpose. In the Software Engineering literature, different evaluation purposes such as to understand, monitor, predict, improve, control, compare and selection are mentioned. These purposes share distinctive and common aspects as we discuss in the present work. A categorization that helps a better understanding of them is also proposed. Additionally, in order to reach the purpose of a goal, a strategy should be selected. We have envisioned different strategies for different purposes. Basically, a strategy describes a course of action by means of the specification of activities and methods. In order to analyze the applicability of a strategy, we illustrate the strategy for the *compare and adopt* purpose for adopting strengths in a new app comparing four social network applications.

Keywords: Goal, Evaluation purpose, Compare and adopt, Strategy.

1 Introduction

In any mature organization, the continuous quality improvement of its resources, processes, products and services should be ensured. Sometimes, software organizations need to develop new applications or to improve the existent ones, so, it would be desirable to take into account, as a reference, others applications to adopt their strengths. For this end, it is not only necessary to have Measurement and Evaluation (ME) activities which allow to understand the actual or estimated situation of an entity through strengths and weakness analysis, but also is important to have activities which allow to do changes oriented to the improvement (MEC will stand for ME and Change).

An engineering way to organize the ME/MEC activities and to manage resources is by project formulation. Considering project definitions given in [6, 12], both agree on that it is necessary to determine a project plan. During planning and resource allocation, the goal and its purpose are considered as well as the suitable strategy is selected aimed at achieving the goal. The selected strategy should describe the course of action for the goal and the particular purpose, by means of the specification of activities and methods. As a consequence, software organizations that consider ME or MEC projects in a systematic way can ensure process repeatability and consistency of results.

Aimed at illustrating the goal, purpose and strategy concepts, Fig. 1 outlines some

relations between them, which will serve as the basis for understanding this work. An organizational goal can be both business and information need goal. The latter supports the former, and consequently has the *analyze* purpose. Also, an information need goal could require a ME information need goal, which involve ME activities. Furthermore, typical evaluation purposes for business goals can be to understand, monitor, control, improve, adopt, among others. However, we observe that in the state-of-the-art literature there is some lack of consensus in their descriptions and/or definitions.

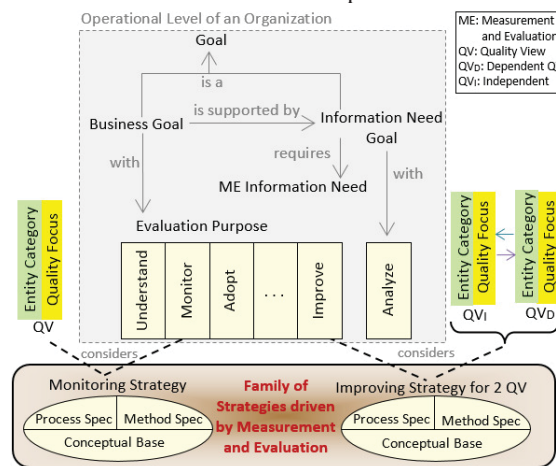


Fig. 1. Schema that relates Business Goal and Information Need Goal with Evaluation Purposes. Also, a Family of Strategies considers different purposes and Quality Views.

On the other side, Fig. 1 shows that a ME/MEC strategy considers an evaluation purpose and, one or more quality views. E.g., the *monitoring* strategy considers the *monitor* purpose and one quality view. According to [11], a strategy should have at least three capabilities such as a domain conceptual base, and the process and method specifications. These allow to have explicitly defined the ME/MEC domain, to know which generic and specific activities intervene, and how to carry out them through methods. As a strategy considers the quality view concept, it is worth saying that it represents an association between an entity category and a quality focus. Examples of quality views with entity category and quality focus names, in addition to the relationships between them are shown in Fig. 2 (see [14] for more details).

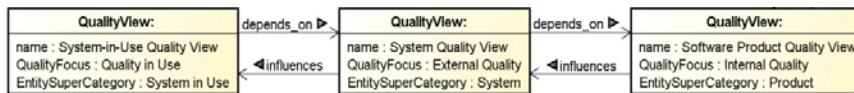


Fig. 2. Example of ‘depends on’ and ‘influences’ relationships between different Quality Views.

Therefore, in a specific ME or MEC project, the more suitable strategy should be selected from a family of strategies, taking into account the evaluation purpose of the goal and the amount and types of quality views.

In summary, the main contributions of this work are: i) to identify and categorize a

set of evaluation purposes for business goals and information need goals at operational level, which can be satisfied through the systematic use of strategies driven by ME, analysis and, eventually, change; and ii) to illustrate the applicability of the *comparing and adopting* strategy with the aim of determining Usability strengths from LinkedIn, Instagram, Twitter and Facebook social network apps to be adopted in a new entity.

The rest of the article is organized as follows. Section 2 defines the main terms for a better comprehension of this work. Section 3 analyses evaluation purposes in the context of related literature and proposes three categories for the considered purposes. Section 4 specifies a strategy for the *compare and adopt* purpose, illustrating it with a concrete evaluation scenario, which then is developed in Section 5. Section 6 discusses related work and, finally, Section 7 summarizes conclusions and future works.

2 Definition of some Terms

The terminology used in this work is grounded on conceptual bases structured in ontologies [2, 14]. Terms presented in Fig. 1 are next defined:

- *Goal*: the statement of the aim to be achieved by an organization which considers the propositional content of a *purpose* in a given time frame.
- *Business Goal*: it is the main *goal* that the organization intends to achieve.
- *Information Need Goal*: it is a *goal* intended to get insight for a given *business goal*.
- *ME Information Need Goal*: it is an *information need goal* driven by ME activities.
- *Purpose*: the rationale for achieving a specific *goal*.
- *Strategy*: principles, patterns, and particular domain concepts and framework that can be specified by a set of core processes, in addition to a set of appropriated methods and tools, as core resources, for helping to achieve the project's *goal purpose*.

3 Evaluation Purposes and Categories

As introduced in Section 1, in order to select a strategy from a family of strategies, three key factors should be taken into account, i.e., the business goal purpose, as well as the amount and type of involved views. In this section we address evaluation purposes.

Analyzing the ME literature, it is important to remark that there is still no consensus about definitions on different purposes, particularly, for those purposes which involve evaluation and change activities. Many works consider that *characterize* and *understand* purposes are the same one, which involve to understand or to take a snapshot of the current situation of an entity for establishing baselines for future assessment [3, 7, 8, 10]. However, the *characterize* purpose could be previous to *understand*, since it would allow the characterization of an entity from the relevant nonfunctional requirements identification standpoint. Once identified the relevant characteristics, the understanding of the current entity state can be done by means of ME activities. On the other hand, the *predict* purpose usually is intended to identify and understand cause/effect relationships between characteristics/attributes of entities [7] with the aim of establishing predictive models that could be useful in MEC activities.

Besides, the *monitor* purpose implies tracking the status or performance of attributes

of an entity [3]. That is, it consists of a continuous understanding of an entity over time for analyzing its evolution and possible trends. Note that monitoring involves the *understand* purpose applied in a repetitive way, with some frequency in a defined interval of time. Sometimes, the *control* purpose is related with *monitor* since control addresses the identification of deviations that influence the status or performance of processes and products for reducing risks [3]. Some authors [3, 8] consider the *control* and *monitor* purposes as a single purpose, due to they involve a continual evaluation and occasionally changes for improving the expected performance.

On the other hand, Precece and Rombach [14] categorized evaluation purposes as passive or active: “*passive purposes are aimed at better understanding or visualizing existing software items without influencing them, whereas active ones are aimed at actually influencing them in some way*”. We can infer that purposes for the active category always imply changes on the target entity and/or its context, aimed at improving it. Conversely, evaluation purposes for the passive category imply no changes.

We observe that both categories, however, do not fit well the intention of all evaluation purposes. For example, the *compare* and *adopt* purpose, which involves the determination of strengths and weakness of a set of preselected representative entities, and the adoption of strengths in a new entity or in one that already exist, could be both active and passive. If it is about adopting detected strengths in an existent entity, this purpose implies changes and falls into the active category; otherwise, if the strengths are adopted in a new entity to be developed, it falls into the passive category. Furthermore, exists the *select an alternative* purpose among preselected competitive entities aimed at adopting and installing it. As a result, for purposes that have entity preselection -independently if purposes are active or passive- a new category can be envisioned. Consequently, we represent three categories for evaluation purposes, namely: *Characterize and Understand*, *Control and Change*, and *Adopt an Alternative*.

The ***Characterize and Understand*** category includes purposes such as to *understand*, *monitor* and *predict*. *Understand* implies getting information about the current state/performance of an entity, taking into account that changes are not performed on this entity. Regarding the *monitor* purpose, as above mentioned, it consists of a continuous understanding over time. The *predict* purpose helps to understand possible cause-effect relationships among characteristics/attributes of entities. This category matches the passive category mentioned by Precece and Rombach.

The ***Control and Change*** category includes purposes such as to *improve* and to *monitor and control*. These purposes are not just intended for understanding the current state or performance of an entity, but also for improving it by introducing changes. In addition, these purposes can be achieved taking into account one or two quality views considering the relationships shown in Fig. 2. The *improve* purpose is related with the identification of problems, vulnerabilities and others opportunities for improving the entity quality. So it always implies changes. In turn, the *monitor and control* purpose implies performing a critical analysis in order to maintain under control the detected performance problems by doing corrective actions, if needed. Control and Change category is similar to the active category considered in [13].

The ***Adopt an Alternative*** category includes purposes such as to *select an alternative* and to *compare and adopt*. These purposes share one activity which is devoted to pre-

select competitive or representative entities. The *select an alternative* purpose implies to understand the current state of each preselected competitive entity and therefore select the one with best performance. Note that the selected entity does not undergo changes, but rather is adopted for installation and/or use. On the other hand, the *compare and adopt* purpose is based on determining strengths and weaknesses of a set of representative entities in order to adopt strengths in an existing entity or in one to be developed. Thus, this purpose may imply changes on the existing entity. In the next Section, the *comparing and adopting* strategy is introduced.

4 One Evaluation Strategy from the Family of Strategies

Given an evaluation goal, the selection of the suitable strategy from a family of strategies is a critical aspect. The strategy defines a specific course of action to be followed for achieving the goal purpose. In order to illustrate a ME/MEC strategy, we use an evaluation scenario for the *compare and adopt* purpose. This scenario includes the business goal purpose, the amount/type of involved views, the description and example of the scenario, and the specification of the generic process of the strategy:

Business goal purpose: Compare and adopt / Amount of views: One / Type: Quality

Scenario description: the business goal purpose at operational level is to compare characteristics and attributes from a set of representative entities, in a given context, with the aim of adopting recommended strengths through the systematic use of a comparing strategy driven by ME, analysis and eventually, change. The measurement activity is performed by quantifying attributes (using metrics) and the evaluation activity is performed by interpreting characteristics and attributes (using indicators). The comparative analysis is based on determining strengths and weaknesses of the preselected entities in a given moment, for recommending and adopting the detected strengths in a new entity or in one that already exists.

Evaluation scenario example:

- *Business goal statement:* compare a set of social network applications for smartphones with the aim of adopting the best Usability characteristics in a new social network application for smartphones.
- *Quality View:* System Quality View / *Entity category:* System / *Quality Focus:* External Quality / *Entity subcategory:* Social network applications for smartphones.
- *Concrete entities:* Facebook, Instagram, Twitter and LinkedIn.
- *Characteristics:* Usability (*Sub-characteristics:* Comprehensibility, Learnability, Operability and User error protection).

Strategy to be applied: GOCAMECom (*Goal-Oriented Context-Aware Measurement, Evaluation and Comparison*).

Generic process specification for GOCAMECom: see Fig. 3.

Process description: The GOCAMECom process begins with the preselection of representative entities to be compared (A0 activity in Fig. 3). The preselection can be based on methods such as expert judgment, case studies, among others. Note that if the strengths will be adopted in an existing entity, this should be one out of the preselected entities. Then, A1 is devoted to define the characteristics and attributes to be evaluated.

In the A2 activity the ME are designed by selecting the most suitable metrics and indicators from a repository. Additionally, the A4 activity is divided into two sub-activities. In A4.1 the analysis is designed, which includes, among others aspects, the establishment of criteria for recommending and adopting. A4.1 can be done in parallel with A3, which involves implementing the ME per each preselected entity (producing measures and indicator's values). In the sequel, A4.2 uses the measures, indicator's values and the analysis specification as input, and produces the recommendation report as output. This report contains the strengths to be adopted in an existing target entity or in a new one to be developed. If the target entity already exists and weaknesses were detected, then changes should be carried out for adopting the strengths of the other compared entities. Consequently, for designing and implementing changes, A5 and A6 must be done respectively. Otherwise, for a new entity, A5 and A6 should not be performed.

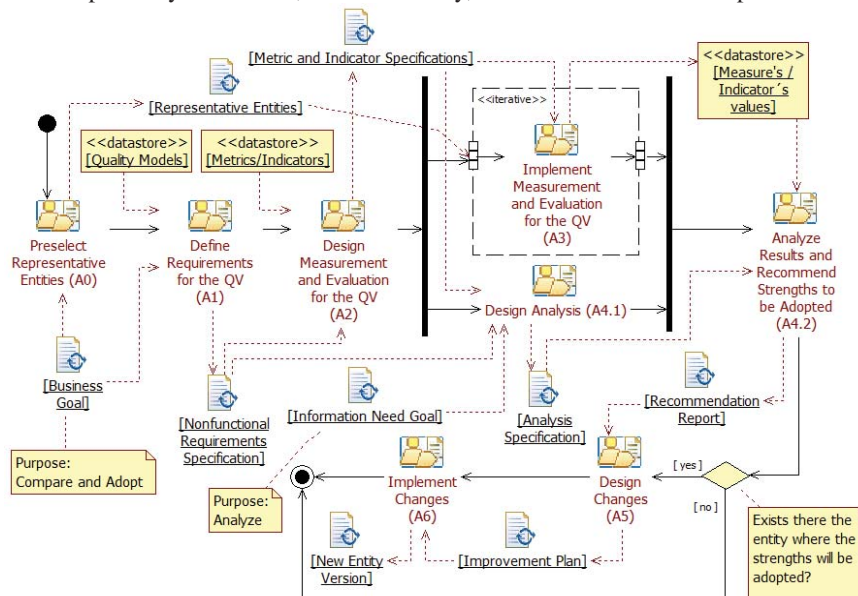


Fig. 3. Generic process specification for the GOCAMECom strategy for one Quality View (QV)

5 Application of the *Comparing and Adopting* Strategy

In the above scenario example, the statement of the business goal contains the *compare and adopt* purpose. This belongs to the *Adopt an Alternative* category. Also, the amount of involved quality views is one: The *System Quality View*. So, GOCAMECom is the strategy to be instantiated for achieving this goal purpose. Let's describe its activities.

The first performed activity was A0, which allowed to select a set of representative entities to be evaluated for analyzing their strengths to be adopted. The selection was done at random from the 15 most popular social network apps listed at www.ebizmba.com/articles/social-networking-websites. As a result, four representative

entities were selected: Instagram (v.10.0.0), Facebook (v. 106.0.0.26.28), Twitter (v. 4.0.100) and LinkedIn (v. 6.28.0). Note that the Android version per each selected app was the most recent at the time of this study (in Dec., 2016).

Name: Performance Level of the Foreign Language Support (P_FLS)	
Author: Santos L.	Version: 1.1
Elementary model: Specification: the mapping is: P_FLS=0 iff FL _{SL} =0; P_FLS=80 iff FL _{SL} =1; P_FLS=100 iff FL _{SL} =2.	
Decision criterion (3 acceptability levels):	
Name 1: Unsatisfactory; Range: if $0 \leq P_FLS < 80$	
Description: Indicates that it is not advisable to be adopted.	
Name 2: Marginal; Range: if $80 \leq P_FLS < 90$	
Description: Indicates that it could be adopted.	
Name 3: Satisfactory; Range: if $90 \leq P_FLS \leq 100$	
Description: Indicates that it is advisable to be adopted.	
Numerical Scale: Value Type: real Scale Type: ratio Unit: Name: percentage Acronym: %	

Fig. 4. Elementary indicator specification for the *Foreign Language Support* attribute. Note that FL_{SL} stands for the *Foreign Language Support Level* metric, with 0, 1 or 2 allowed values.

Table 1. Evaluation results (in [%]) for the Usability characteristics and attributes for the 4 selected apps. The green color means “satisfactory”; yellow “marginal”, and red “unsatisfactory”.

Characteristics and Attributes (in <i>italic</i>)	Facebook	Twitter	Instagram	LinkedIn
1. Usability	87.56 ●	85.65 ●	70.47 ●	69.85 ●
1.1. Understandability	91.15 ●	85.12 ●	89.82 ●	35.88 ●
1.1.1. Familiarity	91.15 ●	85.12 ●	89.82 ●	35.88 ●
1.1.1.1. <i>Global Organization Scheme Understandability</i>	100 ●	100 ●	100 ●	0 ●
1.1.1.2. <i>Control Icon Ease to be Recognized</i>	77.88 ●	87.80 ●	74.56 ●	89.71 ●
1.1.1.2.1. <i>Main Control Icon Ease to be Recognized</i>	85.71 ●	80 ●	80 ●	85.71 ●
1.1.1.2.2. <i>Contextual Control Icon Ease to be Recognized</i>	66.67 ●	100 ●	66.67 ●	95.83 ●
1.1.1.3. <i>Foreign Language Support</i>	100 ●	0 ●	100 ●	0 ●
1.2. Learnability	88.22 ●	81.34 ●	77.10 ●	81.23 ●
1.2.1. Feedback Suitability	86.46 ●	79.38 ●	74.25 ●	87.65 ●
1.2.1.1. <i>Current Location Feedback Appropriateness</i>	66.67 ●	100 ●	80 ●	83.33 ●
1.2.1.2. <i>Alert Notification Feedback Appropriateness</i>	100 ●	62.50 ●	100 ●	87.50 ●
1.2.1.3. <i>Error Message Appropriateness</i>	93.75 ●	75 ●	37.50 ●	92.86 ●
1.2.2. Helpfulness	90 ●	83.33 ●	80 ●	75 ●
1.2.2.1. <i>Context-Sensitive Help Appropriateness</i>	90 ●	83.33 ●	80 ●	75 ●
1.3. Operability	80.96 ●	98.15 ●	71.76 ●	89.19 ●
1.3.1. Data Entry Ease	80 ●	100 ●	60 ●	80 ●
1.3.1.1. <i>Defaults</i>	100 ●	100 ●	100 ●	100 ●
1.3.1.2. <i>Mandatory Entry</i>	50 ●	100 ●	50 ●	50 ●
1.3.1.3. <i>Widget Entry Availability</i>	100 ●	100 ●	0 ●	100 ●
1.3.2. Visibility	77.99 ●	90.92 ●	78.05 ●	86.91 ●
1.3.2.1. <i>Color Visibility Appropriateness</i>	77.99 ●	90.92 ●	78.05 ●	86.91 ●
1.3.2.1.1. <i>Brightness Difference Appropriateness</i>	78.33 ●	91.40 ●	79.70 ●	87.08 ●
1.3.2.1.2. <i>Color Difference Appropriateness</i>	77.65 ●	90.44 ●	76.44 ●	86.74 ●
1.3.3. Consistency	83.45 ●	100 ●	81.13 ●	100 ●
1.3.3.1. <i>Permanence of Controls</i>	83.45 ●	100 ●	81.13 ●	100 ●
1.3.3.1.1. <i>Permanence of Main Controls</i>	84 ●	100 ●	80 ●	100 ●
1.3.3.1.2. <i>Permanence of Contextual Controls</i>	81.25 ●	100 ●	85.71 ●	100 ●
1.4. User Error Protection	91.67 ●	77.50 ●	55 ●	75 ●
1.4.1. Error Management	91.67 ●	77.50 ●	55 ●	75 ●
1.4.1.1. <i>Error prevention</i>	83.33 ●	80 ●	60 ●	50 ●
1.4.1.2. <i>Error recovery</i>	100 ●	75 ●	50 ●	100 ●

Then, the A1 activity produced the “*Specification of nonfunctional requirements*” document. For the *Usability* characteristic, 17 attributes were used -see their definitions in [9]. These are related to some of the following sub-characteristics: *Comprehensibility*, *Learnability*, *Operability* and *User Error Protection*. In the first column of Table 1, the resulting *Usability* requirements tree can be observed.

For the A2 activity, a metric and an elementary indicator from the Metrics/Indicators repository were selected per each attribute, as well as a derived indicator for the *Usability* characteristic. Fig. 4 specifies the elementary indicator for the “*Foreign Language Support*” attribute (coded 1.1.1.3 in Table 1). It is worth mentioning that all indicators have the percentage scale interpreted by three acceptability levels, viz. “*Satisfactory*” in the [90-100%] range; “*Marginal*” [80-90); and “*Unsatisfactory*” [0-80).

In the analysis design (A4.1), a set of criteria was agreed for deciding whether an entity attribute should be considered as a strength to be adopted or not. Thus, an attribute is considered a strength if its indicator value falls in the “*Satisfactory*” acceptability level. When several entities have the same strength attribute, the one with the highest score is recommended to be adopted. In case that they have the same score, any of them could be recommended. For those attributes which are not benchmarked satisfactorily in any selected entity but fall in the “*Marginal*” level, their adoption could be considered taking into account that some kind of improvement must be included.

Concurrently to A4.1, the A3 activity was carried out. This produces the values which were stored in the *Measure’s and Indicator’s values* repository. Table 1 shows all the indicators values for each evaluated social network app.

Lastly, A4.2 produces a “*Recommendation Report*” following the criteria documented in the “*Analysis specification*”, as depicted in Fig. 3. Looking at the resulting *Usability* indicator values in Table 1, no app reached the “*Satisfactory*” acceptability level. Nevertheless, within the “*Marginal*” level, Facebook got the best score (87.56%) followed by Twitter (85.65%). While Instagram and LinkedIn fell into the “*Unsatisfactory*” acceptability level, with 70.47% and 69.85% values respectively.

Looking the analysis at attribute level, only Twitter obtained the maximum score (100%) for *Contextual Control Icon Ease to be Recognized* (1.1.1.2.2.), *Current Location Feedback Appropriateness* (1.2.1.1.) and *Mandatory Entry* (1.3.1.2.). Also, it has the highest score for the *Brightness Difference Appropriateness* (1.3.2.1.1.) and *Color Difference Appropriateness* (1.3.2.1.2.) attributes, which reached 91.40% and 90.44% respectively. These five attributes are strengths in the Twitter app and therefore are recommended to be adopted in the new entity.

On the other hand, Facebook met the highest score for the *Error Message Appropriateness* (1.2.1.3.) and *Context-Sensitive Help Appropriateness* (1.2.2.1.) attributes, i.e., 93.75% and 90% respectively. Thus, these attributes should be adopted from the Facebook mobile app.

In this way, the rest of the attributes were analyzed in order to decide their adoption. Once A4.2 was finished, the GOCAMECom process for this case study finished as well. Recall that change activities (A5 and A6) are just performed when the target entity already exists, as shown in Fig. 3.

6 Related Work and Discussion

Every organization seeks to achieve its business goals successfully. Therefore, it is advisable to have an approach which allows the establishment of goals at different organizational levels, as well as organizing the work by means of projects and strategies for helping to reach goals. For this, it is desirable that the approach includes the following characteristics: i) the definition of business goals and information need goals at different organizational levels; ii) the definition of different evaluation purposes for goals, and; iii) the conception of a family of ME/MEC strategies that helps to achieve goals. In this sense, we have developed the approach called *Holistic Quality Multilevel and Multipurpose Evaluation Approach* [15].

In the present work, we have considered aspects of the two latter approach's features. Regarding the evaluation purposes, many of them are documented in several works [3, 4, 7, 8, 12, 13]. However, there is no broad consensus yet in some purpose definitions. Even more, just in [13] a classification for purposes is presented, but it does not allow to encompass the intentionality of some purposes. Therefore, in Section 3 we proposed three new categories that permit this.

Regarding the third approach feature, in the current literature there is few papers which address the importance of having a family of ME/MEC strategies for helping to achieve different evaluation purposes, considering also quality views. For example, a MEC strategy that integrates simultaneously the domain conceptual base, the process specification and the method specification is presented in [11]. But in this work a family of strategies for different purposes is not considered. Also, in [5] the *Cloud MoS@RT* strategy is presented to *monitor* services in the cloud. This strategy has the process specification modeled in SPEM in addition to a set of methods and models. However, authors do not discuss about a family of strategies for different evaluation purposes.

7 Conclusions and Future Work

In this paper we have analyzed a categorization for a set of evaluation purposes in which business and information need goals can be satisfied through the systematic use of a family of strategies driven by ME, analysis and, eventually, change. Also, the GOCAMECom strategy has been illustrated for the comparison of four social network apps, which helps us to determine the strengths to be adopted in a new smartphone app.

With regard to the first paper contribution, three categories for evaluation purposes have been proposed. The first category, *Characterize and Understand*, includes the understand, monitor and predict purposes. The second one, *Control and Change*, includes the improve, and monitor and control purposes. The last category, *Adopt an Alternative*, embraces the compare and adopt, and select an alternative purposes. We argue that the proposed categories may foster a clearer understanding and communication of the purposes' intention to different stakeholders. We can also highlight that purposes in the *Characterize and Understand* category serve as a basis for the remainder category purposes. For example, it is necessary to understand the current state of an entity before

implementing changes for improvement; likewise, it is necessary to understand the current state of representative/competitive entities before making thoughtful decisions. Regarding the second contribution, the usefulness of a specific strategy has been illustrated for helping to achieve the *compare and adopt* purpose.

Lastly, a limitation of the current work is the lack of a broader validation that the different ME/MEC strategies fit well to the corresponding purpose and category. Hence, as an ongoing work, we are designing surveys in order to confirm their appropriateness with domain experts.

Acknowledgment. This work and line of research are supported by Science and Technology Agency of Argentina, in the PICT 2014-1224 project at UNLPam.

References

1. Basili V., Lindvall M., Regardie M., Seaman C., Heidrich J., Jurgen M., Rombach D., Trendowicz A.: Linking Software Development and Business Strategy through Measurement, *IEEE Computer*, 43:(4), pp. 57–65 (2010)
2. Becker P., Papa F., Olsina L.: Process Ontology Specification for Enhancing the Process Compliance of a Measurement and Evaluation Strategy, *CLEI E.Jnal*, 18:(1), pp. 1-26, (2015)
3. Briand L., Differding Ch., Rombach D.: Practical Guidelines for Measurement-based Process Improvement, *Software Process Improvement and Practice Jnal*, 2:(4), pp. 253-280, (1996)
4. Briand L., Morasca S., Basili V.: An Operational Process for Goal-driven Definition of Measures, *IEEE Transactions on Software Engineering*, 28:(12), pp. 1106-1125, (2002)
5. Cedillo P., Gonzalez-Huerta J., Insrán E., Abrahao, S.: Towards Monitoring Cloud Services Using Models@run time. In *Wkshp on Models@run.time, MODELS*, Valencia, Spain, pp. 31-40, (2014)
6. CMMI: Capability Maturity Model Integration, for Dev. V1.3. CMU/SEI:TR-033, (2010)
7. Fenton N., Pfleeger S.: *Software Metric: a Rigorous and Practical Approach*, 2nd Ed., PWS Publishing Company, (1996)
8. INCOSE Systems Engineering Measurement Primer: A Basic Introduction to Measurement Concepts and Use for Systems Eng., INCOSE-TP-2010-005-02, v2, (2010)
9. Olsina L., Santos L., Lew P.: Evaluating Mobileapp Usability: A Holistic Quality Approach, In: *ICWE 2014*, Casteleyn, Rossi, and Winckler (Eds.): LNCS 8541, pp. 111-129, (2014)
10. OMG: Business Motivation Mode (BMM), V1.3, (2015)
11. Papa M. F.: Toward the Improvement of a Measurement and Evaluation Strategy from a Comparative Study, In: *Current Trends in Web Engineering, ICWE Int'l Workshops*, Grossniklauss and Wimmer (Eds.), LNCS 7703, pp. 189-203, (2012)
12. PMBOK: A Guide to the Project Management Body of Knowledge, 5th Edition, (2013)
13. Preece J., Rombach D.: A Taxonomy for Combining Software Engineering and Human-Computer Interaction Measurement Approaches: Towards a Common Framework. *Int'l Jnal. of Human-Computer Studies*, 41:(4), pp. 553-583, (1994)
14. Rivera B., Becker P., Olsina L.: Quality Views and Strategy Patterns for Evaluating and Improving Quality: Usability and User Experience Case Studies, In: *Journal of Web Engineering*, Rinton Press, USA, 15:(5&6), pp.433-464, (2016)
15. Rivera B., Becker P., Papa F., Olsina L.: A Holistic Quality Evaluation, Selection, and Improvement Approach driven by Multilevel Goals and Strategies. *CLEI E.Jnal.*, 19: 3, (2016)

Una Plataforma para el Desarrollo de LPS en el Dominio Geográfico: Caso de Estudio en el Subdominio de Paleontología

Fiorella Pesce¹ *, Sofia Caballero¹, Agustina Buccella^{1,2}, and Alejandra Cechich¹

¹ GIISCO Research Group

Departamento de Ingeniería de Sistemas - Facultad de Informática

Universidad Nacional del Comahue

Neuquen, Argentina

{fiorella.pesce,sofia.caballero,agustina.buccella,alejandra.cechich}@fi.uncoma.edu.ar

² CONICET

Abstract. La Líneas de Productos de Software (LPS) son un paradigma orientado al reuso de software dentro de dominios particulares. Aspectos claves dentro de este paradigma son las particularidades inherentes de estos dominios y las técnicas aplicadas para sistematizar las formas de maximizar el reuso dentro de los mismos. En este artículo describimos un proceso para la creación de líneas de productos orientado al reuso de jerarquías de dominios, partiendo del dominio geográfico en general y considerando el subdominio paleontológico en particular. Nuestro proceso se basa en estandarizaciones y técnicas previas ya aplicadas para otro subdominio geográfico, el de ecología marina. Aquí mostramos cómo estas técnicas son aplicadas en el subdominio de paleontología, mejorando el reuso sistemático de artefactos software.

Keywords: Líneas de Productos de Software, Reuso de Dominios, Dominio Geográfico, Paleontología

1 Introducción

La Ingeniería de Líneas de Productos de Software (ILPS) [6, 13, 18] propone un proceso de desarrollo de software cuyo principal objetivo es maximizar el reuso de artefactos de manera de obtener aplicaciones más rápidas y de mejor calidad. Como el desarrollo de las LPSs propone un reuso dentro de dominios particulares, el éxito del desarrollo depende de la identificación, uso y administración de los artefactos dentro de cada uno de ellos; y por lo tanto, la aplicación de técnicas específicas para la sistematización del reuso es crucial.

El dominio geográfico [11] es amplio ya que incluye aspectos generales aplicables a todos los productos dentro de este dominio. Aquí, la creación de una LPS

*

puede resultar impracticable (la cantidad de variabilidad que habría que definir dentro de cada servicio se volvería inmanejable), por lo que es lógico pensar en la división en diferentes subdominios donde cada uno posea ciertas características del dominio geográfico común, como realizar zoom en un mapa, obtención de coordenadas, etc; y ciertas características especiales como conocer la profundidad del océano en determinadas zonas. De esta forma se construyen componentes pensados para formar parte de una jerarquía de dominios, es decir, se definen un conjunto de servicios generales a nivel del dominio genérico geográfico aplicables a cualquier línea a ser implementada dentro de los subdominios que incluya.

En la literatura las LPSs son un área de investigación muy activa en donde existen muchas propuestas de técnicas o métodos que mejoren las actividades involucradas en su desarrollo. Esto se puede ver en la cantidad de revisiones literarias o sistemáticas que existen hoy en día enfocadas en técnicas de modelado, implementación, validación, etc. [4, 5, 12, 16]. Enfoques similares a nuestra propuesta orientada a una jerarquía de dominios pueden ser aquellos relacionados con el desarrollo de ecosistemas de software o multi-líneas de software [7, 15]. En general, estos trabajos se enfocan en la problemática de haber desarrollado componentes para dominios diferentes que deben luego trabajar juntos al ser integrados dentro de un mismo desarrollo. En nuestro caso, estos problemas son minimizados ya que los componentes creados son diseñados para formar parte de una jerarquía dentro del dominio geográfico compartiendo estandarizaciones ya definidas.

Una primera aproximación a nuestro enfoque ha sido presentado en [9, 10] en donde hemos construido una LPS dentro del subdominio de ecología marina, pero tomando servicios comunes de dominios de más alto nivel como el oceanográfico y el geográfico. En esos trabajos hemos definido una metodología para la creación de LPSs que diseña e implementa una serie de *artefactos de software* necesarios para comunicar y modelar el dominio basado en estandarizaciones y técnicas bien definidas. En este artículo mostramos el proceso de desarrollo para la creación de LPSs, basado en la metodología previamente presentada [9, 10] para el subdominio de ecología marina, pero enfocándonos en el reuso de los artefactos creados y adaptados al subdominio paleontológico. En particular nos centramos en tres de los artefactos de software creados como parte de la ingeniería de dominio.

Este artículo está organizado de la siguiente manera. En la siguiente sección describimos nuestro proceso de desarrollo de LPSs orientado al dominio geográfico y sus subdominios. En la Sección 3 mostramos un caso de estudio en donde aplicamos dicho proceso al subdominio de paleontología en lo relevante a la información asociada a las excavaciones para encontrar piezas paleontológicas. Luego, realizamos una validación preliminar en donde analizamos mejoras debido al reuso efectuado. Por último describimos las conclusiones y trabajos futuros.

2 Proceso de Desarrollo para la creación de Líneas de Productos de Software Orientada a Subdominios

La Figura 1 muestra los cuatro pasos de nuestro proceso para la ingeniería de dominios en el desarrollo de una LPS. Recordemos que esta ingeniería es la que se encarga de identificar, capturar y organizar toda la información fuente recopilada. Como resultado, genera una plataforma de software con un conjunto de artefactos reusables y configurables que proveen un entorno en común [18].

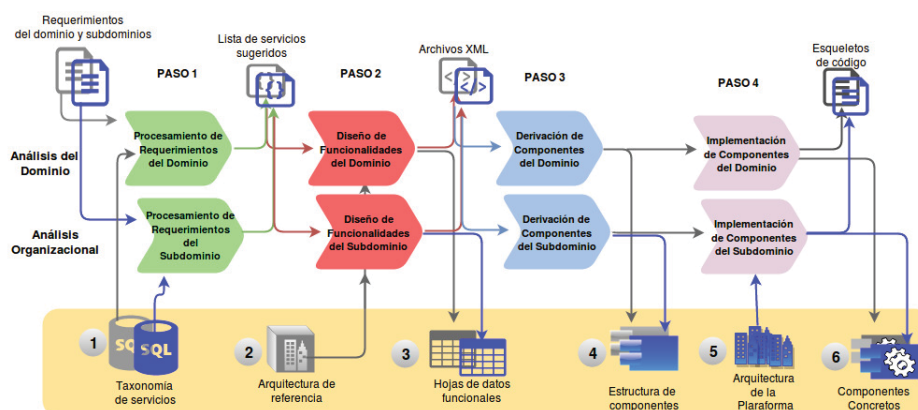


Fig. 1. Proceso de desarrollo como parte de la ingeniería de dominio de las LPSs junto con los artefactos de software creados y utilizados

En nuestra metodología de desarrollo, la ingeniería de dominios se divide en dos tipos de análisis: de *dominio* y *organizacional*. El primero involucra el análisis y diseño de la información dentro de un dominio específico pero enfocándose en una vista general. Luego, el análisis organizacional utiliza la información anterior para adaptarla al contexto de la LPS que está siendo desarrollada. En nuestro caso de aplicación, dentro del análisis de dominio se debe analizar el dominio geográfico general obteniendo servicios comunes al conjunto de subdominios que podrían ser de interés. Luego, en el análisis organizacional, las actividades están enfocadas en el subdominio específico, en nuestro caso, en el subdominio paleontológico.

En la Figura 1 podemos observar los seis artefactos de software que deben ser desarrollados o utilizados en los cuatro pasos involucrados en la ingeniería de dominio. En general cada artefacto es construido durante el análisis del dominio (artefactos en color gris) y refinados y completados en el análisis organizacional (artefactos en color azul). El primero de ellos, el cual es una de las entradas del paso 1, es una *taxonomía de servicios* del dominio analizado construida como una estructura jerárquica, con las categorías por las cuales clasificar los servicios. En trabajos previos, esta taxonomía ha sido creada para el dominio geográfico

considerando el estándar de Servicios de Arquitectura (definido en el OpenGIS Service Architecture)³, y la ISO/DIS 19119⁴ como base para definir los servicios y sus categorías [10]. La taxonomía es creada en base a los requerimientos de los usuarios expertos, en forma general (en el análisis de dominio), y refinada y especializada en el análisis organizacional para contener aquellos servicios particulares del subdominio paleontológico. Luego, la salida de este paso es la lista de servicios requeridos en el subdominio para el diseño de las funcionalidades.

El segundo artefacto es una *arquitectura de referencia* y es utilizada como entrada en el paso 2 de *diseño de funcionalidades*. Dicha arquitectura debe especificar una estructura preliminar para la interacción de los servicios definidos en la taxonomía. Como podemos observar, sólo existe una instancia de la misma en color gris debido a que es extraída y reusada de la especificada en la norma ISO 19119. A su vez, en este paso se deben crear las *hojas de datos funcionales* que poseen cada una de las funcionalidades del dominio geográfico general (en gris) y particular (en azul). Las mismas se diseñan mediante la interacción de los servicios de la taxonomía necesarios. También, cada hoja funcional se representa mediante un conjunto de archivos XML que permiten analizar automáticamente estos modelos en busca de inconsistencias o incompatibilidades a la hora de especificar la variabilidad [8, 20]. Luego, en el paso 3 se debe realizar la *derivación de componentes abstractos* para crear componentes reusables en base a la información definida en las hojas de datos funcionales [1, 19]. Como parte de este paso y el siguiente (*Implementación de componentes*) hemos definido en trabajos previos mecanismos iniciales que asisten a la creación del cuarto artefacto la (*estructura de componentes*) [2].

Luego, los artefactos refinados determinan la estructura del software que conforma el quinto artefacto, la *arquitectura de la plataforma*. Ésta contiene la forma en que cada hoja de datos funcional refinada (en base al refinamiento de la taxonomía) se implementa como componentes de software. Esta plataforma es luego utilizada en la configuración de los productos para crear la arquitectura de una aplicación específica.

3 Un Caso de Estudio en el Subdominio Paleontológico

Primero es necesario introducir algunos aspectos particulares de este subdominio. En la Figura 2 mostramos una parte del modelo conceptual en donde nos enfocamos en las clases que nos sirven para comprender luego las funcionalidades definidas. Para construir el mismo, además de la información elicitada de los usuarios expertos en el subdominio de paleontología, hemos extraído información de las ISO 19109 y 19107 para datos geográficos, y las normas ISO 21127:2014⁵,

³ The OpenGIS Abstract Specification: Service Architecture, 2002.

⁴ Geographic information. Services International Standard 19119, ISO/IEC, 2005.

⁵ Information and documentation – A reference ontology for the interchange of cultural heritage information

LIDO⁶ y CIDOC⁷. Estas tres últimas normas son específicas del subdominio paleontológico ya que definen y clasifican servicios y protocolos requeridos para la gestión de colecciones dentro de los museos y la representación de las piezas paleontológicas encontradas.

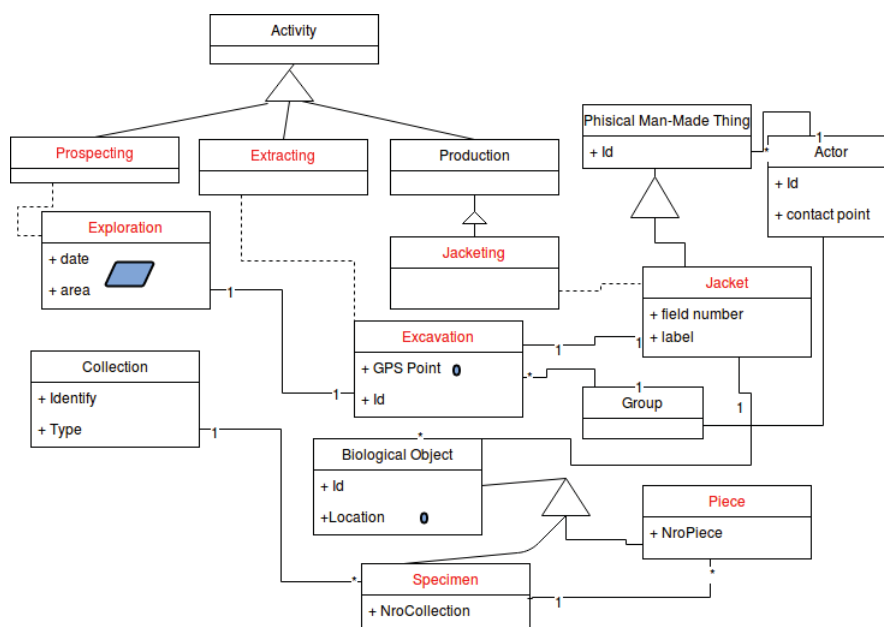


Fig. 2. Parte del Modelo de Datos para el Subdominio Paleontológico

La Figura 2 muestra en color rojo aquellas clases que han sido definidas por los usuarios expertos y en color negro las extraídas de las tres últimas normas. Según la ISO 21127 cualquier acción que se realice sobre un objeto físico, se representa como una actividad. En particular, la actividad de *exploración* (*prospecting*) es la que releva zonas en las que puedan realizarse las *excavaciones* (*extracting*) a cargo de un grupo de profesionales. En el caso de encontrar algún *especimen*, se inicia la actividad de armado de un *bochón* (*jacket*) (mediante la actividad de *jacketing*), el cual puede contener *objetos biológicos* – *especimen* o *pieza*. A éstos se les aplican procedimientos mecánicos y/o químicos para retirar correctamente el objeto encontrado. A su vez, la información que debe ser geográfica, como las áreas y puntos, ha sido representada utilizando el enfoque MADS (Modeling of Application Data with Spatio-temporal features) [17].

⁶ LIDO - Lightweight Information Describing Objects Version 1.0 - <http://network.icom.museum/cidoc/working-groups/lido/what-is-lido/>

⁷ CIDOC Conceptual Reference Model Version 6.0 - <http://www.cidoc-crm.org>

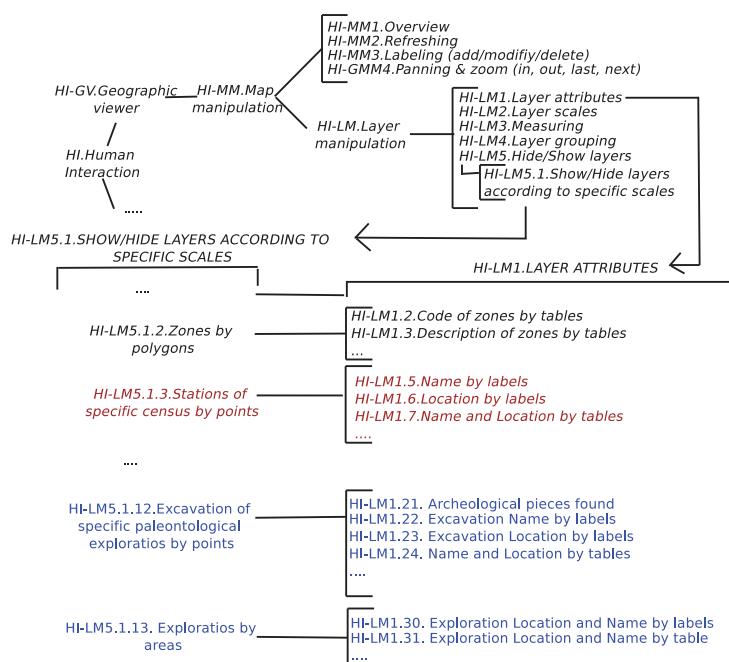


Fig. 3. Algunos servicios de la categoría *interacción humana* (HI) definidos en la jerarquía de dominios

Luego, en base a este modelo conceptual y a los requerimientos funcionales del subdominio creamos la *taxonomía de servicios paleontológicos*. Ésta ha sido definida siguiendo las bases descritas en [10] mediante la realización de un proceso iterativo de trabajo entre todos los participantes del subdominio, usuarios expertos, ingenieros de software y desarrolladores, y utilizando como base la taxonomía geográfica definida en la ISO 19119. También aquí se utilizaron las normas específicas del subdominio (ya mencionadas previamente) y la arquitectura de referencia de la ISO 19119 ya aplicada para la LPS en el subdominio de ecología marina [9].

La taxonomía, que se define en base a las capas de la arquitectura, posee un conjunto de categorías principales en las cuales deben estar incluidos los servicios correspondientes. Así en la Figura 3 podemos observar algunos servicios de la categoría *interacción humana* (HI) ya definidos previamente [10]. Estos servicios pertenecen al dominio geográfico (en color negro y cursiva), y al subdominio de ecología marina (en color rojo y cursiva). A su vez también están definidos los servicios del subdominio paleontológico (en color azul), en donde podemos observar algunos orientados a las formas de visualización de la información. En este caso vemos los atributos de exploraciones y excavaciones que pueden ser mostrados en forma de tablas o por medio de etiquetas directamente sobre un mapa.

De esta manera, como resultado para cada categoría, se obtiene la taxonomía completa dentro del subdominio paleontológico con servicios reusados y heredados del dominio geográfico.

El siguiente paso es la creación de las hojas de datos funcionales (tercer artefacto) con las funcionalidades del subdominio en base a los servicios de la taxonomía paleontológica y la arquitectura de referencia. La confección de estas hojas de datos sigue los mismos lineamientos definidos en trabajos previos [1, 14] y utiliza nuestra herramienta de soporte llamada *Datasheet Modeler* [14]. Ésta permite crear las hojas mediante la utilización de dependencias de variabilidad [1] y servicios de la taxonomía, traduciendo a archivos XML procesables por la computadora. La Figura 4 muestra la funcionalidad *Registrar excavaciones*, la cual permite registrar una nueva excavación dentro de una exploración ya existente en el sistema. Como podemos observar, se realiza una interacción entre los servicios de la taxonomía ya definida de manera de poder llevar cabo la funcionalidad. La variabilidad representada define en este caso que la carga de datos debe contemplar que pueda realizarse en lugares sin conexiones a internet o dispositivos GPS. Por lo tanto, datos de las excavaciones (tales como coordenadas geográficas y objetos encontrados) deben permitirse mediante el ingreso manual de información. A su vez, en caso se requiera en algún producto, también se contemplan las opciones de cargar los datos que provienen de planillas de cálculo, y/o a través de dispositivos que sí contengan GPS.

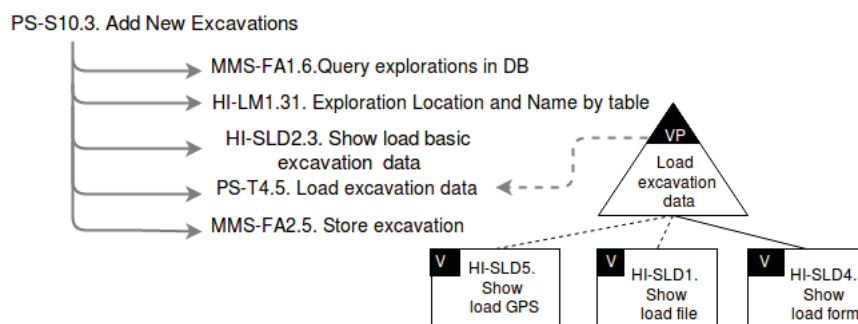


Fig. 4. Representación de la funcionalidad *Registrar Excavaciones* en una Hoja de Datos Funcional

Luego, como salida de la herramienta se crean automáticamente los archivos XML de cada una de estas hojas de datos. Estos archivos nos permiten procesarlas automáticamente para validarlas [19, 20] y crear una estructura preliminar de componentes reusables (artefacto 4) [2, 3], que luego formarán parte de la arquitectura de la plataforma (artefactos 5). Éstos deben ser componentes implementados (artefacto 6) en el último paso.

4 Una Evaluación Preliminar del Reuso Alcanzado

Para realizar una evaluación, hemos realizado un análisis del grado de reuso logrado en dos de los artefactos creados y descriptos en la sección anterior.

Primero hemos analizado el reuso dentro de la taxonomía de servicios considerando la cantidad de servicios que hemos definido para el subdominio paleontológico. Para esta evaluación hemos analizado tres tipos de servicios:

- *Servicios completamente reusados (SCR)* que son aquellos que están dentro del dominio geográfico o de algún subdominio y son utilizados sin modificación dentro de las funcionalidades de la LPS paleontológica. Ejemplos de estos servicios son los que están en color negro y rojo en la Figura 3 y que son de incumbencia para este subdominio.
- *Servicios reusados por especialización (SRE)* que son servicios nuevos pero generados a partir de servicios ya definidos para otros dominios. Ejemplos de estos servicios son los de color azul en la Figura 3.
- *Servicios sin reuso (SSR)* que son aquellos creados completamente nuevos sin ninguna relación con los servicios de la taxonomía previamente desarrollada.

En base a esos tres tipos de servicios, y de un total de unos 120 servicios definidos como útiles, identificamos un 25% como SCR, un 58.3 % como SRE y un 16.6% como SSR. La Figura 5a) muestra estos porcentajes gráficamente.

Luego, para analizar las funcionalidades, hemos simulado la creación de un producto en donde elegimos un conjunto de diez funcionalidades (definidas en las hojas de datos funcionales) como parte de la plataforma LPS y las hemos instanciado aleatoriamente, de manera de generar un nuevo producto. Para ilustrar este proceso, en la Figura 6 mostramos la hoja de datos resultante de la funcionalidad *Registrar Excavaciones* en donde hemos instanciado la variabilidad para permitir además de la carga manual por formularios, también la carga por medio de dispositivos GPS. De esta manera, el producto resultante contendrá ambas opciones para la carga de una excavación.



Fig. 5. Análisis del reuso logrado en la taxonomía y las hojas de datos funcionales instanciadas

Este proceso se efectuó por cada una de las 10 funcionalidades que fueron utilizadas para la creación del producto. Para analizar el reuso hemos contado la

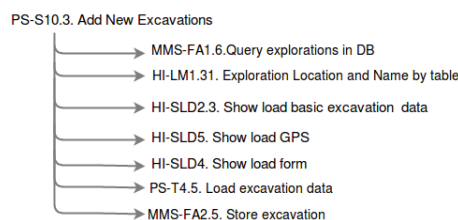


Fig. 6. Funcionalidad *Registrar Excavaciones* instanciada para un producto en particular

cantidad de *servicios de otros dominios (SOD)* utilizados en la funcionalidades y la cantidad de *servicios propios del dominio (SPD)* de la taxonomía. Los SOD son aquellos que hemos clasificado como la suma de los SCR y SRE y los SPD son los SSR. De este análisis, surgen los resultados mostrados en la Figura 5b) en donde podemos observar un porcentaje de aprox. un 65% de SOD y un 35% de SPD. Este resultado resulta razonable considerando que muchas funcionalidades utilizan servicios SRE (especializados para el subdominio).

5 Conclusiones y Trabajo Futuro

En este trabajo hemos mostrado un proceso para el desarrollo de LPS orientado a reuso de artefactos de software dentro de una jerarquía de dominios. A partir de trabajos previos dentro del dominio geográfico, hemos generado y reusado artefactos en un nuevo subdominio en este caso paleontológico. A su vez hemos realizado un análisis preliminar del reuso logrado en la construcción de alguno de esos artefactos mostrando resultados prometedores. Como trabajo futuro, entenderemos el proceso hacia los artefactos que están involucrados en la ingeniería de la aplicación de las LPSs.

References

1. A. Buccella, M. Pol'la, A.C., Arias, M.: A variability representation approach based on domain service taxonomies and their dependencies. In: International Conference of the Chilean Computer Science Society (SCCC). Talca, Chile (2014)
2. Arias, M., Buccella, A., Cechich, A.: Smooth transition from abstract to concrete spl components: a client-server implementation for the geographic domain. In: Proceedings of the IEEE ARGENCON'16. GRSS: IEEE GRSS, Buenos Aires, Argentina (2016)
3. Arias, M., Buccella, A., Cechich, A.: Un marco de trabajo para la gestión de requerimientos de software en líneas de productos de software. In: CLEI 2017 : XLIII Conferencia Latinoamericana en Informática. SLISW - Simposio Latinoamericano de Ingeniería de Software. To appear (2017)
4. Bashroush, R., Garba, M., Rabiser, R., Groher, I., Botterweck, G.: Case tool support for variability management in software product lines. ACM Comput. Surv. 50(1), 14:1–14:45 (Mar 2017), <http://doi.acm.org/10.1145/3034827>

5. Benavides, D., Segura, S., Ruiz-Cortés, A.: Automated analysis of feature models 20 years later: A literature review. *Information Systems* 35, 615–636 (September 2010)
6. Bosch, J.: Design and use of software architectures: adopting and evolving a product-line approach. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA (2000)
7. Bosch, J.: From software product lines to software ecosystems. In: Proceedings of the 13th International Software Product Line Conference. pp. 111–119. SPLC '09, Carnegie Mellon University, Pittsburgh, PA, USA (2009)
8. Braun, G., Pol'La, M., Cecchi, L., Buccella, A., Fillottrani, P., Cechich, A.: A dl semantics for reasoning over ovm-based variability models. In: 30th International Workshop on Description Logics (DL17). To appear, France (2017)
9. Buccella, A., Cechich, A., Arias, M., Pol'la, M., Doldan, S., Morsan, E.: Towards systematic software reuse of gis: Insights from a case study. *Computers & Geosciences* 54(0), 9 – 20 (2013), <http://www.sciencedirect.com/science/article/pii/S0098300412003913>
10. Buccella, A., Cechich, A., Pol'la, M., Arias, M., Doldan, S., Morsan, E.: Marine ecology service reuse through taxonomy-oriented SPL development. *Computers & Geosciences* 73(0), 108 – 121 (2014), <http://www.sciencedirect.com/science/article/pii/S0098300414002155>
11. Burrough, P., McDonnell, R.: Principles of Geographical Information Systems. Oxford University Press (1998)
12. Capilla, R., Snchez, A., Dueas, J.C.: An analysis of variability modeling and management tools for product line development. In: Proceedings of the Software and Services Variability Management Workshop: Concepts, Techniques, and Tools. pp. 32–47 (2007)
13. Clements, P.C., Northrop, L.: Software Product Lines : Practices and Patterns. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA (2001)
14. Mancuso, M., Buccella, A., Cechich, A., Arias, M., Pol'la, M.: Datasheet modeler: Una herramienta de soporte para el desarrollo de funcionalidades en líneas de productos de software. In: Proceedings of the CACIC'15: XXI Argentine Congress of Computer Science. Junin, Buenos Aires, Argentina (2015)
15. Manikas, K., Hansen, K.M.: Software ecosystems a systematic literature review. *Journal of Systems and Software* 86(5), 1294 – 1306 (2013)
16. Munir, Q., Shahid, M.: Software Product Line: Survey of Tools. Master's thesis, Linkopings University, Linkopings, Sweden (2010)
17. Parent, C., Spaccapietra, S., Zimányi, E.: Spatio-temporal conceptual models: data structures + space + time. In: Proceedings of the GIS '99: 7th ACM International Symposium on Advances in Geographic Information Systems. pp. 26–33. ACM Press, New York, NY, USA (1999)
18. Pohl, K., Böckle, G., Linden, F.J.v.d.: Software Product Line Engineering: Foundations, Principles and Techniques. Springer-Verlag New York, Inc., Secaucus, NJ, USA (2005)
19. Pol'la, M., Arias, M., Buccella, A., Cechich, A.: Un sistema de anotaciones para la especificación de componentes de una línea de productos de software. *Revista Tecnología y Ciencia de la Universidad Tecnológica Nacional* 6(1), 116–122 (2015)
20. Pol'la, M., Buccella, A., Arias, A., Cechich, A.: Sevatax: Service taxonomy selection & validation process for spl development. In: XXXIV International Conference of the Chilean Society of Computer Science (SCCC'15). IEEE Computer Society Press, Santiago, Chile (2015)

Una Revisión Sistemática de la Literatura en Pruebas de Compatibilidad Web

Leandro N. Sabaren, Maximiliano A. Mascheroni, Cristina L. Greiner, Emanuel Irrazábal,

Departamento de Informática. Facultad de Ciencias Exactas y Naturales y Agrimensura.
Universidad Nacional del Nordeste. Corrientes, Argentina

leans177@gmail.com, {mascheroni, cgreiner, eirrazabal}@exa.unne.edu.ar

Abstract. Los usuarios acceden a aplicaciones web desde diferentes navegadores esperando obtener la misma experiencia de usuario. Sin embargo diversas causas producen defectos de compatibilidad que afectan negativamente las funcionalidades y la interfaz gráfica. En este trabajo se busca identificar, analizar y sintetizar las técnicas, herramientas y desafíos encontrados en artículos científicos publicados sobre pruebas de compatibilidad web. La metodología elegida es la revisión sistemática de la literatura. Se analizaron y sintetizaron los hallazgos para responder los interrogantes propuestos. La técnica de pruebas más elegida es el análisis visual. El principal desafío detectado es la detección de elementos variables. La compatibilidad web es cada vez más importante de acuerdo al crecimiento de las publicaciones, pero aún no se han desarrollado técnicas adecuadas soportadas por herramientas automatizadas.

Keywords: aplicación web, prueba de compatibilidad web, pruebas cruzadas de navegador, revisión sistemática de la literatura.

1 Introducción

Al desarrollar una aplicación web, un objetivo es que sea visualizada correctamente por la mayor cantidad de usuarios posible [1], [2]. Debido a la naturaleza de la distribución de las aplicaciones web, basadas en una arquitectura cliente servidor [3], los usuarios pueden acceder a la página web desde una gran variedad de navegadores, en dispositivos y plataformas distintas. Sin embargo, las diferencias entre los navegadores y su interpretación del código de la página web pueden causar defectos de incompatibilidad, que afectan negativamente la experiencia de usuario. Es trabajo del desarrollador el proveer una experiencia aceptable a los usuarios, aun cuando ésta no sea exactamente la misma para todos.

La compatibilidad web es la característica de una página web que permite que la aplicación funcione correctamente en un número aceptable de navegadores web [4]. Resulta imposible probar el funcionamiento de la aplicación para el 100% de navegadores, en todas las plataformas [4]. El sitio web será compatible si se logra la compatibilidad web con el conjunto de navegadores acordado en los requerimientos.

Existen herramientas industriales para las pruebas, como CrossBrowserTesting¹ o BrowserStacks², aunque es necesaria la inspección visual de las capturas que generan éstas por parte de un usuario. Cada vez más, investigadores han propuesto técnicas y herramientas para realizar las pruebas de compatibilidad en los últimos años. Mientras el estado del arte crece y se diversifica, surge la necesidad de revisar y resumir sistemáticamente las soluciones presentadas. Entonces, se propone un estudio en el dominio de pruebas de compatibilidad web llevando a cabo una revisión sistemática de la literatura, orientado a encontrar las técnicas de pruebas diseñadas, las herramientas implementadas y los desafíos encontrados.

2 Metodología de revisión

La metodología elegida es la revisión sistemática de la literatura (RSL), propuesta por Kitchenham y Charters [5]. Se identificó la necesidad para la revisión y se formularon las preguntas de investigación (PI). El protocolo de revisión se observa en la Fig. 1.

El objetivo es analizar trabajos académicos relacionados al tema de compatibilidad web, haciendo énfasis en las técnicas y herramientas propuestas. Las PI servirán como guía y el análisis del conocimiento encontrado buscará responder los interrogantes propuestos. Una síntesis del estado del arte servirá para trabajos y proyectos futuros en esta línea de investigación. Para ello se formulan las siguientes PI:

- PI 1 – ¿Cuáles son las técnicas propuestas para realizar pruebas de compatibilidad web?
- PI 2 – ¿Qué herramientas se proponen para detectar defectos de incompatibilidad en aplicaciones web?
- PI 3 – ¿Qué desafíos se enfrentaron al implementar las pruebas?

Para los términos de búsqueda en los repositorios, se trabajó de forma iterativa. Las palabras clave se obtuvieron de las PI y se tuvieron en consideración sinónimos que contribuyan al resultado final. Se realizaron búsquedas sucesivas, para ajustar los parámetros usados. La expresión de búsqueda obtenida es la siguiente: (web OR website) AND ("cross-browser" OR "cross browser" OR crossbrowser) AND (test OR testing OR defect OR failure OR issue) AND (technique OR method OR tool).

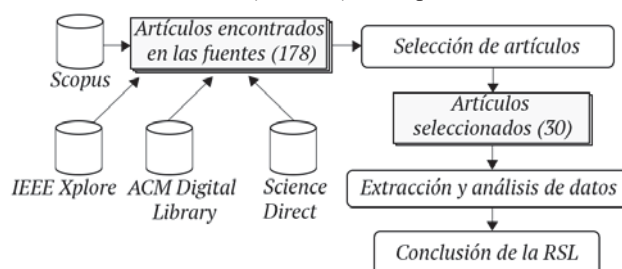


Fig. 1. Diagrama del protocolo de revisión utilizado en esta RSL.

¹ <https://crossbrowsertesting.com/>

² <https://www.browserstack.com/>

En la Fig. 1 se detallan las actividades propuestas del protocolo de revisión. Seleccionadas las fuentes se comienza la búsqueda con los términos definidos. A continuación se seleccionan los estudios mediante la aplicación de los criterios de inclusión y exclusión. Una vez obtenidos los artículos que se consideran relevantes, se extraen y analizan los datos para contestar las PI formuladas. Finalmente, se elabora una síntesis y las conclusiones de la RSL.

Para la recopilación de artículos, se seleccionaron repositorios utilizados en [6], [7], [8], [9]: Scopus, ACM Digital Library, IEEE Xplore, Science Direct.

Para descartar estudios irrelevantes se aplicaron criterios de inclusión y exclusión en cada publicación revisada. Se examinó cada estudio para decidir su inclusión. Se tomó el título, el resumen y las palabras clave como punto de partida, luego se revisó la introducción y conclusión. Los criterios usados fueron los siguientes:

- Examinar la relevancia del artículo en el área de pruebas de compatibilidad.
- Evaluar si el artículo provee información que sea de real utilidad para resolver las cuestiones propuestas con las PI.

3 Artículos seleccionados

Luego de realizar la búsqueda en los repositorios académicos, se encontraron 178 artículos. Al aplicar los criterios de inclusión y exclusión se seleccionaron 30 artículos para el análisis guiado por las PI propuestas. En la Tabla 1 se detallan los artículos.

Tabla 1. Artículos científicos seleccionados para esta RSL.

ID	Ref.	Título
S1	[10]	A Cross-browser Web Application Testing Tool
S2	[11]	Webdiff: Automated Identification of Cross-browser Issues in Web Applications
S3	[12]	Detecting cross-browser issues in web applications
S4	[13]	Automated cross-browser compatibility testing
S5	[14]	Cross Browser Incompatibility Reasons and Solutions
S6	[15]	CrossCheck: Combining crawling and differencing to better detect cross-browser incompatibilities in web applications
S7	[16]	WebMate: A Tool for Testing Web 2.0 Applications
S8	[17]	Visual testing of Graphical User Interfaces; An exploratory study towards systematic definitions and approaches
S9	[18]	Measuring and Improving Website User Experience using UX Methodologies; A Case Study on Cross Browser Compatibility Heuristic
S10	[19]	X-PERT: Accurate identification of cross-browser issues in web applications
S11	[20]	WebMate: Generating test cases for web 2.0
S12	[21]	Browserbite: Accurate cross-browser testing via machine learning over image features
S13	[22]	X-PERT: A web application testing tool for cross-browser inconsistency detection
S14	[23]	Crawl-based analysis of web applications; Prospects and challenges
S15	[24]	Cross Browser Testing Using Automated Test Tools
S16	[25]	Modeling web application for cross-browser compatibility testing
S17	[26]	Finding HTML presentation failures using image comparison techniques
S18	[27]	Adaptive random testing for image comparison in regression web testing

Tabla 1. (Continuación)

ID	Ref.	Título
S19	[28]	A crowdsourcing framework for detecting cross-browser issues in web Application
S20	[29]	An oracle based on image comparison for regression testing of web applications
S21	[30]	Detection and Localization of HTML Presentation Failures Using Computer Vision-Based Techniques
S22	[31]	Browserbite: cross-browser testing via image processing
S23	[32]	Cross-Browser Compatibility Testing Based on Model Comparison
S24	[33]	Static Analysis Technique of Cross-Browser Compatibility Detecting
S25	[34]	A Survey on Cross Browser Inconsistencies in Web Application
S26	[35]	X-Check A Novel Cross-browser Testing Service based on Record/Replay
S27	[36]	Using Visual Symptoms for Debugging Presentation Failures in Web Applications
S28	[37]	An Automated Approach for Cross-Browser Inconsistency (XBI) Detection
S29	[38]	Detect cross-browser issues for javascript-based web applications based on record-replay
S30	[39]	Detection of Cross Browser Inconsistency by Comparing Extracted Attributes

4 Resultados

4.1 PI 1 – ¿Cuáles son las técnicas propuestas para realizar pruebas de compatibilidad web?

En total 28 artículos presentaron técnicas para realizar pruebas (93.3%), algunos de ellos proponen una combinación de éstas. A continuación se resumen las técnicas:

Análisis de modelos DOM. Document Object Model (DOM) es una interfaz multi-plataforma e independiente del lenguaje para representar documentos HTML, XHTML o XML como estructuras de árboles [40]. Cada nodo del árbol es un objeto de la aplicación. La técnica consiste en comparar los modelos DOM de una página web en distintas configuraciones. Se define una configuración como una combinación navegador web-sistema operativo [21].

Análisis visual. Se comparan imágenes que son obtenidas mediante capturas de pantalla de la aplicación. En general, se comparan de a pares. Una imagen se considera una representación correcta de la página web, la segunda captura será tomada en una configuración distinta en la que se quiere probar la compatibilidad.

Análisis de modelos de navegación. Primero se produce un modelo de comportamiento de la aplicación mediante rastreadores web (Web Crawlers), éste es expresado como máquinas de estados finitos. Estos programas realizan una exploración de los posibles estados en que pueda encontrarse la página web. Luego se analizan las diferencias entre los modelos producidos.

Grabación/reproducción. Se ejecuta en etapas, en primer lugar un usuario realiza un conjunto de acciones en la aplicación, que son grabadas para luego ser reproducidas automáticamente en una configuración diferente. Luego se compara los resultados obtenidos para encontrar diferencias que provoquen incompatibilidades.

Análisis estático. Se realiza un análisis directo del código de la aplicación en lugar

de evaluar la página mientras esta es procesada por un navegador. Se centra en detectar posibles elementos en conflicto, muy ligados al estándar HTML5. El análisis de código se maneja mediante detección de expresiones regulares.

Comparación de atributos. Comienza generando gráficos que contienen atributos propios de los elementos de una página web en diferentes configuraciones mediante rastreadores web. Luego compara los atributos de los mismos elementos en los distintos gráficos para detectar incompatibilidades.

Evaluación heurística. Es un método de inspección para realizar una evaluación de usabilidad en aplicaciones, enfocado en detectar problemas en las interfaces de usuario (UI). Se basa en evaluadores que examinan una interfaz emitiendo un juicio de acuerdo a si se cumplen ciertos requerimientos establecidos.

En la Tabla 2 se resumen las técnicas propuestas y la cantidad de artículos por cada una. Las técnicas más elegidas fueron el análisis visual, análisis de modelos DOM y análisis de modelo de navegación.

Tabla 2. Resumen de técnicas propuestas por cada artículo seleccionado.

Técnica	Artículos científicos	Nº art.
Análisis de modelos DOM	(S1, S2, S6, S10, S13, S3, S4, S7, S14, S19, S26)	11
Análisis visual	(S1, S2, S6, S10, S13, S3, S8, S12, S17, S21, S18, S20, S22, S26, S27)	15
Análisis de modelo de navegación	(S4, S7, S11, S10, S13, S14, S16, S23)	8
Grabación/reproducción	(S19, S26, S29)	3
Análisis estático	(S24)	1
Comparación de atributos	(S28, S30)	2
Evaluación heurística	(S9)	1

En la Fig. 2 se observa la tendencia de uso de las técnicas en 2010-2017. El análisis de modelos DOM fue la más usada en los primeros años. Se produjo un incremento de popularidad del análisis de modelos de navegación, aunque tuvo una disminución en los últimos dos años. El análisis visual es la técnica más elegida, además su uso se incrementó a lo largo del tiempo. El análisis estático y la comparación de atributos son las técnicas más nuevas.

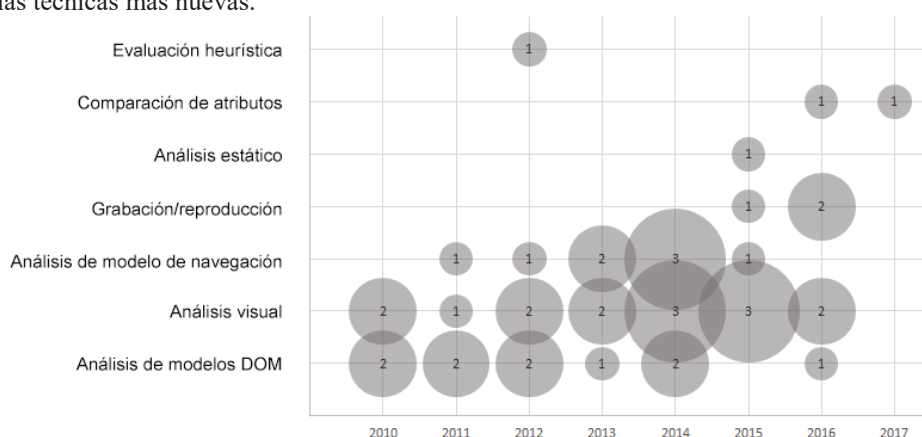


Fig. 2. Cantidad de artículos que seleccionaron cada técnica de pruebas por año de publicación.

4.2 PI 2 – ¿Qué herramientas se proponen para detectar defectos de incompatibilidad en aplicaciones web?

Se analizaron dos tipos de artículos: los que proponen una herramienta de desarrollo propio de los autores (16 artículos, 53.3%), y los que analizan herramientas comerciales (3 estudios, 10%). Un artículo (3.3%) propone una técnica de pruebas utilizando una herramienta ya desarrollada (S23).

En la Tabla 3 se listan las herramientas desarrolladas por los autores de los artículos, que además proponen alguna técnica de pruebas de compatibilidad web.

Tabla 3. Herramientas propuestas por los autores para implementar las técnicas elegidas.

Herramienta	Ref.	Técnicas de prueba
Webdiff	(S1, S2, S3)	Análisis de modelos DOM, análisis visual
CrossT	(S4)	Análisis de modelos DOM, análisis de modelo de navegación
CrossCheck	(S6)	Análisis de modelos DOM, análisis visual
WebMate	(S7, S11)	Análisis de modelos DOM, análisis de modelo de navegación
X-PERT	(S10, S14)	Análisis de modelos DOM, análisis visual, análisis de modelo de navegación
Browserbite	(S12, S22)	Análisis visual
WebSee	(S21)	Análisis visual
Crowdcheck	(S19)	Grabación/Reproducción, análisis de modelos DOM
Crawljax	(S23)	Análisis de modelo de navegación
X-Check	(S26, S29)	Grabación/reproducción, análisis de modelos DOM, análisis visual
FieryEye	(S27)	Análisis visual

En la Tabla 4 se detallan las herramientas comerciales propuestas. Algunas ya no tienen mantenimiento y otras solo funcionan en ciertas configuraciones.

Tabla 4. Herramientas comerciales para pruebas de compatibilidad web.

Herramienta	Observaciones	Ref.
Adobe Browserlab	Obsoleto desde el 13 de Marzo, 2013.	(S5)
IE Netrenderer	Sólo para Internet Explorer. Gratuita.	
Browsera	Realiza pruebas de sitios completos.	
Litmusapp	Obsoleto desde el 28 de Febrero, 2017.	
Browsrcamp	Obsoleto.	
IETester	Sólo para Internet Explorer.	(S15)
SuperPreview	De Expression Web. Ya no desarrollada.	
BrowserStack		
BrowserShots	Herramienta gratuita.	
CrossBrowserTesting		
Browser Sandbox	Herramienta gratuita.	(S25)
IE Tab	Extensión para Firefox y Chrome.	
BrowserCam	Obsoleto.	
Browserseal	Obsoleto.	

4.3 PI 3 – ¿Qué desafíos se presentan al implementar las pruebas?

En la Fig. 3 se aprecia que la detección de elementos variables fue uno de los desafíos

más mencionados. Se refiere a regiones en las aplicaciones web que no permanecen estáticas, incluyen animaciones, estadísticas y principalmente publicidades que pueden cambiar con sucesivas cargas de la página. En (S1, S2, S20) se maneja esta problemática con una estrategia de detección de regiones variables. Se carga la misma página múltiples veces en una misma configuración y se verifica qué sectores presentan cambios, para descartarlos de las pruebas.

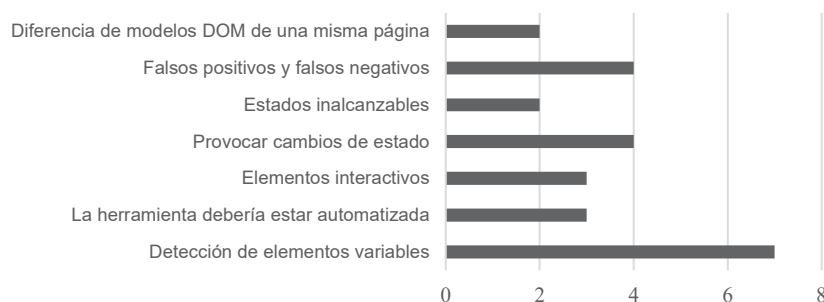


Fig. 3. Desafíos más enfrentados según los artículos científicos seleccionados.

Falsos positivos se refieren a pruebas que resultan positivas cuando no deberían, ya que existe compatibilidad web. Falsos negativos son pruebas que resultan negativas aun con la existencia de incompatibilidades web. Este problema se da en todas las pruebas al utilizar cualquiera de las técnicas mencionadas. Según (S6), información obtenida de modelos DOM puede conducir a falsos positivos, por lo que se emplea ésta técnica en conjunto con el análisis visual. En (S12, S22) se detalla un módulo de clasificación de potenciales incompatibilidades en capturas de pantalla. Se clasifican en verdaderos positivos y falsos positivos. Para utilizar este criterio de clasificación se emplea una técnica de aprendizaje automático con redes neuronales. El aprendizaje automático es una rama de la inteligencia artificial que permite a las computadoras aprender comportamiento a partir de datos empíricos [15]. También se usan técnicas de aprendizaje automático en (S6), para construir un detector de diferencias visuales más preciso. En (S29) se señala que la técnica de análisis de modelo de navegación es propensa a generar falsos positivos y negativos. Al rastrear la aplicación con los web crawlers, se ignoran fuentes de acciones no determinísticas, es decir, sólo se consideran actividades explícitas cargadas por los usuarios.

Provocar cambios de estado está relacionado al análisis de modelo de navegación. Los artículos (S7, S14) mencionan la dificultad que resulta de la necesidad de pasar de estados al elaborar el modelo de navegación. Particularmente cuando los eventos son muy numerosos, donde cualquier clic puede desencadenar un estado nuevo. En (S16) se explica que acciones ejecutadas en distinto orden producen diferentes resultados.

Elementos interactivos son una problemática recurrente en los estudios elegidos (S7, S16). Tecnologías web 2.0 promueven comportamiento dinámico con código que se ejecuta del lado del cliente, como JavaScript y HTML5 [41].

La necesidad de que la herramienta para las pruebas de compatibilidad web sea automatizada se hace evidente en varios estudios (S6, S9, S29). Se menciona lo costoso que resulta tener un usuario que realice acciones de forma manual (S6). Aun así, (S8, S9) proponen técnicas totalmente manuales para las pruebas.

Estados inalcanzables es una problemática de los análisis de modelo de navegación. Según (S14, S29) existen estados que no pueden alcanzarse desde enlaces en la página. Entonces, el modelo producido estará incompleto, dejando parte del sistema fuera de las pruebas.

5 Conclusiones y trabajos futuros

Actualmente la plataforma elegida para la comunicación y explotación de las aplicaciones es la web. Debido a la naturaleza de las aplicaciones, asíncronas y dinámicas, realizar pruebas efectivas se vuelve un desafío para los desarrolladores. La rápida evolución y popularidad de tecnologías han fomentado la producción de aplicaciones complejas y dinámicas [41], [43]. Esto hace que la usabilidad, la compatibilidad y la disponibilidad sean los criterios clave para el éxito de un sitio web [1]. Pero distintos navegadores producen contenido diferente [44], [45], [46]. De aquí surge la necesidad de realizar pruebas de compatibilidad.

Se estudiaron las técnicas con las cuales se diseñan pruebas de compatibilidad. El análisis visual resultó la técnica más elegida, quedando el análisis de modelos DOM y el análisis de modelos de navegación en segundo y tercer lugar, respectivamente. Asimismo, se revisaron las aplicaciones propuestas para desarrollar las pruebas. Se las clasificó según el origen: desarrolladas por los autores de los estudios (11, 44%) o aplicativos comerciales que el artículo evaluó (14, 56%).

Finalmente se listaron sistemáticamente los desafíos encontrados en los artículos revisados. Una de las dificultades más mencionadas fue la presencia de elementos variables, abundantes en las aplicaciones web actuales [43]. Otra de las problemáticas encontradas fue la necesidad de automatizar las pruebas. Si bien se proponen herramientas y técnicas con distintos niveles de automatización (algunas soluciones son totalmente manuales), se observa el alto costo asociado al requerir la participación constante de un usuario. Dependiendo de la popularidad de un sitio web, el número de configuraciones que se deben probar para cubrir el 90-95% de usuarios asciende a 20-30 [31]. Por esto es necesario automatizar el proceso de pruebas.

Como trabajo futuro se propone diseñar y desarrollar una herramienta para pruebas de compatibilidad web, intentando resolver los desafíos encontrados en la RSL.

Referencias

1. Dustin, E., Rashka, J., McDiarmid, D.: *Quality Web Systems: Performance, Security, and Usability*. Addison Wesley, Boston, Massachusetts, USA (2001)
2. Moustakis, V., Litos, C., Dalivigas, A., Tsironis, L.: *Website Quality Assessment Criteria*. IQ, 59-73 (2004)
3. Kurose, J., Ross, K.: *Redes de computadoras. Un enfoque descendente*. Pearson (2010)
4. Mozilla Corporation In: *Mozilla Developer Network*. (Accessed 2017) Available at: https://developer.mozilla.org/en-US/docs/Learn/Tools_and_testing/Cross_browser_testing
5. Kitchenham, B., Charters, S.: *Guidelines for performing Systematic Literature Reviews in Software Engineering*. Keele University, Durham University (2007)
6. Zakaria, Z., Atan, R., Ghani, A., Sani, N.: *Unit Testing Approaches for BPEL: A Systematic*

- Review. In : Software Engineering Conference. Asia-Pacific, Malaysia (2009)
7. Nabil, E. I.: Specifications for Web Services Testing: A Systematic Review. In : 2015 IEEE World Congress on Services, New York, New York, USA (2015)
 8. Mendes, E.: A systematic review of Web engineering research. In : International Symposium on Empirical Software Engineering, Noosa Heads, Queensland, Australia (2005)
 9. Dogan, S., Betin-Can, A., Garousi, V.: Web application testing: A systematic literature review. *Journal of Systems and Software* 91, 174-201 (2014)
 10. Choudhary, S., Varsee, H., Orso, A.: A cross-browser web application testing tool. In : 2010 IEEE International Conference on Software Maintenance, Timișoara, Romania (2010)
 11. Choudhary, S., Varsee, H., Orso, A.: WEBDIFF: Automated Identification of Cross-browser Issues in Web Applications. In : 26th IEEE International Conference on Software Maintenance, Timișoara, Romania (2010)
 12. Choudhary, S.: Detecting Cross-browser Issues in Web Applications. In : 2011 33rd International Conference on Software Engineering (ICSE), Honolulu, Hawaii, USA (2011)
 13. Mesbah, A., Prasad, M.: Automated Cross-Browser Compatibility Testing. In : 2011 33rd International Conference on Software Engineering, Honolulu, Hawaii, USA (2011)
 14. Ochin, J.: Cross Browser Incompatibility: Reasons and Solutions. *International Journal of Software Engineering & Applications*, July 2011 Vol.2(No.3) (2011)
 15. Choudhary, S., Prasad, M., Orso, A.: CROSSCHECK: Combining Crawling and Differencing To Better Detect Cross-browser Incompatibilities in Web Applications. In : IEEE Fifth International Conference on Software Testing, Verification and Validation, Canada (2012)
 16. Dallmeier, V., Burger, M., Orth, T., Zeller, A.: WebMate: A Tool for Testing Web 2.0 Applications. In : Workshop on JavaScript Tools, Beijing, China (2012)
 17. Issa, A., Sillito, J., Garousi, V.: Visual Testing of Graphical User Interfaces: an Exploratory Study Towards Systematic Definitions and Approaches. In : 2012 14th IEEE International Symposium on Web Systems Evolution, Trento, Italy (2012)
 18. Sivaji, A., Ramli, N., Nor, Z., Chuan, N.-K., Wan, F., Wan, A., Shi-Tzuaan, S.: Measuring and Improving Website User Experience using UX Methodologies: A Case Study on Cross Browser Compatibility Heuristic. In : Southeast Asian Network of Ergonomics Societies, Langkawi, Kedah, Malaysia (2012)
 19. Choudhary, S., Prasad, M., Orso, A.: X-PERT: Accurate identification of cross-browser issues in web applications. In : 2013 35th International Conference on Software Engineering, San Francisco, California, USA (2013)
 20. Dallmeier, V., Burger, M., Orth, T., Zeller, A.: WebMate: Generating Test Cases for Web 2.0. In : International Conference on Software Quality, Software Quality. Increasing Value in Software and Systems Development, Vienna, Austria (2013)
 21. Semenenko, N., Dumas, M., Saar, T.: Browserbite: Accurate Cross-Browser Testing via Machine Learning over Image Features. In : 2013 29th IEEE International Conference on Software Maintenance, Eindhoven, Netherlands, Netherlands (2013)
 22. Choudhary, S., Prasad, M., Orso, A.: X-PERT: a web application testing tool for cross-browser inconsistency detection. In : 2014 International Symposium on Software Testing and Analysis, San Jose, California, USA (2014)
 23. Deursen, A. v., Mesbah, A., Nederlof, A.: Crawl-based analysis of web applications: Prospects and challenges. *Science of Computer Programming* 87, 173-180 (2014)
 24. Kaalra, B., Gowthaman, K.: Cross Browser Testing Using Automated Test Tools. *International Journal of advanced studies in Computer Science and Engineering* 3(10), 7-12 (2014)
 25. Li, X., Zeng, H.: Modeling web application for cross-browser compatibility testing. In : 2014 15th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, Las Vegas, Nevada, USA (2014)
 26. Mahajan, S., Halfond, W.: Finding HTML presentation failures using image comparison techniques. In : 29th ACM/IEEE international conference on Automated software

- engineering, Vasteras, Sweden (2014)
27. Selay, E., Zhou, Z., Zou, J.: Adaptive Random Testing for Image Comparison in Regression Web Testing. In : 2014 International Conference on Digital Image Computing: Techniques and Applications, Wollongong, New South Wales, Australia (2014)
 28. He, M., Tang, H., Wu, G., Zhong, H.: A Crowdsourcing framework for Detecting Cross-Browser Issues in Web Application. In : the 7th Asia-Pacific Symposium on Internetware, Wuhan, China (2015)
 29. Hori, A., Takada, S., Tanno, H., Oinuma, M.: An oracle based on image comparison for regression testing of web applications. In : 27th International Conference on Software Engineering and Knowledge Engineering, Pittsburgh, USA (2015)
 30. Mahajan, S., Halfond, W.: Detection and Localization of HTML Presentation Failures Using Computer Vision-Based Techniques. In : 2015 IEEE 8th International Conference on Software Testing, Verification and Validation, Graz, Austria (2015)
 31. Saar, T., Dumas, M., Kaljuve, M., Semenenko, N.: Browserbite: cross-browser testing via image processing. *Software—Practice & Experience* 46(11), 1459-1477 (2015)
 32. Shi, H., Zeng, H.: Cross-Browser Compatibility Testing Based on Model Comparison. In : 2015 International Conference on Computer Application Technologies, Matsue, Japan (2015)
 33. Xu, S., Zeng, H.: Static Analysis Technique of Cross-Browser Compatibility Detecting. In : 2015 3rd International Conference on Applied Computing and Information Technology/2nd International Conference on Computational Science and Intelligence, Okayama, Japan (2015)
 34. Barskar, N., Patidar, C. P.: A Survey on Cross Browser Inconsistencies in Web Application. *International Journal of Computer Applications* 137(4), 37-41 (2016)
 35. He, M., Wu, G., Tang, H., Chen, W., Wei, J., Zhong, H., Huang, T.: X-Check: A Novel Cross-browser Testing Service based on Record/Replay. In : 2016 IEEE International Conference on Web Services, San Francisco, California, USA (2016)
 36. Mahajan, S., Li, B., Behnamghader, P., Halfond, W.: Using Visual Symptoms for Debugging Presentation Failures in Web Applications. In : 2016 IEEE International Conference on Software Testing, Verification and Validation, Chicago, Illinois, USA (2016)
 37. Sharma, M., Patidar, C.: An Automated Approach for Cross-Browser Inconsistency (XBI) Detection. In : 9th Annual ACM India Conference, Gandhinagar, India (2016)
 38. Wu, G., He, M., Tang, H., Wei, J.: Detect Cross-Browser Issues for JavaScript-Based Web Applications Based on Record/Replay. In : 2016 IEEE International Conference on Software Maintenance and Evolution, Raleigh, North Carolina, USA (2016)
 39. Patidar, C. P., Sharma, M., Sharda, V.: Detection of Cross Browser Inconsistency by Comparing Extracted Attributes. *International Journal of Scientific Research in Computer Science and Engineering* 5(1), 1-6 (2017)
 40. W3C Document Object Model. (Accessed 2005) Available at: <https://www.w3.org/DOM/>
 41. O'Reilly, T.: What Is Web 2.0 - Design Patterns and Business Models for the Next Generation of Software. Available at: <http://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html>
 42. Powers, D.: Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation. *Journal of Machine Learning Technologies* 2(1), 37-63 (2011)
 43. Jazayeri, M.: Some Trends in Web Application Development. In : Future of Software Engineering, 2007. FOSE '07, Minneapolis, Minnesota, USA (2007)
 44. Kiciman, E., Livshits, B.: AjaxScope: a platform for remotely monitoring the client-side behavior of web 2.0 applications. In : Proceedings of twenty-first ACM SIGOPS symposium on Operating systems principles, Stevenson, Washington, USA (2007)
 45. Ricca, F., Tonella, P.: Web testing: a roadmap for the empirical research. In : Seventh IEEE International Symposium on Web Site Evolution, Budapest, Hungary (2005)
 46. Ramler, R., Weippl, E., Winterer, M., Schwinger, W., Altmann, J.: A Quality-Driven Approach to Web Testing. In : Ibero American Conference on Web Engineering, Santa Fé, Argentina (2002)

Utilización de SysML y Papyrus para el desarrollo del software de vuelo del Cubesat FS2017 haciendo uso del enfoque MBSE

Emmanuel Arias, Eduardo Sufan, Javier Uranga, and Carlos Barrientos

Universidad Nacional de La Matanza - Unidad de Formación Superior, CONAE,
Ruta Prov. C45, Km 8, Falda del Cañete, Córdoba, Argentina

{emarias, esufan, juranga,
}@alumno.unlam.edu.ar
carlos.barrientos@conae.gov.ar
<http://ufs.conae.gov.ar/>

Resumen La ingeniería en sistemas permite la realización correcta de sistemas, para ello es importante enfocarse en los requerimientos que provienen de los stakeholders y en las funcionalidades, la documentación, el diseño, la verificación, la validación y el test del mismo, siempre estudiando al sistema de manera completa. Actualmente, existe un gran variedad de metodologías y estándares que ayudan en el proceso de ingeniería de sistemas. Este trabajo se enfoca en demostrar las bondades que brinda la Ingeniería Basada en Modelos, en conjunto con el lenguaje de modelado SysML, y la herramienta *open source* Papyrus. Esta combinación logró el diseño del software de vuelo de la misión satelital argentina Formador Satelital 2017 con éxito.

Keywords: Diseño y Modelado, SysML, Papyrus, CubeSat, Software de Vuelo, Ingeniería de Sistemas.

1. Introducción

Este trabajo tiene como objetivo, demostrar la importancia tanto de la ingeniería de sistemas, como del enfoque (y diseño) basado en modelos para el desarrollo de sistemas complejos (y críticos). La ingeniería de sistemas permite la realización correcta de sistemas, esto ayuda a llevar a cabo un sistema operable capaz de cumplir con los requerimientos teniendo en cuenta las restricciones, y además lograr hacer frente a los riesgos de cualquier índole, que pueden afectar al proyecto. Los modelos presentados en este trabajo son llevados a cabo utilizando SysML. Este es un lenguaje de modelado para la aplicaciones de ingeniería de sistema, y es ampliamente utilizado en proyectos de grandes magnitudes debido a su versatilidad, y su sencilla sintaxis y semántica.

1.1. Contexto

En el marco del Plan Nacional de Actividades Espaciales[1], la Comisión Nacional de Actividades Espaciales (CONAE) creó la Unidad de Formación Su-

perior (UFS), con el propósito de generar y capacitar recursos humanos. Esto permite desarrollar profesionales con un alto nivel de conocimientos tanto científicos, como tecnológicos, que se desarrollarán en el ámbito espacial. Así de este modo, se logra satisfacer las necesidades de las instituciones y empresas argentinas que trabajan en el área espacial. La UFS, en conjunto con tres Universidades Nacionales, entre ellas la Universidad Nacional de Córdoba, Universidad Tecnológica Nacional (Facultad Regional Córdoba y Facultad Regional Mendoza) y la Universidad Nacional de La Matanza, han implementado cuatro nuevas maestrías: Maestría en Aplicaciones de Información Espacial (MAIE), Maestría en Tecnología Espacial (MTS), Maestría en Instrumentos Satelitales (MIS) y Maestría en Desarrollo Informático de Aplicación Espacial (MDIAE), para cubrir diferentes áreas del desarrollo de una misión satelital.

En este contexto, las tres maestrías, desarrollan en conjunto, una misión satelital llamada Formador Satelital 2017 (FS2017), con el principal objetivo de proveer a los maestrandos entrenamiento práctico en los aspectos técnicos y programáticos de una misión espacial.

Se ha decidido desarrollar una misión Cubesat debido a las ventajas que presentaban. Esta misión tiene dos cargas útiles: Una cámara NIR/SWIR perteneciente a la misión argentina SABIA-MAR, la cual se buscará medir la calidad del detector en el espacio; y un Data Collection System que permitirá monitorear una gran cantidad de estaciones, y proveer servicios a instituciones que necesitan obtener información de zonas aisladas del país.

2. Ingeniería en Sistemas

Definir exactamente qué es la ingeniería en sistemas (SE) es una tarea compleja [4]. La SE es un enfoque interdisciplinario, metódico para el diseño, la realización, gestión técnica, operación y el retiro de sistemas [2]. El proceso de SE tiene una metodología iterativa que soporta el descubrimiento, el aprendizaje y la mejora continua.

La SE permite la realización correcta de sistemas. Para poder lograr ese objetivo se debe enfocar en las necesidades de las partes interesadas y en las funcionalidades, ciclo de desarrollo, documentación, requerimientos, diseño, verificación y validación del sistema, siempre teniendo en cuenta el problema completo, esto incluye además operaciones, costos, cronogramas, performance, entrenamiento y soporte, test, fabricación y retiro (o disposal) del sistema [3].

Se afirma además que la SE es el arte y la ciencia de desarrollar un sistema operable capaz de cumplir con los requerimientos teniendo en cuenta las restricciones (constraints) impuestos en los requerimientos [2].

Una de las importancias de llevar a cabo la SE es que permite reducir el riesgo que es inherente a todo sistema. Este riesgo puede ser de cualquier índole: riesgos técnicos, riesgos financieros, riesgos ambientales, etc.

Otro beneficio de utilizar la SE es que permite mejorar la performance de los proyectos llevados a cabo. Esto se demuestra en un estudio realizado por la National Defense Industrial Association, el Institute of Electrical and Electronic

Engineers (IEEE) y el Carnegie Mellon. Este trabajo demostró que en proyectos donde hay un bajo nivel de SE, solo el 15 % de estos proyectos tiene un alto nivel de performance. En cambio, aquellos que aplican en mayor medida la SE, la cantidad de proyecto con una alta performance crece hasta los 57 % [5].

La SE es utilizada en proyectos satelitales debido a que brinda importantes beneficios en la gestión y en el desarrollo de las misiones espaciales. La National Aeronautics and Space Administration (NASA) define un ciclo de vida de los programas/proyectos, divididos en fases, los cuales se encuentran separados por Key Decision Points (KDPs) los cuales son reuniones con expertos y autoridades que deciden si el programa/proyecto es lo suficientemente maduro para pasar a la siguiente fase del proyecto (decisiones GO o NO-GO) [2]. Las fases de un proyecto siguiendo el lineamiento de NASA, son las siguientes:

- Pre-fase A: se llevan a cabo estudios de conceptos, se estudian las diferentes alternativas para la misión satelital.
- Fase A: se determina la viabilidad del sistema propuesto y se desarrollan los requerimientos a nivel de sistemas y el concepto de misión.
- Fase B: se lleva a cabo un diseño preliminar. Se desarrollan los requerimientos de la estructura final del producto y se genera un diseño preliminar del sistema.
- Fase C: se desarrolla el diseño final y la fabricación del producto. Se completa el diseño detallado del sistema completo y se comienza a codificar el software.
- Fase D: se ensamblan e integran las partes para crear el sistema final.
- Fase E: comienza la operación y mantenimiento del sistema.
- Fase F: se debe implementar el retiro del sistema.

Esto se basa en el ciclo de vida propuesto por ISO/IEC [6] el cual consiste en: concepto, desarrollo, producción, utilización, soporte y retiro.

La SE identifica cuatro grupo de procesos: procesos técnicos, procesos técnicos de gestión, proceso de acuerdo, procesos de organización de proyectos. En este trabajo nos basamos en los procesos técnicos. Esto incluye, entre otras cosas, la definición de requerimientos, definición de la arquitectura y definición del diseño. Los procesos técnicos permiten transformar requerimientos en productos que los cumplan [3].

3. Ingeniería en Sistemas Basada en Modelos

La SE utiliza una gran cantidad de modelos como pueden ser modelos funcionales para el desarrollo de requerimientos, simulación de modelos para analizar el comportamiento del sistema y otros tipos de modelos para analizar diferentes aspectos del mismo, como la confiabilidad, seguridad, propiedades de masa, etc. En la actualidad, se sigue utilizando una metodología orientada a documentos, donde toda la información referente al sistema o proyecto, tal como los requerimientos, arquitectura, diseño, se almacenan en documentos. Esto dificulta el mantenimiento y la reutilización de información [7].

Los diseñadores de satélites, misiones espaciales, stakeholders, crean modelos y artefactos del mismo sistema utilizando diferentes procesos, herramientas y representaciones. La descoordinación crea importantes barreras de comunicación entre los diseñadores y los stakeholders. La misma información que es capturada múltiples veces, en múltiples lugares, con múltiples representaciones crean problemas de mantenimiento [11].

Para hacer frente a estos problemas, se ha comenzado a utilizar modelos para lograr comunicar y documentar los proyectos. Esto permite conocer con anticipación las necesidades y el comportamiento del sistema antes que el proceso de desarrollo se lleve a cabo. Además, permite una mejor distribución de las tareas entre el equipo de trabajo, maximizando la productividad y minimizando los errores [3].

Existen diferentes tipos de modelos, cada uno se enfoca en diferentes aspectos del sistema. Según [3] los modelos se pueden clasificar como:

- Maquetas físicas, que es una representación del sistema actual.
- Modelos abstractos, pueden utilizar diferentes maneras para representar el sistema. Estos pueden ser modelos formales o informales.
- Modelos informales, estos pueden realizarse con alguna herramienta de dibujo. Estos modelos son menos claros que los modelos formales, y pueden presentar ambigüedades.
- Modelos formales, estos pueden ser clasificados como: *modelos geométricos* para representar la geometría o la relación espacial del sistema; *modelos cuantitativos*, para representar relaciones cuantitativas del sistema; y *modelos lógicos*, que se utilizan para representar relaciones lógicas del sistema, como las relaciones entre las partes, relaciones entre las actividades, etc.

Esto es lo que forma la base de la **metodología de la SE basada en modelos** (MBSE). Esta se define como “la aplicación formal del modelado para brindar soporte a los requerimientos del sistema, y a las actividades de diseño, análisis, verificación y validación, que comienzan en la fase de diseño conceptual y continúan durante el desarrollo, hasta el final del ciclo de vida” [3]. MBSE brinda grandes ventajas en contraste con el enfoque “tradicional” orientado a documentos, donde es necesario almacenar grandes volúmenes de información referente al sistema en documentos. Esto dificulta el mantenimiento y la sincronización, afectando en gran medida la calidad del producto.

Existen varias metodologías MBSE que se pueden aplicar tales como: *IBM Telelogic Harmony-SE*, el cual implementa una V clásica como ciclo de vida y asume que los modelos y requerimientos se encuentran centralizados en un repositorio; *INCOSE Object-Oriented Systems Engineering Method (OOSEM)* el cual utiliza un enfoque top-down basado en modelos que utiliza SysML [8] para la especificación, análisis, diseño y la verificación del sistema; *IBM Rational Unified Process for Systems Engineering (RUP SE)* este es un proceso que se basa en el RUP tradicional para el desarrollo de proyectos de software; *JPL State Analysis (SA)* que aprovecha el control de arquitecturas basado en modelos y estados [10].

3.1. Estándar para el modelado

Existe una variedad estándares que se pueden utilizar para apoyar el análisis, especificación, diseño, verificación y validación del sistema. En [7] se brinda una lista de estándares de modelado, que suelen ser utilizados en el desarrollo de sistemas. En este trabajo utilizamos SysML.

SysML. Este es un lenguaje de propósito general, creado por el OMG (Object Management Group) para el modelado gráfico de arquitecturas para aplicaciones de SE. SysML soporta la especificaciones, análisis, diseño, verificación y validación de sistemas (y sistemas de sistemas) [8]. SysML, que actualmente se encuentra en su versión 1.4, unifica los diversos lenguajes de modelado que son utilizados por los ingenieros en sistemas, de manera similar a UML en la industria del software, además utiliza una subconjunto de UML 2. SysML es producto de una iniciativa en conjunto del OMG y el INCOSE [4]. Este lenguaje especifica nueve tipo de diagramas y un tipo de notación tabular que es denominado *tabla de asignación*. De estos diagramas, cuatro son considerados como *diagramas estructurales o estáticos* (definición de bloques, bloques internos, diagramas paramétricos y diagramas de paquetes) y otros cuatro se los conocen como *diagramas de comportamiento o dinámicos* (actividad, secuencia, máquinas de estados y casos de uso), y el diagrama restante es utilizado para especificar los *requerimientos* [8]. Existen muchos casos de éxito de la utilización del SysML [11] [12] [13] [14].

4. Herramienta

Es importante seleccionar la correcta herramienta que se va a utilizar en todo el proyecto. La misión Cubesat FS2017 exigía que la herramienta a utilizar sea estable, y que permita el trabajo colaborativo de todo el equipo. Existen varias alternativas en la actualidad [8]. Luego de un estudio preliminar que hemos realizado se decidió continuar con el software Papyrus.

Papyrus. Esta es una herramienta *open source* de MBSE y pertenece a la Fundación Eclipse. Existen diferentes casos de éxito de la utilización de Papyrus, incluyendo el desarrollo del software de vuelo del satélite TARANIS, perteneciente a el Centro Nacional de Estudios Espaciales (CNES) de Francia o en la empresa aeronáutica Airbus. Esta herramientas fue la utilizada en el proyecto FS2017 [9].

5. Diseño del Software de vuelo del FS2017

5.1. Requerimientos

La misión Cubesat FS2017 siguiendo la filosofía de NASA, comienza con la definición de las expectativas de los stakeholder (partes interesadas). Estos

- Command and Data Handling (C&DH)
- Attitude Determination and Control System (ADCS)
- Electrical Power Subsystem (EPS)
- Telemetry, Telecommand and Communication Subsystem (TT&C)
- Payload, o carga útil, en esta parte se encuentran todos los componentes necesarios para que la misión se cumpla:
 - Cámara NIR/SWIR (Near-Infrared/ Short-wavelength infrared).
 - Una computadora a bordo (OBC) para el manejo de los componentes del Payload.
 - DCS Tranceiver, Una antena para descargar información de un DCS (Data Collection System) que la misión lleva a bordo.

Con la utilización de SysML, esta descripción se puede comprimir a un modelo, dónde toda la información, tanto sistémica como técnica, se encuentra embebidos en el modelo. Siguiendo el MBSE se puede representar cada subsistema como cajas, con información importante, mostrando la manera en que se relacionan Figura 2. Las herramientas de modelado permiten almacenar información relevante dentro de estos modelos, los cuales estos se pueden mostrar u ocultar dependiendo de la cantidad de información que se quiere mostrar.

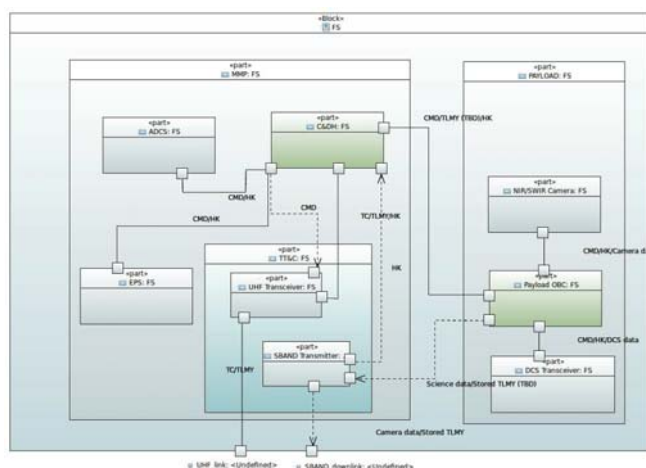


Figura 2. Arquitectura general de la misión FS2017

Desarrollar visualmente la arquitectura que se espera como resultado, nos permitió conocer cuáles son las interfaces esperables para cada uno de los módulos (o subsistemas). De esta manera, los diferentes equipos que trabajan sobre los subsistemas, lo pueden hacer libremente, ya que las interfaces o el flujo entre los subsistemas ya se encuentran preestablecidos en el modelo. Así, todos los modelos que se desarrollan tendrán una compatibilidad desde el primer momento,

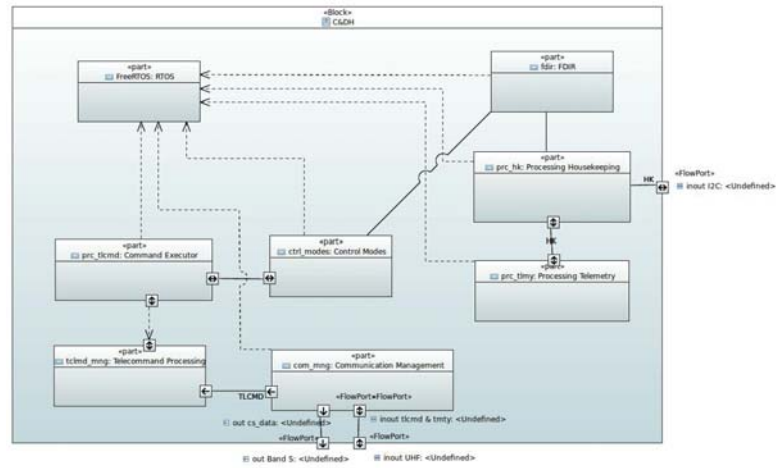


Figura 4. Diagrama de bloques internos del subsistema Command Data Handling

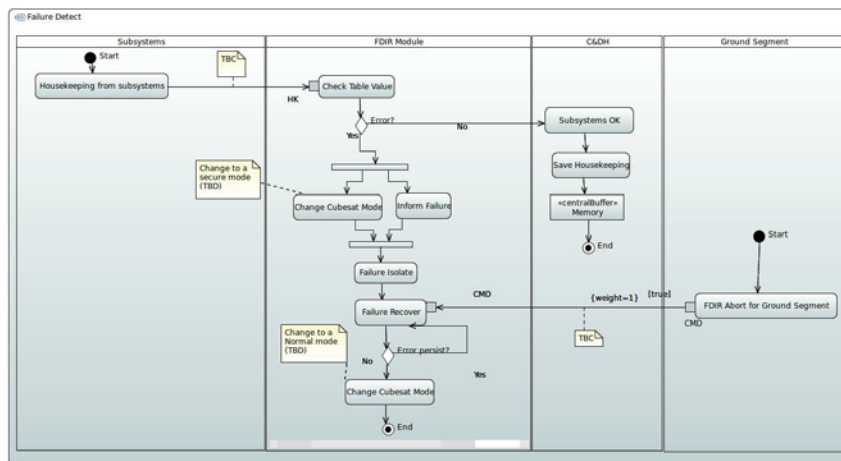


Figura 5. Diagrama de actividades de la detección de fallas perteneciente al módulo de FDIR del subsistema Command and Data Handling

aquellos documentos que son vitales para la continuidad de la misión, como así también utilizar los diagramas como documentación.

La herramienta de modelado Papyrus nos permitió aplicar correctamente el MBSE y llevar a cabo la trazabilidad de los requerimientos del proyecto. Papyrus, fue utilizado en múltiples oportunidades en el desarrollo de importantes proyectos a nivel mundial [9]. Su principal ventaja es ser *open source*, lo cual nos permite modificar a nuestra conveniencia los modelos, y/o utilizar modificaciones de estándares propios del equipo de desarrollo.

Referencias

1. CONAE: Plan Espacial Nacional Argentina en el espacio 2004-2015. Argentina (2010)
2. NASA: Systems Engineering Handbook. NASA/SP-2007-6105 Rev1. Washington, D.C. (2007)
3. Walden, D.D., Roedler, G.J., Forsberg, K.J., Hamelin, R.D., Shortell, T.M.: Systems Engineering Handbook. A guide for system life cycle processes and activities (4th ed.) International Council on System Engineering. John Wiley & Sons, Inc. San Diego, CA (2015)
4. Holt J., Perry, S.: SysML for Systems Engineering. The Institution of Engineering and Technology. 4ed. Wiley. United Kingdom (2008).
5. Elm J.P., Goldenson, D.R.: The Business Case for Systems Engineering Study: Results of the Systems Engineering Effectiveness Survey. CERT Program. Carnegie Mellon University. Pittsburgh (2012).
6. ISO/IEC TR: Systems and software engineering - Life cycle management. Technical report, Switzerland (2010)
7. BKCASE: Guide to the Systems Engineering Body of Knowledge. BKCASE Editorial Board (2017).
8. SysML, <http://sysml.org/>
9. Papyrus, <https://www.eclipse.org/papyrus/>
10. Estefan J.: Survey of Model-Based Systems Engineering (MBSE) Methodologies, Rev. B, Section 3.2. NASA Jet Propulsion Laboratory (2008)
11. Wang, L.: HESTIA Model Based System Engineering with SysML. NASA Johnson Space Center; Houston, TX, United States (2016)
12. Breckenridge, J.T., Johnson, S.B.: Implementation of a Goal-Based Systems Engineering Process Using the Systems Modeling Language (SysML). Boston, MA; United States (2013)
13. Mandutianu, S., Morillo, R., Simpson K., Liepack O., Bonanne K.: Modeling Complex Cross-Systems Software Interfaces using SysML. Jet Propulsion Laboratory/Caltech Boston, MA; United States (2013)
14. Dutenhoffer, C., Tirona, J.: The Value of SysML Modeling During System Operations: A Case Study. Jet Propulsion Laboratory (2013)
15. Roggero, E., Alberto, N., Cara, L., Gonzalez, E., Vanni, M.: Formador Satelital 2017, An opportunity to train student in sciences and space technologies. Unidad Formación Superior, CONAE. Argentina (2013)

Wide Evaluation Processes

Leonardo Anibal Fernández¹, Rodolfo Bertone²

¹ leonardoanibalfernandez@gmail.com, ² pbertone@lidi.info.unlp.edu.ar
Facultad de Informatica - Universidad Nacional de La Plata
Calle 50 s/n, 1900 La Plata, Buenos Aires, Argentina

Abstract: The evaluation process in software quality is very important because it helps to identify strengths, weakness and to provide information that can help to improve the developer's efforts on application development. In this paper, we describe the functionalities of the web system EVAP COMPETISOFT (Wide processes evaluation based on COMPETISOFT project) that implements the evaluation methodology described on COMPETISOFT project and we introduce some results of applying it in some development enterprises.

Keywords: Quality evaluation, software quality, COMPETISOFT, SMEs.

1 Introduction

To get quality, is required the use of methodologies and standardized process for the analysis, design, development and software testing obtaining like this an uniform way to work guaranteeing more credibility and maintainability on our software and getting at the same time more productivity, not only at the development process also in the software's quality control.

Is important to know that adopting good practices in quality helps to get software quality but it does not guarantee it. Therefore, besides knowing the quality criteria is important to evaluate the fulfillment of them.

The evaluation process in software quality is very important because it helps to identify strengths, weakness and to provide information that can help to improve the developer's efforts on application development.

2 COMPETISOFT Project

COMPETISOFT is the result of different Ibero-American investigation groups. It was funded by CYTED (Ibero-american project of science and technology for development) and it has as main objective increase the competitive level of all ibero-american software developer SME (Small and Medium Enterprises) through its methodological framework that could turn into a base to get evaluation and certification process in software industry recognized in all Ibero-American[1].

This project also offers to the software industry a model based on the best international practices. The main characteristics that it has is the facility to understand it, its facility to apply it, that is not expensive on its adoption and maybe the most

important is that it a base to get a successful evaluation with other models like ISO 9000:2000 o CMM®1 V1.1.

2.1 Processes Model

The COMPETISOFT Processes Model is the one defined by MoProSoft¹. It counts with three process categories: High Direction, Management and Operation. These categories with their respective processes reflect an organization structure. The figure 1 represents the different categories with their processes and sub-processes:

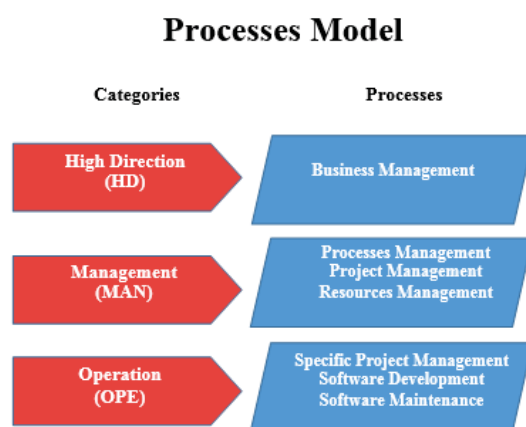


Fig. 1 Processes Model

2.2 Processes Evaluation Methodology

The main objective that tries to obtain the processes evaluation methodology (EvalProSoft)² in software industry is to deliver to the applicant organization the profile of the level capacity on the different implanted process and a level of maturity on the capacity of the organization.

The capacity level got from the evaluation result it could be a number in a scale zero to five. The zero level is the lowest capacity level the process can obtain and it means that the process does not fulfill with the process purpose. The fifth is associated with the highest capacity level and it means that the actual and projected business objectives are fulfilling through the optimization and the continued improvement.

On table 1 are described the different capacity levels.

Table 1: Capacity level description

¹ Página oficial del modelo. <http://www.moprosoft.com.mx/>

² Método de Evaluación de procesos para la industria de software (EvalProSoft) - Versión 1.1 -Marzo 2004
http://materias.utags.edu.mx/claroline/backends/download.php?url=L0FydGJdWxvcy9FdmFsUHJvU29mdHYxMS5wZGY%3D&cidReset=true&cidReq=CALS_IVET

Level	Description
Incomplete	The process was not implanted or it fails to reach the process purpose.
Completed	The process reaches its purpose.
Administered	The process "Completed" was implanted in an organized way and their work products are properly established, controlled and maintained.
Established	The process "Administered" was implanted through a defined process that is capable to obtain the process results.
Predictable	The process "Established" works inside the limits to obtain the results.
Optimized	The process "Predictable" is improvement continuingly to reach the actual business goals and the future relevant ones.

To determine the level of compliance it is used an ordinal scale which ranges are described on table 2.

Table 2: Compliance scale process attribute

Result	Description	Value range scope
N	Not Reached	Between 0 and 15%
P	Partially Reached	> 15 % until 50 %
W	Wide Reached	> 50 % until 85 %
C	Completely Reached	> 85 % until 100 %

Once that the evaluation method finish it obtains a report with the results that are send to the applicant organization. On it, is documented the profile of the capacity level on the different process, the maturity level of the capacity and a resume of the findings detected. On the statistics report is also provided general information about the evaluated organization, the results of the evaluation and also the learned lessons about the evaluation methodology.

2.3 Improvement Processes

PMCOMPETISOFT (COMPETISOFT Improvement Processes) based on Agile SPI³, is an agile improvement process that can be used as a guide on the improvement process program of small and mediums enterprises. The main objective of this process is to give a guide on the improvement software process to improve some of the following aspects: productivity increase, quality improvement, adequacy to software process standards, improvement on the customer satisfaction and improvement on the process perception inside the organization.

³Mejora de Procesos de Software Ágil con Agile – Spi Process.
<http://dyna.unalmed.edu.co/ediciones/164/articulos/a25v77n164/a25v77n164.pdf>

PMCOMPETISOFT is composed for five macro-activities: Installation, Diagnostic, Formulation, Improvement and Program Revision.

3 EVAP Competisoft

AECP (Application to capability evaluation process)[2] is a tool to make self-evaluation through the Evaluation Process Model proposes on COMPETISOFT project. The application was developed to implement the two first steps of the improvement process model that proposes COMPETISOFT: Program Installation and Diagnostic, leaving the other three (Formulation, Improvement and Revision) in charge of the organization through the app diagnostic.

But AECP counts with some failures as the following:

- It is not possible to evaluate the processes from all the categories presents on the Processes Model because only implements the self-evaluation for “Operation” category.
- It does not count with an historical of previous evaluations. The detailed result is available to see it once finished the evaluation, after that is not possible to do it.
- Questions are not flexible in the possible answers. It only accepts “YES/NO” answer but it does not accept a range of acceptance as answer.
- The security scheme to handle the passwords does not count with web security aspects.

EVAP (Wide Process Evaluation), which takes as bases the web system AECP, fixed the failures mentioned before and some others to make it the most complete and efficient as possible.

- The app was modified to allow evaluate the processes from the other two categories that present COMPETISOFT project: “High Direction” and “Management”.
- The original database was modified to allow the different organization to answer the processes they want to evaluate. In other words, the new version allows evaluating all the processes but in case that the organization doesn’t want to evaluate some of them now it is possible.
- But the most important modification made to the application was to adapt the questionnaire to allow a range of acceptance on the answers.

Originally the app was designed to accept “YES/NO” answers making the evaluation results processes faster once that the self-evaluation was finish. But, considering questions as the following: “*Are activities planned taking into account the resources available?*” we determine that everyone plans, good or bad, with sheets or mentally but everything is plan someway which in most of the cases the answer will be “YES”. Now, taking as objective reach a better approximation to reality it was analyzed the questionnaire of the different processes and they were adapted to accept a range of acceptance.

As example we show the adaptation of the previous question: “*In which way the activities are planned respecting the resources available?*”.

In this way questions that before just accepted “YES/NO” as answer now can accept range of acceptance.

To make it more practical and not to make the self-evaluation a tedious task it was defined the range of acceptance in four groups:

Table 3: Range of acceptance in answers

Percent	Description
1-25%	[Really Little]
26-50%	[Little]
51-75%	[Almost Everything]
76-100%	[Everything]

Furthermore, was made a categorization of the evaluation process questions, so in the final report that presents the evaluation results, the good quality practices that the enterprise doesn't fulfill on the different process can be grouped by category, getting on this way a better visualization of those aspects to improve to get better results in future evaluations.

4 Results

EVAP COMPETISOFT was applied on some software development organizations from different areas, from public organizations until private companies. Before they answer the new questionnaire there was a previous conversation about the objectives of their help and explain them about the COMPETISOFT project and the benefits that it can generate in quality subject on small and medium enterprises.

The idea of applying the app on real software development organization was to show that COMPETISOFT is not just a theoretical model, is a useful framework that can help to improve quality practices. Another objective was discover which of the good practices that the questionnaire presents are more difficult to satisfy considering the reality of the small and medium enterprises. Furthermore, to get an overview about the reality of this type of organization respect of quality practices.

EVAP Competisoft's validation was made on 10(ten) enterprises and development groups, among them: 1(one) big enterprises, 2(two) medium enterprises, 5(five) SMEs and 2(two) public organizations.

4.1 Processes evaluated

It was given a new profile to each organization in the application to evaluate the different categories processes that make up the COMPETISOFT process model leaving to them to choose the process to evaluate.

The results of the different evaluated processes by each organization are described below:

Table 4: Processes Evaluated

Category	Process	Level						Total
		0	1	2	3	4	5	
Management	Processes Management	3	4	-	-	-	-	7
	Resources Management	3	1	1	-	-	-	5
	Human Resources Management	1	1	-	-	1	-	3
	Goods, Services and Infrastructure Management	2	-	1	-	-	-	3
Operation	Specific Project Management	-	2	1	-	-	-	3
	Software Development and Software Maintenance	2	5	1	-	-	-	8
	Total	11	13	4	-	1	-	29

As we can see from all the processes evaluated:

- 11 of them are on level 0 or Incomplete → represents the 38% of the total.
- 13 of them are on level 1 or level Completed → represents the 45% of the total.
- 4 of them are on level 2 or level Administrated → represents the 14% of the total.
- 1 of the are on level 4 or level Predictable → represents the 3% of the total.

The following table shows the results of the different evaluated processes grouped in the other by category, process and level compliance.

Table 5: Level compliance of the evaluated processes

Level	Level Compliance	Management	Operation	Total
-------	------------------	------------	-----------	-------

		PM	RM	HRM	GSIM	SPM	SD-SM	
1	Between 0 and 15%	-	1	-	2	-	1	4
	> 15 % up to 50 %	3	2	1	-	-	1	7
	> 50 % up to 85 %	4	1	1	-	2	5	13
	> 85 up to 100 %	-	1	1	-	2	1	5
2	Between 0 and 15%	-	-	-	1	-	-	1
	> 15 % up to 50 %	2	4	2	1	1	1	11
	> 50 % up to 85 %	5	1	-	-	3	7	16
	> 85 up to 100 %	-	-	1	-	-	-	1
3	Between 0 and 15%	-	1	-	2	-	-	3
	> 15 % up to 50 %	2	3	1	-	1	5	12
	> 50 % up to 85 %	5	1	1	-	1	2	10
	> 85 up to 100 %	-	-	1	-	2	1	4
4	Between 0 and 15%	2	-	2	1	-	5	10
	> 15 % up to 50 %	-	4	-	1	1	-	6
	> 50 % up to 85 %	-	-	-	-	2	-	2
	> 85 up to 100 %	5	1	1	-	1	3	11

4.3 Quality practices not implemented

With the result of the evaluations carried out, it was possible to identify those best practices proposed by COMPETISOFT that are more difficult to meet by small and medium-sized companies.

The table below summarizing the results founded.

Table 6: Summary of categories of quality practices not implemented by process

Type of Question	Management				Operation		Total
	PM	RM	HRM	GSIM	SPM	SD-SM	
Knowledge Base	-	-	-	-	1	-	1
Automation	-	4	-	1	9	-	14
Documentation	11	12	11	6	99	-	139
Control	1	-	2	-	-	7	10
Definitions	8	10	11	1	47	17	94
Reviews	2	-	2	5	-	-	9
Validations	-	-	2	2	-	-	4
Management	2	-	-	-	-	-	2
Capacitation	3	16	7	23	2	20	71
Total							344

As can be seen in the table above, the weakest aspects of the evaluated processes are mainly lack of documentation, lack of definitions and lack of training.

Below the details in percentages:

- Knowledge Base → 0.29%
- Automation → 4.06%
- Documentation → 40.40%
- Control → 2.90%
- Definitions → 27.35%
- Evaluations → 2.61%
- Validations → 1.16%
- Management → 0.58%
- Training → 20.63%

4.4 Comments about the evaluations.

After the evaluations each organization gave their opinion about the quality practices proposed by COMPETISOFT and if they believe that are hard to apply them considering the reality of small and medium enterprises. Next are detailed those comments:

Company 1:

- “It wasn’t difficult to understand the questions, I believe that to all the people what work on this area should know what they mean”.

- “The implementation of the quality practices will depend on what kind of organization you are working with, what I mean with this, in public administration is very difficult because unfortunately is led by politics and they are changing permanently, they change from one day to another. I see it easier in private organizations with enough people in the required areas because it would be an overwork on one person and in the end things would be made in hurry and in a wrong way.”
- “There must be a lot of emphasis on it because time is a factor that always goes against the organizations and with our economy it would take more time to develop a system and what the price of the system today won't be the same in two months. This also means that a lot of steps that we have to do are not made, of course this is more easy for a big company, where all these steps are established. ”

Company 2:

- “The questions weren't difficult to understand and to answer them.”
- “Actually I'm working in a small area and almost new so most of the practices we don't apply it.”
- “I think it wouldn't be difficult to apply the practices that the application mentions, even more, it would be really useful if we could do it”

Company 4:

- “The organization where I'm working it doesn't work that much in quality aspects, is quite little or nothing at all.”
- “I think it would be great and that we have to implement them here, but it would demand too much work and time, but it would be good to implement them because it would fix a lot of problems that we have now.”
- “I think for a big company it would be easier to implement these practices than in a small one like this.”

Company 5:

- "I consider the questions excellent although I do not think they are feasible to be fully implemented in small companies. I also consider that the demands will come from the companies that implement it. "
- "I also found that the question form was very long, it had too many questions."
- "As I mentioned earlier, the questions are well written, but some I had to read twice to know what they were referring to, I think a context should be put in place to make them understand them better."

Company 6:

- "In my case it is not feasible because we are still very young."
- "I think in the case of large software companies this type of practices is easier to implement."
- "The questions are understood and clear. They ignore ambiguities when it comes to answering. "
- "It is almost impossible for a SME to adopt as well as the proposed questions. I think we should lower the requirements for this type of company. "

Company 9:

- "The questions are easy to understand but I think you should use a language more accessible and not so technical."
- "I think the practices mentioned are very idealistic about the reality of a software SME, but being able to implement them would help considerably in the performance of companies."

5 Conclusions and Future Guidelines

Evaluations carried out by software development companies have highlighted the lack of some aspects that companies should make and place more emphasis, such as documentation, definitions and training.

The lack of attention in these aspects was reflected in the high percentage of processes that obtained low levels of capacities, 83% of the processes evaluated are at level zero or level one.

At the same time it was observed that companies are aware of the lack of quality practices and the main reason is due to the lack of time and staff to be able to adjust this reality.

It was also observed that the incorporated modifications helped not only to have questions that are closer to the reality of the evaluated processes, but also to determine the causes or factors that influenced the result through the categorization of the questions of the questionnaire.

As future guidelines for the tool it is proposed:

1. Within the profile of the administrator add the management of questions, activities, items, processes and categories introduced by COMPETISOFT.
2. Within the manager's profile add the generation of statistical reports based on the results of the evaluations.
3. Within the profile of the user modify that a same user can be the owner of more than one company and answer questionnaires of any of the registered companies.
4. Within the profile of the user modify so that the questionnaire is not restricted to answer everything at one time. It is not currently possible to continue the questionnaire on another day than the day the questionnaire was started. That way you will gain flexibility with the users' time, especially in long questionnaires such as the Specific Project Management.
5. Within the profile of the user modify so that the report with the results of the evaluation contains recommendations based on the missing practices. Currently the report generated after the evaluation of the process contains a sector dedicated to the shortcomings found that are those practices that the company responded that do not comply. It is proposed to modify the sector of failure of the report so that it not only shows the practices not performed but also recommendations on how to improve these missing aspects.

References

1. COMPETISOFT. Mejora de Procesos para Fomentar la Competitividad de la Pequeña y Mediana Industria del Software de Iberoamérica. Versión 0.2. Proyecto COMPETISOFT 506PI287.CYTED.http://artemisa.unicauca.edu.co/~ecaldon/docs/spi/COMPETISOFT_v02_27-
2. Aplicación para la Evaluación de la Capacidad de Procesos basado en el Modelo COMPETISOFT. Lic. Leonardo A. Fernández, Mgter. Gladys Da Pozo, Mgter. Somia Mariño. Facultad de Ciencias Exactas, Naturales y Agrimensura - Universidad Nacional del Nordeste.
3. Pressman, R.S. (1998). "Ingeniería del software. Un enfoque práctico". 4ª Edición. Mc Graw Hill.
4. Observatorio Permanente de la Industria del Software y Servicios Informáticos (OPSSI). "Reporte anual sobre el Sector de Software y Servicios Informáticos de la República Argentina". Abril 2016. Publicado en: <http://www.cessi.org.ar/descarga-institucionales-2007/documento2-130347cd83ae771a9f3db3da5407269a>

XII Workshop Arquitectura, Redes y Sistemas Operativos (WARSO)

Coordinadores

Ardenghi, Jorge – Buckle, Carlos – Micolini, Orlando – Arroyo,
Marcelo

Comité Científico

Acosta, Nelson (Argentina)	Marrone, Luis (Argentina)
Ardenghi, Jorge (Argentina)	Matrángolo, Carlos (Argentina)
Bevilacqua Roberto (Argentina)	Mercado, Gustavo (Argentina)
Buckle, Carlos (Argentina)	Micolini, Orlando (Argentina)
Castro Lechtaler, Antonio (Argentina)	Molinari, Lía (Argentina)
Díaz F. Javier (Argentina)	Padovani, Hugo (Argentina)
Echaiz, Javier (Argentina)	Steinmetz, Ralf (Alemania)
Fischer, Stefan (Alemania)	Tarouco, Liane (Brasil)
García Garino, Carlos (Argentina)	Ultreras, Florencio (Chile)
García Villalba, Luis Javier (España)	Vendrell Vidal, Eduardo (España)

A novel performance metric for Virtual Network Embedding combining aspects of Blocking Probability and Embedding Cost

Enrique Dávalos and Benjamín Barán

Universidad Nacional de Asunción, Facultad Politécnica
San Lorenzo, Paraguay

{edavalos, bbaran}@pol.una.py

<http://www.pol.una.py>

Abstract. Network Virtualization offers a solution for Future Internet and it is a key enabler for cloud computing applications. *Virtual Network Embedding* (VNE) problem deals with resource allocation of a physical infrastructure to *Virtual Network Requests* (VNRs). Several performance metrics are employed in order to evaluate the efficiency of specific VNE approaches. These existing metrics, mostly related to Infrastructure Provider profit, are computed at the end of the VNE process, after embedding many VNRs. This work proposes a novel performance metric, VNE-NP (*VNE Normalized Profit*) which combines aspects of the two metrics most used in the literature: *Blocking Probability* and *Embedding Cost*.

Keywords: network virtualization; virtual network embedding; performance metrics

1 Introduction

Network architecture is critical for cloud computing applications, which require distributed datacenters interconnected through high data transfer and low delay networks [1]. Besides, a high level of virtualization in datacenters resources and networking devices are required for the provisioning of cloud services, which explains the increasing importance of technologies related to Network Virtualization [2].

The problem of optimally assigning resources of a physical network to *Virtual Network Requests* (VNRs) is usually named *Virtual Network Embedding* (VNE) problem and it is the main resource allocation challenge in Network Virtualization [3].

Several performance metrics have already been used to compute how optimal or efficient a specific VNE approach is. Surveys as [3] and [4] present these metrics, which are computed at the end of a VNE process (a large amount of VNR mappings), and are related to Quality of Services (QoS) or profit issues.

In this work, we present an overview of the performance metrics that are used to evaluate profit-related aspects. Next, a new performance metric is proposed

for the first time, named *VNE Normalized Profit* (VNE-NP), considering both aspects of efficiency in finding a possible mapping of a specific VNR as well as physical network resources utilization.

The rest of this work is organized as follows: Section 2 presents a VNE formulation and main performance metrics, while the performance metrics related to economical profit are discussed in Section 3. Section 4 proposes the novel VNE-NP metric and a VNE example shows its viability. Section 5 concludes this work.

2 Virtual Network Embedding (VNE) problem

2.1 Physical network

When considering a VNE problem, a physical network is modeled as an undirected graph $G^P = (N^P, L^P)$ where N^P is the set of physical nodes and L^P is the set of bi-directional physical links.

Each physical node $n^P \in N^P$ can be considered as connected to a datacenter whose available computing capacity is represented by a single parameter c^n , usually called *computing resource capacity*. At the same time, the physical links $l^P \in L^P$ interconnect physical nodes. Each physical link is characterized by its available bandwidth indication, *i.e.* its capacity to support data traffic.

2.2 Virtual network requests

Each virtual network request (VNR) is also usually modeled as an undirected graph $G^V = (N^V, L^V)$ where N^V is the set of virtual nodes and L^V is the set of virtual links. The requirements of each virtual node and each virtual link are typically related with the same parameters that characterize physical nodes and links respectively, *i.e.* each virtual node requires a specific number of computing resource units, while each virtual link requires a specific bandwidth.

2.3 VNE problem formulation

A given VNE algorithm should process one VNR at a time, trying to efficiently allocate virtual nodes into physical nodes and virtual links into paths formed by physical links. Therefore, a VNR allocation process can be divided in two mapping functions:

1. *Virtual Node Mapping* (VNM), which assigns each virtual node n^V to a physical node n^P :

$$VNM := f(n^V \rightarrow n^P) \quad (1)$$

2. *Virtual Link Mapping* (VLM), which assigns to each requirement of virtual link l^V a path of consecutive physical links:

$$VLM := f(l^V \rightarrow p^P) \quad (2)$$

$p^P \in P^P$ where P^P is the set of all possible paths (or lightpaths in the case of a physical optical network) between a pair of physical nodes which host the corresponding pair of virtual nodes interconnected by l^V .

A feasible solution must comply with specified restrictions, mostly related to the availability of physical resources to host virtual nodes or links, and the nature of the physical network.

Each VNR allocation may be chosen to optimize a single objective function, or multiple objective functions, in order to choose the most convenient solution among all feasible alternatives. In this sense, each VNR allocation can be considered as an optimization problem, with a discrete and finite number of solutions that satisfy specific restrictions.

If the VNE process succeeded in finding a feasible (and ideally optimal) solution, then the VNR is *embedded* and the physical resources required are assigned to it. Otherwise, the VNR is *blocked*. There is not partial VNR allocation.

2.4 VNE global performance metrics

The individual result in the allocation or blocking of a single VNR does not define the real efficiency of the whole VNE process. The VNE process comprises multiple VNR allocations and it is evaluated by global, long-term *Performance Metrics*, which can only be computed at the end of a whole VNE process, typically given by a period of working (or simulation) under a required load environment.

VNE performance metrics can be classified into those related to: (1) resources spending or economical profit, (2) Quality of Service issues, and (3) others metrics [3]. The main performance metrics, utilized in almost all VNE approaches, are those related to resource spending or economical profit. Therefore, they will be explained in detail in next section.

3 Resources spending or economical profit metrics

These type of metrics try to measure how efficient is the VNE method in the utilization of physical resources or to accept the maximum number of VNRs. In this context, the most referenciated/used performance metrics are:

- *Accepted Ratio*[5];
- *Embedding Cost*[6];
- *Embedding Revenue*[7]; and
- *Revenue/Cost Relation*[8].

3.1 Accepted Ratio

Accepted Ratio performance metric [5] measures the relation between the number of accepted (not blocked) VNRs (VNR_{Acc}) over the total number of VNRs considered in the process (VNR_{Tot}):

$$Accepted\ Ratio = \frac{\sum VNR_{Acc}}{\sum VNR_{Tot}} \quad (3)$$

The values of *Accepted Ratio* is in the range [0,1] and higher values of this metric indicate better performance of a specific approach. For online problems, this metric is usually considered as its complement and called "*Blocking Rate*" or "*Blocking Probability*":

$$Blocking\ Rate = \frac{\sum VNR_{Block}}{\sum VNR_{Tot}} \quad (4)$$

where $\sum VNR_{Block}$ is the number of blocked VNRs, and therefore:

$$Blocking\ Rate = 1 - Accepted\ Ratio \quad (5)$$

Accepted Ratio represents the efficiency in finding a solution for allocating a VNR. However, an approach specially designed to decline this metric, can accept all VNRs with fewer amount of required resources and with simpler topologies, while blocking more complex VNRs.

3.2 Embedding Cost

Embedding Cost [6] or *Physical Resource utilization* performance metric computes total cost of embedding, in terms of physical resources that were needed to embed accepted requests.

The embedding cost of an individual VNR_i ($Cost_i^{VNR}$) can be considered as proportional to physical resources needed to perform virtual node and virtual link mappings:

$$Cost_i^{VNR} = (\gamma \sum_{N^P} c_{n,i}^P + \delta \sum_{L^P} B_{l,i}^P) \cdot HT_i \quad (6)$$

where $c_{n,i}^P$ represents physical node resources assigned to virtual nodes of VNR_i ; $B_{l,i}^P$ are physical link resources (bandwidth) assigned to virtual links of VNR_i ; while γ and δ are coefficients that relate real costs per time unit of node resources and bandwidth resources of physical paths. At the same time, HT_i is the Holding Time (lifetime) of a single request VNR_i . If the problem is a static VNE, HT_i can be considered as equal to 1, or it may not be considered at all.

The *Embedding Cost* performance metric can be defined as the sum of all the embedding costs corresponding to accepted (not blocked) VNRs:

$$Embedding\ Cost = \sum_{Acc} Cost_i^{VNR} \quad (7)$$

The values of this metric are not normalized and the VNE strategy must try to *minimize* its value, looking for an efficient utilization of physical resources. However, this metric does not consider neither the number of blocked VNRs, nor the number of total resources successfully embedded, so it is also a partial view of VNE efficiency.

3.3 Embedding Revenue

Embedding Revenue considers that the revenue of a VNR is proportional to the number of total node and link resources requested by virtual networks [7], [9].

The potential revenue of a single VNR_i , or $Revenue_i^{VNR}$ (which can be effective, once embedded) can be defined as

$$Revenue_i^{VNR} = (\alpha \sum_{N^V} c_{n,i}^V + \beta \sum_{L^V} B_{l,i}^V) \cdot HT_i \quad (8)$$

where α and β are coefficients related to real revenue per time unit, of each unit of computing capacity resources $c_{n,i}^V$ required by a virtual node and amount of spectrum / wavelength $B_{l,i}^V$ required by virtual links (network resources) of VNR_i .

The *Embedding Revenue* performance metric can be defined as the sum of all potential revenues corresponding to all accepted VNRs:

$$Embedding Revenue = \sum_{Acc} Revenue_i^{VNR} \quad (9)$$

The *Embedding Revenue* is a metric that must be *maximized*, looking for better profit of the *Infrastructure Provider* or *InP* and, as the previous metric, their values are not normalized.

3.4 Revenue/Cost Relation

Revenue/Cost Relation performance metric indicates the ratio between *Embedding Revenue* and *Embedding Cost* [8], [10]:

$$Rev/CostRelation = \frac{Embedding Revenue}{Embedding Cost} \quad (10)$$

This VNE performance metric does not express global efficiency of a VNE approach, because it is focused, as *Embedding Cost* and *Embedding Revenue* metrics, only in accepted VNRs. A VNE strategy rejecting a lot of VNRs may be very efficient in physical resources utilization, and therefore with a good *Revenue/Cost Relation*, but with a high blocking probability.

Revenue/Cost Relation performance metric must be *maximized*, for a good profit of the *InP*. Furthermore, if the values of parameters α, β, γ , and δ are taken equal to one, this metric can be considered as normalized, with its values varying between 0 (in the worst case) and 1 (for a perfect VNE process with all virtual links mapped to single physical links) [8].

4 A novel profit-related performance metric: VNE-NP

As we have seen, the diversity in the topology of each VNR and the resources that requires, makes that none of the above mentioned performance metrics

is able to reflect all the different aspects of the problem, failing in adequately comparing the efficiency of two or more different approaches.

The *Accepted Ratio* metric evaluates only how many VNRs were accepted, without taking into account neither the potential lost revenue when large VNRs were blocked nor the efficiency in the utilization of physical resources in the VNE process. In terms of *InP* profit, there can be a considerable difference in blocking a relative large VNR, than blocking a small one that only requires a small number of resources.

The *Embedding Cost* metric only evaluates how efficiently is the VNE approach in physical resources utilization, with no consideration about the number of blocked VNRs and their potential lost revenue.

On the other hand, the rest of the considered metrics: *Embedding Revenue* and *Revenue/Cost Relation* do consider the revenue generation of accepted requests but gives a partial evaluation since only consider the cost of accepted VNRs.

To mitigate this situation with the already published metrics, a novel profit-related performance metric is proposed in this work, named *Virtual Network Embedding Normalized Profit* (VNE-NP). First, *Potential Revenue* is defined as the Total Revenue of *both* accepted and blocked VNRs (not only accepted VNRs as in *Embedding Revenue*):

$$Potential\ Revenue = \sum_{Tot} Revenue_i^{VNR} \quad (11)$$

The VNE-NP metric is defined by:

$$VNE - NP = \frac{Embed.Revenue - Embed.Cost}{Potential\ Revenue} \quad (12)$$

It is important to notice that while *Embedding Revenue*, *Embedding Cost* and *Revenue/Cost Relation* are computed over only the accepted (not blocked) VNRs as can be seen in equations (7), (9) and (10), the *Potential Revenue* (11) considers both accepted and blocked VNRs.

The VNE-NP performance metric expresses *Net Profit* of a single successfully embedded VNR as the difference between its Revenue and Cost, which are normalized by the total *Potential Revenue* of all VNRs in the process (blocked and accepted VNRs).

An additional benefit of VNE-NP is its normalization over potential total revenue, which permits a better comparison among different instances of the problem.

This way, VNE-NP takes into account all important aspects of profit for a VNE process, including:

1. the ability of finding a feasible solution for any VNR and therefore to accept the larger possible potential source of profit (*Potential Revenue*);
2. the efficiency of a VNE algorithm in physical resources utilization in the embedding process; and
3. the ability in generating more revenue, accepting more complex requests.

Table 1. A comparison among the five VNE performance metrics studied

Performance Metric			Affected by		
Name	Max/ Min	Norma- lized	Lost potential profit of blocked VNR	Efficiency in utilization of physical res.	Efficiency in getting more revenue
<i>Accepted Ratio</i>	Max	Yes	YES	NO	NO
<i>Embedding Cost</i>	Min	No	NO	YES	NO
<i>Embedding Revenue</i>	Max	No	YES	NO	YES
<i>Revenue/Cost Relation</i>	Max	No	NO	YES	YES
<i>VNE-NP</i>	Max	Yes	YES	YES	YES

The above advantages are shown in Table 1, which compares the five metrics in these three above mentioned aspects.

In what follows, an example is presented to show the difference of using VNE-NP with respect to other known metrics and how it can represent a better figure of merit for *InP* profit for different situations.

4.1 An example of VNE-NP application

In Figure 1 (a) it can be seen a physical network with 5 nodes (n_1^P to n_5^P) and 6 links (l_1^P to l_6^P). Each physical node has 3 available CPU units and each physical link has 2 bandwidth units (indicated in parenthesis). Figure 1 (a) also shows three VNRs (VNR 1, VNR 2 and VNR 3) which must be embedded in the physical network using a VNE approach. Each virtual link requires one bandwidth unit, and each virtual node requires one CPU unit, which are not indicated in the figure for simplicity.

We also consider three different VNE algorithms (VNE algorithms A, B and C) which offers three different embedding solutions showed at Figures 1 (b), (c) and (d) respectively.

As we can see in Figure 1 (b), VNE Algorithm A only was able to embed VNRs 1 and 2 while VNR 3 could not be allocated, because there were not enough physical link resources. On the other hand, in Figure 1(c), Algorithm B succeeded in embedding VNRs 2 and 3, but it could not embed VNR 1. However, Algorithm C was successful in embedding the three VNRs, as it is showed in Figure 1(d).

Notice that the three VNRs require different numbers of virtual nodes and virtual links and the total demand of resources are different, and that is why they represent different amount of profit for the *InP*.

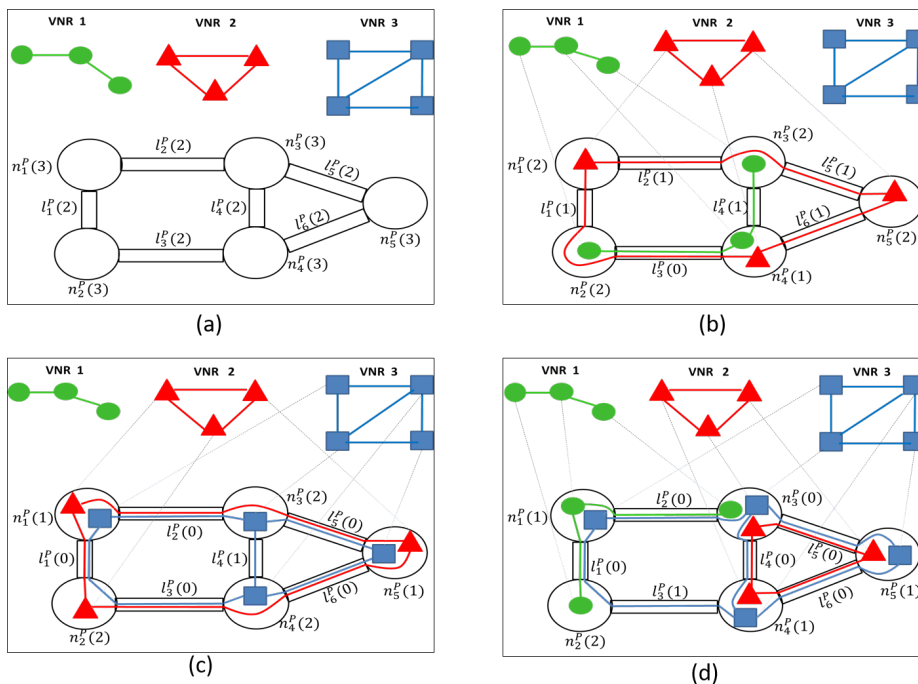


Fig. 1. Embedding of VNE algorithm C.

Table 2. Values of Performance Metrics for the 3 studied VNE algorithms

Performance Metric	VNE Algorithms		
	Algorithm A	Algorithm B	Algorithm C
<i>Accepted Ratio</i>	0,67	0,67	1,00
<i>Embedding Cost</i>	13,00	18,00	21,00
<i>Embedding Revenue</i>	22,00	30,00	40,00
<i>Revenue/Cost Relation</i>	1,70	1,67	1,90
<i>VNE-NP</i>	0,23	0,30	0,48

Let us calculate and compare the values of the five performance metrics related to InP profit already defined: *Accepted Ratio*, *Embedding Cost*, *Embedding Revenue*, *Revenue/Cost Relation* and *VNE-NP*.

Taking $\alpha = \beta = 2$ (The revenue of embedding a virtual node or link gives to the InP an equivalent revenue of two physical units), and $\gamma = \delta = 1$ (CPU units cost the same as Bandwidth units), Table 2 shows the values of the five different performance metrics for the three VNE algorithms.

4.2 A discussion of the results

An initial and important observation is that Algorithm C reaches best values for 4 out of 5 metrics: *Accepted Ratio*, *Embedding Revenue*, *Revenue/Cost Relation* and *VNE-NP*. That was an expected result, since Algorithm C succeeded in embedding the three VNRs, reaching the best revenue and revenue/cost relation in the process. Only *Embedding Cost* found Algorithm A as the best. This indicates that this metric can be strongly influenced by the diversity of resource demands for different VNRs and can fail in performing a good evaluation.

Now, we can obviate Algorithm C and compare only results for Algorithms A and B for the four remaining performance metrics. We can notice some important aspects:

- *Accepted Ratio* shows the same results for both algorithms (0,67). Although both algorithms accepted two VNRs and blocked one, the potential revenue of VNR 2 is larger than VNR 1 given that VNR 2 needs more resources. This important aspect is not considered when using *Accepted Ratio* performance metric.
- *Revenue/Cost Relation* found better results for Algorithm A, while the two remaining metrics: *Embedding Revenue* and *VNE-NP* coincide in finding Algorithm B better than Algorithm A. The difference in complexity between blocked VNRs indicates that this is the expected result. That suggests that *Revenue/Cost Relation* fails in adequately compare these two algorithms.

This analysis indicates that for the above example, *Embedding Revenue* and *VNE-NP* best express the economical benefit of the *InP* and the ability in embedding more complex VNRs, with a clear advantage in using *VNE-NP* given that it is normalized while *Embedding Revenue* is not.

5 Conclusion

Virtual Network Embedding (VNE) problem leads with the efficient allocation of physical network resources to virtual network requests. A VNE approach is evaluated with global, long-term performance metrics, that are computed at the end of a VNE process (or simulation).

Most used VNE performance metrics are related to economical profit of Infrastructure Provider (InP), being "*Embedding Cost*" and "*Accepted Ratio*" the most cited in the literature. However, they can lose objectivity when the resources requested by virtual network requests (VNRs) are heterogeneous, as shown in Section 4.

In this work, we propose a novel performance metric, named *Virtual Network Embedding-Normalized Profit* (VNE-NP), presenting an example where the proposed metric proved to give a better figure of merit than any other metric. The proposed performance metric offers a good, normalized representation of the profit reached by an *InP* in the VNE process, and considers both the savings

in physical resources utilization as well as the ability of finding feasible efficient solutions for a maximum number of VNRs.

In future works, the authors plan to implement VNE-NP to network simulators, in order to evaluate VNE instances jointly with other performance metrics. Besides, it would be interesting to study real economical values of physical network resources (nodes and links resources) in cloud applications, aiming a more real evaluation of VNE-NP metric.

References

1. Alshaer, H.: An overview of network virtualization and cloud network as a service. *International Journal of Network Management*. 25, 1–30 (2015)
2. Nejabati, R., Peng, S., Simeonidou, D.: Role of optical network infrastructure virtualization in data center connectivity and cloud computing. In: *Optical Fiber Communication Conference and Exposition and the National Fiber Optic Engineers Conference (OFC/NFOEC)*, pp. 1–3. (2013)
3. Fischer, A., Botero, J.F., Beck, M.T., De Meer, H., Hesselbach, X.: Virtual network embedding: A survey. *IEEE Communications Surveys & Tutorials*. 15, 1888–1906 (2013)
4. Belbekkouche, A., Hasan, M.M., Karmouch, A.: Resource discovery and allocation in network virtualization. *IEEE Communications Surveys & Tutorials*. 14, 1114–1128 (2012)
5. Chowdhury, N. M. M. K., Rahman, M.R., Boutaba, R.: Virtual network embedding with coordinated node and link mapping. In: *IEEE INFOCOM*, pp. 783–791. (2009)
6. Yu, H., Qiao, C., Anand, V., Liu, X., Di, H., Sun, G.: Survivable virtual infrastructure mapping in a federated computing and networking system under single regional failures. In: *IEEE Global Telecommunications Conference GLOBECOM*, pp. 1–6. (2010)
7. Chowdhury, M., Rahman, M.R., Boutaba, R.: Vineyard: Virtual network embedding algorithms with coordinated node and link mapping. *IEEE/ACM Transactions on Networking*. 20, 206–219 (2012)
8. Lischka, J., Karl, H.: A virtual network mapping algorithm based on subgraph isomorphism detection. In: *Proceedings of the 1st ACM workshop on Virtualized infrastructure systems and architectures*, pp 81–88. ACM (2009)
9. Rahman, M.R., Boutaba, R.: SVNE: Survivable virtual network embedding algorithms for network virtualization. *IEEE Transactions on Network and Service Management*. 10, 105–118 (2013)
10. Dávalos, E., Aceval, C., Franco, V., Barán, B.: A multi-objective approach for VNE problems using multiple ILP formulations. *CLEI Electronic Journal*. 19, 1–15 (2016)

An IPC Software Layer for Building a Distributed Virtualization System

Pablo Pessolani¹, Toni Cortés², Fernando G. Tinetti^{3,4} and Silvio Gonnet^{1,5},

¹FRSF, Universidad Tecnológica Nacional, Santa Fe, Argentina

²Barcelona Supercomputing Center & UPC, Barcelona, España

³III-LIDI Facultad de Informática – UNLP, La Plata,

⁴Comisión de Investigaciones Científicas, Buenos Aires, Argentina

⁵INGAR - CONICET, Santa Fe, Argentina

Abstract. Hardware virtualization technologies were principally designed for server consolidation, allowing multiple Operating Systems instances to be co-located on a single physical computer. But, IaaS providers always need higher levels of performance, scalability and availability for their virtualization services. These requirements could be met by a distributed virtualization technology, which extends the boundaries of a virtualization abstraction beyond a host. As a distributed system, it depends on the communications between its components scattered in several nodes of a virtualization cluster. This work contributes M3-IPC, an IPC software layer designed to facilitate the development of an OS-based distributed virtualization.

Keywords: Distributed Systems, Virtualization, IPC.

1 Introduction

Hardware virtualization is the most common technology that provides Virtual Machines (VM) as abstractions. Operating System (OS) level virtualization is another technology that is located at a higher level, encapsulating user-space applications within Containers or Jails or Virtual Operating Systems (VOS) [1]. On the other hand, in the 1980s, several Distributed Operating Systems (DOS) [2, 3, 4, 5] were developed as a consequence of the limited CPU performance of a single host and the need for scalability and higher computing power.

Thinking of a distributed virtualization technology seems to make sense to achieve higher performance and increase service availability. OS-based virtualization and DOS technologies lead authors to think about their convergence to achieve these goals, extending the boundaries of the virtualized abstraction to multiple hosts and thereby running multiple isolated instances of DOSs sharing the same cluster.

A distributed OS-based virtualization approach will explore aggregation with partitioning. In such systems, a set of server processes constitutes a distributed OS running within an execution context which authors call “*Distributed Container*” (DC). Processes belonging to a DC may be scattered on several nodes of a cluster (aggregation); and processes of different DCs could share the same host (partitioning).

The development of such distributed systems would be facilitated with an IPC infrastructure that provides uniform semantics without considering process location. But, as it is also a virtualization system, this IPC infrastructure should support confinement to isolate communications among processes of different DCs. In addition to these features, other characteristics are required which are related to generic distributed systems. They refer to keeping IPC operational even during process migration and supporting fault-tolerant applications which use the *Primary-Backup* approach [6]. Therefore, there is also a need to redirect communications addressed to the failed *Primary* towards one of its *Backup* processes that becomes the new primary. All these features should be complemented by a suitable performance to make Distributed Virtualization a feasible approach.

This paper proposes M3-IPC, a general purpose IPC software layer which is the foundation of a "*Distributed Virtualization System*" (DVS) [7], a new model of OS-level virtualization for Cloud Computing (IaaS). Distributed Virtualization must not be confused with Clustered Virtualization [8, 9] in which an application could run in a distributed way across a group of Containers located on several nodes of a cluster. On such systems, the boundaries of each container are limited by the node where they run, and applications must be developed using special middleware to extend APIs, which avoid the direct migration of legacy applications.

M3-IPC was developed to communicate distributed components of a VOS running within a DC, but it can be used as a powerful infrastructure for designing generic distributed applications. Although the IPC mechanism could be embedded within the DVS, the authors considered useful to build it as an independent software module. Before M3-IPC design, several IPC mechanisms were evaluated, but none of them meet the aforementioned requirements.

The next sections present some works related to IPC mechanisms, followed by a sketch of DVS components, M3-IPC concepts, design goals and implementation issues. Then, results of performance tests are discussed in Evaluation Section. Finally, Conclusions and Future Works Sections summarize the main features of the resulting software, including future improvements.

2 Motivation and Related Works

M3-IPC focuses on providing an IPC software layer to enable N-to-N communications between Client and Servers without any intermediary broker. It should facilitate the development of complex distributed systems such as an OS-level DVS. A great deal of effort had been already taken in providing IPC mechanisms for distributed systems, some of them being integrated as a component within classical DOS [10, 11] and others as added software through libraries, patches, modules, etc. [12, 13, 14, 15, 16]. Their features, semantics and performance, were previously evaluated and during M3-IPC design, implementation, and testing stages. For space limitation reasons, only a representative set of these suitable IPC software is considered here:

- Synchronous Inter-process Messaging Project for Linux (SIMPL) [12].
- Send/Receive/Reply (SSR) [13].
- Distributed IPC (DIPC) [14].

- Remote Procedure Call (RPC) [15].
- Telecommunications IPC (TIPC) [16].

All of the above presented IPC mechanisms were used for performance comparison against M3-IPC as it is shown in Evaluation Section. URPC [17] was not considered because, although it may reach a good performance, it is weak on issues related to security. Messages Queues, Pipes, FIFOs, and Unix Sockets do not have the ability to communicate with remote processes. DIPC presents interesting features, but it is no longer maintained.

One of the most important features needed to build a DVS is IPC isolation, which includes the following properties:

1. *Confinement*: A process running within a DC cannot communicate with any process within another DC.
2. *Private Addressing*: An endpoint number (an M3-IPC address which identifies a process or thread) allocated to a process running within a DC could be allocated to other processes running within other DCs.
3. *Virtualization Transparency*: A process does not need to know that it is a member of a DC or the node in which is running, or in which nodes other processes of the same DC are running. To allow DVS management, privileged processes could allocate endpoints and DCs for other non-related processes.

None of the evaluated IPC software meet all of these requirements considering that a DC could span several nodes of a cluster.

3 M3-IPC Design

This section describes the design outline and rationale for M3-IPC to provide local and remote IPC for centralized and distributed systems. It presents the DVS topology model for which it was originally designed, its design goals, main concepts and implementation issues.

3.1 DVS Components

A DVS consists of the following components (Fig. 1):

- *DVS*: It is the top level layer that assembles all cluster nodes and it embraces all DCs.
- *Node*: It is a computer that belongs to the DVS where processes of several DCs are able to be run. All nodes are connected by a network infrastructure.
- *DC*: It is the group or set of related processes that might be scattered on several nodes. M3-IPC only allows communications among processes that belong to the same DC. The boundary of each DC can be based on administrative boundaries. A DC hides its internals from the outside world and hides network communication issues from its processes.
- *Endpoint*: An endpoint is an integer number which identifies a process/thread registered in a DC. Endpoints are unique and global within a DC, but could be repeated within other DCs.

M3-IPC does not impose a network/transport protocol to be used for inter-node communications. It allows programmers to choose the protocol that best fit their needs. Nodes communicate among them through *proxies*.

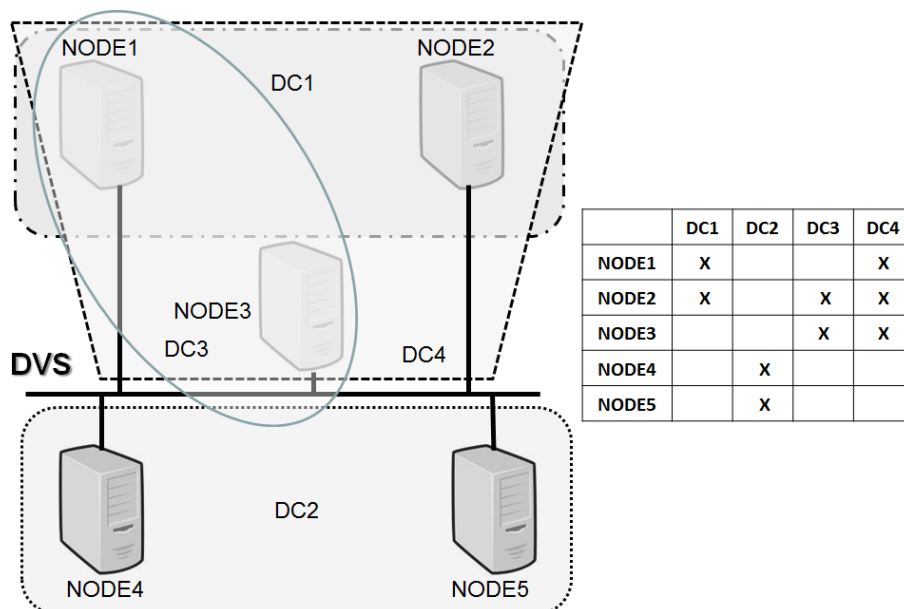


Fig. 1. DVS topology model.

3.2 M3-IPC Design Goals

The following design goals were established:

- *Easy-to-use*: Defining, implementing, and deploying applications using M3-IPC should be straightforward.
- *Location Transparency*: The application should not need to consider the location of the involved processes in order to facilitate programming.
- *Confinement*: IPC must be confined so as to achieve a proper degree of isolation.
- *Concurrency*: It should support threads to maximize throughput and efficiency.
- *Transparent Live Migration Support*: The application would use process migration [18] as a mechanism to achieve workload balancing and downtime avoidance.
- *Transparent Fault Tolerance Support*: IPC must remain operational even during process switching from Primary to Backup [6] without the awareness of the other peer processes.
- *Performance and Scalability*: Performance goals are considered satisfied when meeting minimal thresholds established for local and remote transfers. IPC throughput between processes located on different nodes must be limited by the network capability (latency and bandwidth) and the transport protocol being used.
- *Communication Protocol Agnostic*: Developers can choose the Network/Transport Protocol to be used by proxies that best satisfies their needs.

- *Modular and Customizable*: It should be able to be used by any kind of application with similar requirements.
- *Client/Server and RPC Oriented*: M3-IPC must cover typical communication patterns between multiple clients against multiple servers.
- *Security*: Basic security principles must be applied to its design and implementation. M3-IPC is based on the assumption of a cluster of network nodes that are connected to a switched LAN where the packet loss rate is low and the available bandwidth is high.

3.3 M3-IPC Main Concepts

M3-IPC was designed to emulate the semantics of a multi-server OS within the Linux kernel providing user-space APIs. Furthermore, it expands its use to processes located on remote nodes of a cluster, and includes features that could be required by distributed applications such as application isolation, location transparency, and message redirection on process migration or process switching on replication.

As M3-IPC supports threads to maximize concurrency, the following references to processes are also valid for threads.

M3-IPC APIs are classified as:

- *Communication APIs*: Those APIs emulate Minix 3 [19] IPC primitives. They are related to message and data transfers among processes. They do not include any reference to the DC a process belongs to, or to the node where it resides.
- *Management APIs*: Those related to DVS, DCs, proxies, nodes, and process management, which allow the mapping of applications to nodes and DCs.

A process which will use M3-IPC must be previously registered at the kernel to bind a DC to it. The process can register by itself, or by another local process with management privileges. Every process bound to a DC will have a unique *endpoint* number with backup endpoints as exceptions.

Management APIs consider: *Local endpoints* are those allocated to processes running on the same node; and *Remote endpoints* are those allocated to processes running on other nodes.

Every remote server should be first registered in the client's node specifying in which node is running. Once a remote endpoint is registered by the kernel, all local registered processes (including other clients) within the same DC are able to refer to that server endpoint in communication APIs. Remote client binding is easier because M3-IPC can automatically bind remote unprivileged endpoints when their messages arrive to the local node.

M3-IPC supports message redirection on node switching when a Primary process (on node A) is replaced by a Backup process (on node B). The local M3-IPC kernel automatically unlinks the endpoint from the previous *Primary* process (on node A) and links it to the new *Primary* process (on node B). A local process can be registered as a *Backup* of a remote *Primary* process with the same endpoint in the same DC, but messages sent by other local processes addressed to that endpoint will be sent to the *Primary* process. The *Backup* process will not be able to communicate with any other process until it is promoted as the new *Primary* by a privileged process. Message replication among *Primary* and *Backup* processes does not concern M3-IPC.

M3-IPC can keep communications in operational state and can automatically redirect messages on live process migration. Before a process begins to migrate, a management application must call *mxn_migr_start()*. At that moment, all messages addressed to the

migrating endpoint will be queued up. Once the process has successfully migrated, the management application must call `mnx_migr_commit()`. All queued messages will be sent to the process on its new location. If the migration fails, the management application must call `mnx_migr_rollback()`; then the process can resume its execution as if nothing had happened by receiving the queued messages.

3.4 M3-IPC Implementation Issues

This section describes the implementation details of the M3-IPC core within the Linux kernel.

M3-IPC APIs use a software interrupt vector or interrupt gate that differs from that used by Linux system calls. This issue allows both APIs to share the same kernel with minimal interference or crosstalk. It also facilitates future maintenance of M3-IPC in current and future Linux kernels.

Two important decisions made before developing M3-IPC were: its construction being based on software components that are available in the Linux kernel, and taking advantage of the parallelism in SMP and multi-core systems. Before starting the development, several Linux kernel synchronization mechanisms and mutual exclusion facilities were evaluated. *Kernel semaphores* and *RCU* (Read-Copy-Update) were too slow; and *spinlocks* were somewhat slower than Read/Write locks (*rwlocks*). Finally, *mutexes* were used for mutual exclusion, but since they have a similar performance to that of *rwlocks*. A compile option is available for system programmers to choose which synchronization mechanisms and mutual exclusion facilities want to be used.

Another design decision was related to the mechanism for data transfer from user-space to kernel-space and vice versa. *Netlink sockets* [20] have a quite complex start-up which do not meet DVS project needs. *Efficient Capability-Based Messaging* (ECBM) [21] has an impressive performance, but basic security issues were neglected. Finally, functions `copy_to_user()` and `copy_from_user()` provided by the Linux kernel were used. A custom function named `copy_usr2usr()` was built to copy data from the user-space buffer of a source process to the user-space buffer of a destination process.

The most outstanding characteristics of how M3-IPC was finally implemented are summarized in the following list:

- APIs support threads and were implemented using Linux kernel provided mutual exclusion facilities; Task Queues and Event Waiting were used for process synchronization; and Reference Counters were used to hold a count of processes from which a data structure is referenced.
- The granularity of internal critical sections was maximized at the process/thread level, allowing parallel message transfers among multiple pairs of processes.
- Higher performance is achieved because DCs do not share any data structure among one another during concurrent message transfers between pairs of processes that belong to different DCs.
- Data structures that are frequently used for registered processes (*struct proc*) are aligned with L1 cache lines to reduce access time.
- An Affinity Inheritance Protocol (similar to the well-known Priority Inheritance Protocol) was implemented to minimize performance impact of cache ping-pong. Thus, the facility provided on Linux to specify process affinity with a set of processors/cores was used. With this approach, there is a greater chance for the

message to remain in L1 cache when the destination process is scheduled, thus reducing access time.

- A CPU mask could be allocated by each DC to specify on which CPUs its local processes are able to run (only meaningful within the each node).
- The copy of data blocks and message transfers among co-located processes are made from the source process address space to the destination process address space without any intermediate copy through the kernel. The copy is made by the Linux kernel through the copy-on-write mechanism.
- Debugging information is sent to the Linux kernel ring buffer and can be shown by means of *dmesg* command.
- Data blocks which are page-aligned and whose lengths are greater than or equal to the page size are copied using the kernel provided *page_copy()* function, which is very efficient because it uses MMX instructions.
- Information about configuration, status, and statistics of M3-IPC abstractions is presented as directories and files within Linux */proc* filesystem.

M3-IPC has been implemented in C programming language on Linux for Intel x86 32-bit and it is distributed as a kernel patch, a kernel module, and a set of libraries.

Proxies and IPC through the Network

M3-IPC uses application level proxies for communications between nodes, but also support kernel-level proxies too. Therefore, there is no restriction about the network/transport protocol proxies can use.

M3-IPC APIs provides a function to get messages from the kernel that need to be sent to remote nodes, and another function to insert messages into the local kernel coming from remote nodes to local processes as destinations.

The current M3-IPC distribution provides the use of one pair of proxies (sender-receiver) for each remote node. Proxies exchange *proxy messages* which consist of fixed length headers and, eventually, variable length payloads (data blocks).

The provided proxies send/receive messages and data through the network without taking account to which DC those messages or data belong to. Custom proxies may consider implementing encryption, compression, filtering, message logging, QoS, etc.

4 Evaluation

This section is devoted to M3-IPC performance evaluation against other IPC mechanisms. The compliance of the other design goals was verified during and after the development stage using several testing scenarios.

Two types of communication tests are presented in Fig. 2: 1) between co-located processes; 2) between processes located on different nodes. Furthermore, two types of micro-benchmarks were performed on each one: a) message transfer; b) data copy. The following common communication pattern was used: a server waits to receive a message from a client, and immediately replies. Once the client sends the request to the server, it waits for the reply. The reply may be a message (36 bytes) or a block of data (36 bytes to 64 Kbytes).

Although several performance metrics were evaluated as latency, CPU usage, network usage, by space limitations, only the throughput results are presented here.

The tests were made on a cluster of 8 (eight) quad-core Intel(R) i5 CPU 650@3.20GHz with a memory bandwidth of 20,841 Gbytes/s for 64 Kbytes blocks (reported by *bm_mem*), linked by a 1 Gbps dedicated LAN switch.

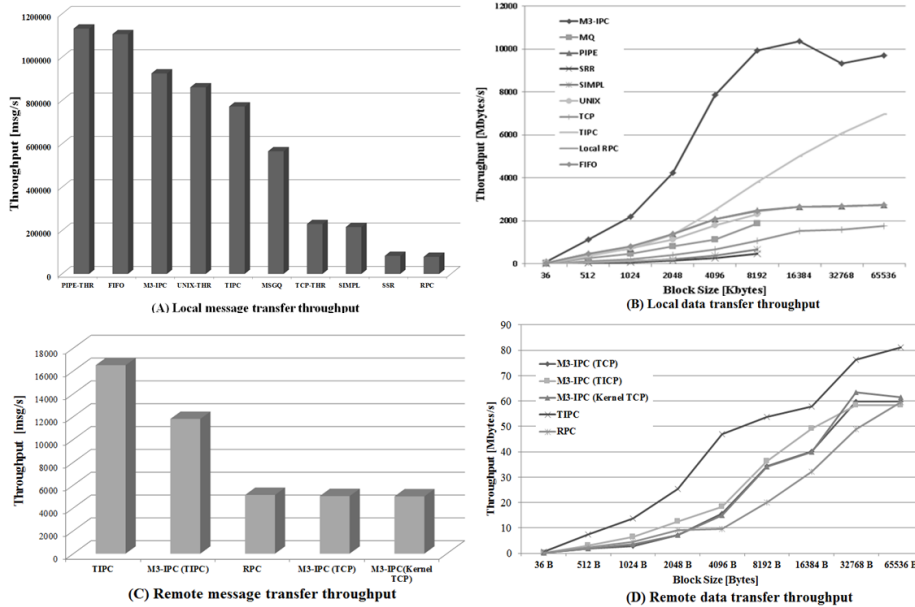


Fig. 2: Results of performance tests.

4.1 Tests between Co-located Processes

Tests between co-located processes allow the comparison of M3-IPC performance versus other IPC mechanisms available on Linux.

One of the design goals states that the expected performance should be as good as the fastest IPC mechanisms available on Linux. The following IPC mechanisms were tested using custom and [22, 23] provided micro-benchmarks: *Message Queues*, *RPC*, *TIPC*, *FIFOs*, *pipes*, *Unix Sockets*, *TCP Sockets*, *SRR*, *SIMPL*.

4.1.1 Message Transfer Performance

The presented results (Fig. 2-A) summarize the best throughput achieved by the IPC mechanisms running a single pair of client/server processes. Linux IPC mechanisms with the highest performance were pipes and named pipes (or FIFOs) followed by M3-IPC (925,314 [msg/s]).

Another micro-benchmark of message transfers between multiple pairs of Client/Server processes was run to evaluate performance in concurrency. The highest average throughput was 1,753,206 [msg/s], which was reached with 4 pairs of Client/Server processes (4 cores).

4.1.2 Data Copy Performance

As it is shown in Fig. 2-B, M3-IPC performance surpasses other IPC mechanisms on Linux. The reasons of this behavior are: 1) M3-IPC performs a single copy of data between address spaces while the others perform at least two copies (Source to Kernel, Kernel to Destination); 2) it requires a lower number of context switches; 3) it uses the Linux kernel provided *page_copy()* function which uses MMX instructions.

4.2 Tests between Processes Located on Different Nodes.

This section presents performance results of M3-IPC against RPC and TIPC.

M3-IPC does not consider flow control, error control, or congestion control. Those issues are delegated to proxies and the protocol that they use. Reference implementations of M3-IPC proxies use TCP and TIPC as transport protocols.

4.2.1 Message Transfer Performance

As it can be seen in Fig. 2-C, TIPC has the highest throughput. M3-IPC using TCP on proxies has throughput similar to RPC.

The remarkable performance of TIPC suggested that it could be a good option to be used by M3-IPC proxies as transport protocol. M3-IPC versatility and flexibility in proxy programming allowed authors to modify the source code of proxies in a few minutes so as to use TIPC instead of TCP. These changes result in an improvement of performance, emphasizing the impact that the transport protocol has on its throughput.

4.2.2 Data Copy Performance

As shown in Fig. 2-D, TIPC presents the highest throughput of 81[MB/s] that confirms results presented in [16]. The highest throughput achieved by RPC and M3-IPC was about 60[MB/s]. Fig.2-D also shows that there is no noticeable difference in performance when using TIPC instead of TCP as transport protocol on M3-IPC proxies to copy data blocks.

5 Conclusions and Future Works

IaaS providers always need higher levels of performance, scalability and availability for their virtualization services. These requirements can be met by a distributed virtualization technology. As a proof of concept, a DVS prototype using M3-IPC was developed.

M3-IPC addresses some issues about thread support, location transparency, message redirection on process migration, network-transport protocol agnostic, and IPC confinement for virtualization. The results show that M3-IPC achieves all its performance related design goals, with a high throughput for both intra-node and inter-node messages and data transfers.

To improve M3-IPC security and isolation issues, its integration to Linux Capabilities and *cgroups* [24] are being considered for further research and future works.

Until the development of M3-IPC, to the best of authors' knowledge, there was not a modular IPC software for Linux that fully satisfied the particular communication needs

of a DVS. This suggests that many questions remain about software for programming communications related to virtualization.

References

1. D. Hall, D. Scherrer, J. Sventek: A Virtual Operating System. Journal Communication of the ACM, 1980
2. David R. Cheriton; Willy Zwaenepoel: The Distributed V Kernel and its Performance for Diskless Workstations. Proceedings of the 9th ACM Symposium on Operating Systems Principles (1983)
3. Morin, C., et al.: Kerrighed and data parallelism: cluster computing on single system image operating systems. IEEE Computer Society (2004)
4. Barak A.; La'adan O; Shiloh A.: Scalable Cluster Computing with MOSIX for Linux. Proc. 5th Annual Linux Expo Raleigh (1999)
5. Pfister, Gregory F.: In Search of Clusters. Prentice Hall 1998, ISBN 0-13-899709-8.
6. Navin Budhiraja, Keith Marzullo, Fred B. Schneider, and Sam Toueg: The primary-backup approach. In Distributed systems (2nd Ed.), Sape Mullender (Ed.). ACM Press/Addison-Wesley Publishing Co., New York, NY, USA 199-216, 1993
7. Pablo Pessolani, Toni Cortes, Silvio Gonnet, Fernando G. Tinetti: Sistema de Virtualización con Recursos Distribuidos. (Spanish) WICC 2012. Argentina, 2012
8. J. Turnbull: The Docker Book. ISBN 978-0-9888202-0-3. 2014
9. Hindman, Benjamin, et al.: Mesos: A Platform for Fine-Grained Resource Sharing in the Data Center. In: NSDI. 11: 22-22. Retrieved 12 January 2015.
10. Andrew S. Tanenbaum, Gregory J. Sharp and De Boelelaan A: The Amoeba Distributed Operating System. 1992
11. Barrera, J.: A Fast Mach Network IPC Implementation. In Proceedings of the USENIX Mach Symposium, November 1991.
12. John Collins and Robert Findlay: Programming the SIMPL Way. ISBN 0557012708, 2008.
13. SRR- QNX API compatible message passing for Linux. <http://www.opcdatahub.com/Docs/booksr.html>
14. M. Sharifi, K. Karimi: DIPC: A Heterogeneous Distributed Programming System. In Proceedings of the 3rd Annual Computer Conference of the Computer Society of Iran, 1997
15. Andrew D. Birrell and Bruce Jay Nelson: Implementing remote procedure calls. ACM Transactions on Computer Systems, 1984
16. Jon Paul Maloy: TIPC: Providing Communication for Linux Clusters. Proceedings of the Linux Symposium, 2004
17. Brian N. Bershad, Thomas E. Anderson, Edward D. Lazowska, Henry M. Levy: User-Level Interprocess Communication for Shared Memory Multiprocessors. 1991
18. Dejan S. Milojevic, Fred Douglass, Yves Paindaveine, Richard Wheeler, and Songnian Zhou. "Process migration". ACM Computer Survey 32, 3, September 2000
19. Tanenbaum A., Woodhull A.: Operating Systems Design and Implementation. Third Edition Prentice-Hall. 2006
20. Ryan Slominski: Fast User/Kernel Data Transfer; Master Thesis. April 20. 2007
21. Luca Veraldi: Efficient Capability-Based Messaging (ECBM). 2003 (Italian)
22. Michael Kerrisk: The Linux Programming Interface; No Starch Press, ISBN 978-1-59327-220-3, 2010
23. ipc-bench: <http://www.cl.cam.ac.uk/research/srg/netos/ipc-bench/>.
24. CGROUPS: <https://www.kernel.org/doc/Documentation/cgroup-v1/cgroups.txt>

Caracterización de Tráfico-Distribución de Johnson SB

Luis Marrone

Calle 50 y 120, 2do piso, La Plata, Buenos Aires, Argentina
Laboratorio de Investigación en Nuevas Tecnologías Informáticas
Facultad de Informática-Universidad Nacional de La Plata
lmarrone@linti.unlp.edu.ar

Resumen. Obtener el modelo de tráfico resulta crucial a la hora de evaluar la performance de una red como así también disponer del mismo en la etapa de desarrollo e implementación de la misma. El punto de partida para obtener el modelo es contar con la caracterización del tráfico a cursar por la red. Caracterizar el tráfico redundante en obtener la distribución que mejor se corresponda con el mismo (“fitting distribution”). En este trabajo presentamos una distribución, “Johnson SB” normalmente no utilizada en la construcción de estos modelos pero que como veremos presenta resultados promisorios.

Palabras clave: Redes. Tráfico. Modelos. Distribuciones. Validación

1. Introducción

Caracterizar el tráfico de una red es proveer la descripción completa de los elementos que lo definen como ser entre otros, el tiempo de arribo entre mensajes, su longitud y el tiempo de servicio. Para disponer de esos elementos como paso previo a su descripción es que tomamos muestras. Muestras que se corresponden a variables aleatorias por lo cual no nos queda otro remedio que hacer un análisis estadístico de las mismas. La estadística descriptiva es la disciplina que nos otorga ese análisis en una primera etapa, dándonos valores del promedio, varianza, desviación estándar, curtosis y asimetría. Estas propiedades de las muestras tomadas nos ayudan a elegir las distribuciones más adecuadas que se correspondan con las muestras. El promedio y la desviación estándar junto con la varianza están relacionados en cuanto al aporte dado que valores de la desviación estándar cercanos al promedio poco nos pueden aportar para la elección de la distribución, dada la dispersión que presentan las muestras. Tal vez un mayor aporte lo representan la asimetría y kurtosis aunque normalmente son poco empleadas en este tipo de análisis. En esta primera etapa también se suele acudir a representaciones gráficas de las variables, en particular el histograma que resulta valioso por cuanto permite acotar rápidamente un conjunto posible de distribuciones adecuadas. Obteniendo así como resultado de la primera etapa un conjunto de distribuciones posibles pasamos a la segunda etapa donde parametrizamos las distribuciones que elegimos en la primera y comparamos su comportamiento con el de las muestras tomadas. Esta etapa normalmente se conoce como adaptación/ajuste/sintonía de la distribución a las muestras (“fitting distribution”). Esta

sintonía se puede realizar en forma gráfica y/o analítica, de hecho se complementan. Los métodos gráficos tradicionales son el de Q-Q-Plot y P-P-Plot. Q-Q-Plot es un gráfico que compara dos distribuciones de probabilidades, la empírica obtenida a partir de las muestras y la teórica a partir de las distribuciones elegidas como las más adecuadas en la primer etapa y cuyos parámetros fueron estimados a partir de las muestras. Específicamente se grafican los percentiles en Q-Q plot donde la mayor correspondencia ente ambas distribuciones se dará si la gráfica resultantes es la recta $y=x$. Para el caso de P-P Plot se grafican ambas funciones de distribución acumulativas donde la mayor correspondencia se dará si la gráfica resultante se da con una pendiente d 45° . El método analítico comprende un conjunto de tests bajo el nombre de Bondad de Ajuste. Los tests comúnmente empleados son el de Kolmogorov-Smirnov, Anderson-Darling y Chi-Cuadrado.

En lo que sigue (Sección 2) detallamos elementos de la estadística descriptiva mencionados anteriormente junto con la distribución de Johnson SB [1, 2, 3], comúnmente no empleada en el análisis de tráfico. En la Sección 3 detallamos el escenario en el que se tomaron las muestras y la metodología empleada para caracterizarlo, continuando (Sección 4) con los resultados y conclusiones (Sección 5)

2. Elementos de Estadística

Detallamos aquí elementos de estadística descriptiva que colaboran eficazmente a la tarea de elegir posibles distribuciones que se correspondan con la realidad.

2.1. Curtosis[4]

La curtosis determina el grado de concentración/amplitud de pico que presentan las muestras en la región central de la distribución, comparada contra una normal. Así puede ser:

Leptocúrtica.- Existe una gran concentración. Supera a la normal, valores positivos

Mesocúrtica.- Existe una concentración normal. Valor cero

Platicúrtica.- Existe una baja concentración. Está por debajo de una normal.

Valores negativos.

Para una muestra de n valores el coeficiente de curtosis está dado por:

$$g_2 = \frac{m_4}{m_2^2} - 3 = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^2} - 3$$

(1)

Los m_i corresponden a los momentos de orden i. A la relación de momentos se le resta 3 por cuanto es el valor que corresponde a una distribución normal.

2.2. Asimetría (“skewness”)

Nos da una medida de la inclinación hacia derecha (>0) o izquierda (<0) de una distribución comparada con la normal.

Para una muestra de n valores la asimetría está dada por:

$$g_1 = \frac{m_3}{m_2^{3/2}} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{3/2}}$$

(2)

2.3. Distribución de Johnson SB

En realidad constituyen un conjunto de cuatro distribuciones, “SL” o log-normal, ”SU” sin límites, “SB”, acotada y “SN”, el caso particular de una normal. Creada por Norman Johnson en 1940 quien la creó con el objeto de aplicar los métodos y teoría de la distribución normal a un amplio rango de distribuciones no normales a través de transformaciones computables a partir de distribuciones como la exponencial y seno hiperbólico. Dada su flexibilidad esta familia de distribuciones se emplea en varios campos como el de química atmosférica [5], ingeniería biomédica [6], economía [7], gerenciamiento [8], ciencia de los materiales [9] y análisis forestal [10], [11].

La distribución se define como:

$$f(x) = \frac{\delta}{\lambda \sqrt{2\pi z(1-z)}} e^{-\frac{1}{2}(\gamma + \delta \ln(\frac{z}{1-z}))^2}, \quad z \equiv \frac{x - \zeta}{\lambda}$$

(3)

Donde:

γ : Factor de forma

δ : Factor de forma ($\delta > 0$)

λ : Factor de escala ($\lambda > 0$)

ζ : Factor de Localización

No tenemos registro de su empleo en el análisis de tráfico de redes de datos.

3. Escenario y metodología empleados

El escenario de prueba fue extremadamente simple por cuanto se quiso ver la factibilidad del empleo de la distribución de Johnson SB en redes de datos. Se realizó la captura con Wireshark 2.0.3 de sesiones de https entre una PC del laboratorio con el servidor del proyecto Gutenberg (<http://www.gutenberg.org/ebooks/author/85>)

De las muestras obtenidas se tomó el tiempo de servicio de cada segmento de datos recibido. Ese parámetro va a ser la componente de tráfico a caracterizar.

Aisladas entonces las muestras de ese tiempo se procedió a estimar los parámetros de distribuciones Normal, Beta y Johnson a los efectos de comparación con la nueva distribución propuesta.

Dados los estimadores se realizaron los tests de Kolmogorov y Anderson Darling para la Bondad de Ajuste.

Los cálculos estadísticos fueron realizados con Matlab R2017a y la Bondad de Ajuste con Johnson Curve Toolbox for Matlab [12]

4. Resultados

Presentamos los resultados producto del ajuste de las muestras a las distribuciones Normal, Beta y JohnsonSB

4.1. Distribución Normal

Estimamos los parámetros con la funcionalidad provista por Matlab

```
>>pd = fitdist(A, 'normal')
```

Siendo A el archivo de muestras. La estimación resultó

Tabla 1. Estimación de parámetros para una distribución normal

μ	σ
782.031 [769.863, 794.199]	697.628 [689.13, 706.34]

A continuación graficamos la distribución Normal(curva) comparada con el histograma(barras), Figura 1. La función de distribución acumulativa vs. La empírica, Figura2. El Q-Q Plot, y P-P Plot Figura 3,.

De los gráficos se desprende que asumir una distribución Normal para las muestras sería estar muy lejos de la realidad.

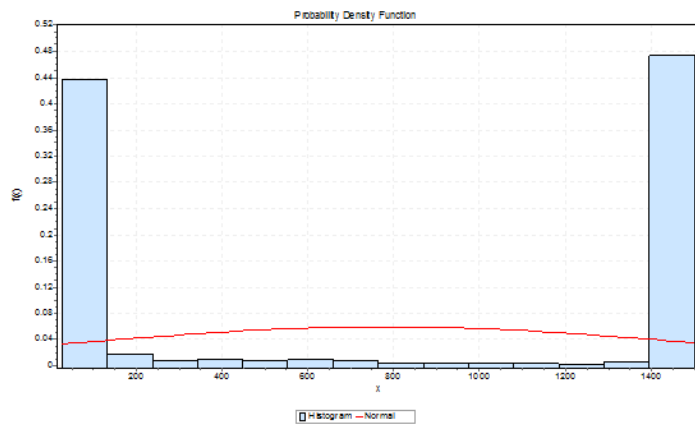


Fig. 1. Histograma y distribución Normal

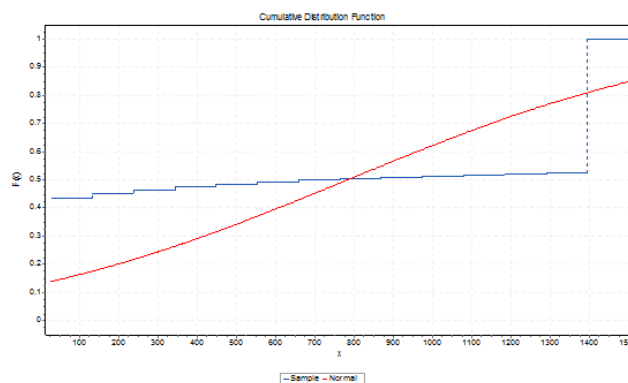


Fig. 2. Función de distribución acumulativa vs. Empírica

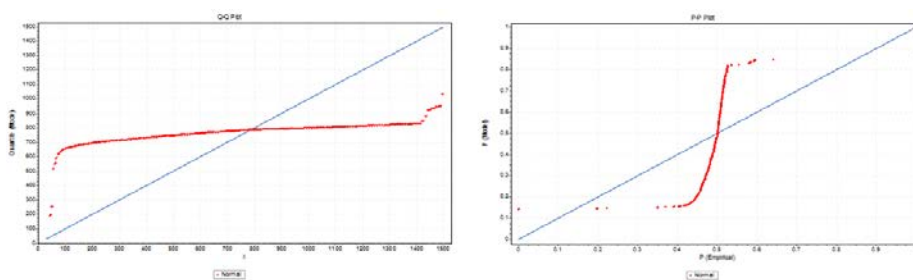


Fig. 3. Q-Q Plot y P-P Plot Muestras vs. Normal

En la Tabla 2 vemos los resultados de la Bondad de Ajuste

Tabla 2. Bondad de Ajuste para distribución Normal

Kolmogorov-Smirnov					
Sample Size	12630				
Statistic	0.29269				
P-Value	0				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	0.0955	0.01088	0.01208	0.01351	0.0145
Anderson-Darling					
Sample Size	12630				
Statistic	1828.2				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	1.3749	1.9286	2.5018	3.2892	3.9074
Chi-Squared					
Deg. of freedom	13				
Statistic	14714.0				
P-Value	0				
Rank	24				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	16.985	19.812	22.362	25.472	27.688

4.2. Distribución Beta

Estimación de parámetros

Tabla 3. Estimación de parámetros distribución Beta

α_1	α_2	a	b
0.0707	0.05544	26.769	1500.0

Presentamos las mismas figuras que para el caso de la distribución Normal para terminar con la Bondad de Ajuste correspondiente.

Del análisis resulta que esta distribución resulta más realista, los puntos del Q-Q-Plot se aproximan a la recta de 45°.

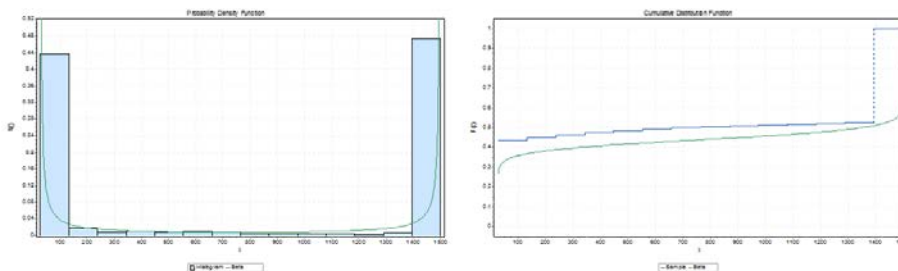


Fig. 4. Histograma- Dist.Beta(izq.) y Distribución Beta acumulativa-Empírica(der.)

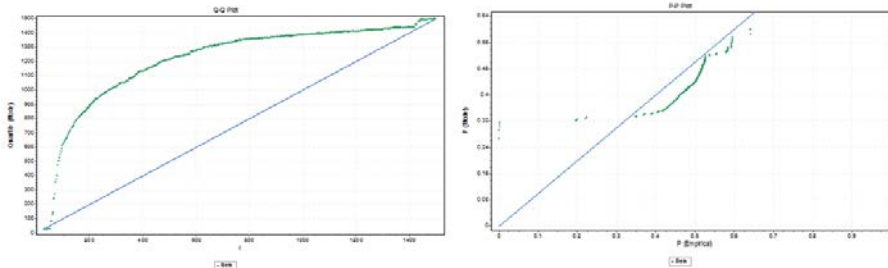


Fig. 5. Q-Q-Plot(izq.) y P-P Plot(der.) para la distribución Beta

Tabla 4. Bondad de Ajuste para distribución Beta

Beta					
Kolmogorov-Smirnov					
Sample Size	12630				
Statistic	0.31512				
P-Value	0				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	0.0095	0.0108	0.0120	0.0135	0.014
	5	8	8	1	5
Anderson-Darling					
Sample Size	12630				
Statistic	13743.0				

α	0.2	0.1	0.05	0.02	0.01
Critical Value	1.3749	1.9286	2.5018	3.2892	3.9074

4.3. Distribución Johnson SB

La estimación resultó:

Tabla 5. Estimación de parámetros distribución

γ	δ	λ	ζ
-0.01187	0.04707	1450.7	49.841

Las figuras y tabla de ajuste resultantes indican la mejor correspondencia de las muestras con esta distribución.

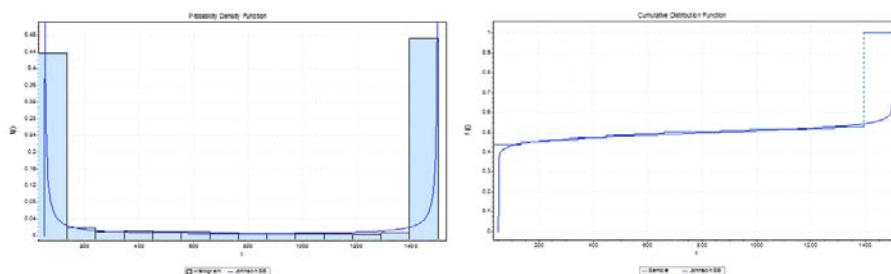


Fig.6. Histograma-Distribución(izq.),CDF-Empírica (der.) JohnsonSB

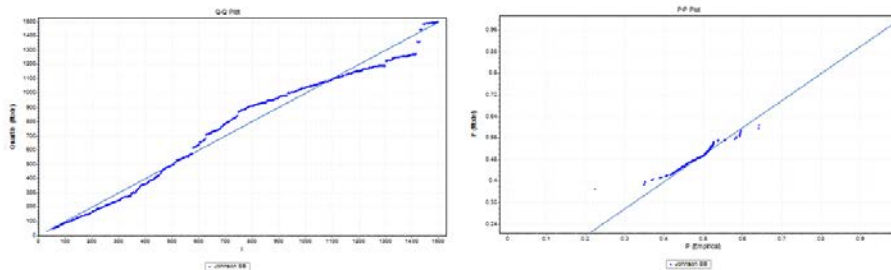


Fig. 7. Q-Q-Plot(izq.) y P-P Plot(der.) para la distribución JohnsonSB

Tabla 6. Bondad de Ajuste para distribución JohnsonSB

Johnson SB					
Kolmogorov-Smirnov					
Sample Size	12630				
Statistic	0.35916				
P-Value	0				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	0.009 55	0.010 88	0.012 08	0.013 51	0.014 5
Anderson-Darling					
Sample Size	12630				
Statistic	9358.0				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	1.374 9	1.928 6	2.501 8	3.289 2	3.907 4

5. Conclusiones

En el presente trabajo hemos puesto en evidencia el recorrido necesario a cumplimentar para poder obtener una caracterización de tráfico que permita construir un modelo válido del mismo. Entendemos por modelo válido aquel que nos permita diagnosticar problemas de performance en la red y/o aquel que nos permita predecir comportamiento y definir recursos necesarios en la implementación de una red de datos. Particularmente hemos apuntado a una etapa crítica del recorrido que es la elección de la distribución de la variable aleatoria objeto de análisis y que necesitamos definir para construir el modelo. Finalmente resaltamos la potencialidad de la distribución de JohnsonSB como elección válida. Sin duda no es una elección de carácter general, (con el tráfico de datos no es dable esa generalidad), pero si a tener en cuenta cuando tenemos una variable con dos valores de mayor recurrencia como fue el caso analizado. Somos concientes de que el trabajo no termina en esa elección. Sobre todo desde hace ya algo más de una década el tráfico ha cambiado su comportamiento, presentando un

caracter autosimilar del cual no hemos abordado su tratamiento por cuanto está fuera de los objetivos planteados del trabajo.

6. Referencias

- [1] Hill, I.D., Hill, R., and Holder, R.L. (1976). Fitting Johnson curves by moments. *Applied Statistics*. AS99.
- [2] Johnson, N.L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika*, **36**. 149-176.
- [3] Wheeler, R.E. (1980). Quantile estimators of Johnson curve parameters. *Biometrika*. **67-3** 725-728
- [4] Karl Pearson (1905) Das Fehlergesetz und seine Verallgemeinerungen durch Fechner und Pearson. A Rejoinder, *Biometrika*, **4**, 169-212,
- [5] Y.-N. Lee, X. Zhou, et. al (1998), Atmospheric chemistry and distribution of formaldehyde and several multioxygenated carbonyl compounds during the 1995 Nashville/Middle Tennessee Ozone Study, *Journal of Geophysical Research*, Vol 103, Issue D17:22449-22462.
- [6] F. George, K.M.Ramachandran (2009), *Analysis of Microarray Data for Gene Selection*, 25th Southern Biomedical Engineering Conference 2009; 15-17 May:237-238
- [7] Lu, Y., O. A. Ramirez, R. M. Rejesus, T. O. Knight, and B. J. Sherrick. 2008. Empirically evaluating the flexibility of the Johnson family of distributions: a crop insurance application. *Agricultural & Resource Economics Review* 37(1): 79-91.
- [8] Alexopoulos, C., D. Goldsman, J. Fontanesi, D. Kopald, and J. R. Wilson. (2008). Modeling patient arrivals in community clinics. *Omega* 36: 33-43
- [9] Edward Prince (2012). *Mathematical Techniques in Crystallography and Material Science*, Springer Science & Business Media.
- [10] Fonseca, T.F., Marques, C.P., Parresol, B.R.(2009).: *Describing maritime pine diameter distributions with Johnson's S B distribution using a new all-parameter recovery approach*. For. Sci. 55(4), 367–373 (2009)
- [11] Ayana Mateus, Margarida Tomé (2013). *Fitting Johnson's SB Distribution to Forest Tree Diameter*. Advances in Regression, Survival Analysis, Extreme Values, Markov Processes and Other Statistical Applications pp 289-296. Springer
- [12] Jones, D. L. (2014). *Johnson Curve Toolbox for Matlab: analysis of non-normal data using the Johnson family of distributions*. College of Marine Science, University of South Florida, St. Petersburg, Florida, USA

Consideraciones sobre el comportamiento del protocolo TCP en sus variantes Vegas, Reno, Cubic y Westwood ante errores en ráfaga en una topología híbrida

Diego R. Rodríguez Herlein, Carlos A. Talay, Claudia N. González y Franco A. Trinidad
Licenciatura en informática, UNPA-UARG
Río Gallegos, Argentina
Campus universitario - Oficina B 18
diegorh@gmail.com, carlostalay@yahoo.com.ar, claudiagonzález@yahoo.com, talejandro.franco@gmail.com

Luis A. Marrone
L.I.N.T.I. – Universidad Nacional de La Plata
La Plata, Argentina
Calle 50 y 120 – 2do. Piso – Edificio Bosque Oeste
lmarrone@linti.unlp.edu.ar

Abstract. Las comunicaciones que utilizan el protocolo TCP han sido y son ampliamente estudiadas debido a la heterogeneidad de los medios y condiciones en las que tienen que desempeñarse. Esto ha dado lugar a diferentes variantes de este protocolo, que tienen que afrontar distintos tipos de escenarios con sus propias características, tales como errores en ráfaga. Estos problemas, entre otros, son característicos de los enlaces inalámbricos y tienen como consecuencia una merma en el rendimiento de las comunicaciones. Ante este inconveniente, es interesante ver cómo el protocolo recupera el funcionamiento normal después de sufrir este evento. Por lo tanto, este documento tiene como objetivo realizar un análisis de cómo las transmisiones de datos son afectadas al sufrir errores en ráfaga, cuando se utiliza alguna de las 4 variantes estudiadas, mediante simulaciones con el ns-2

Keywords: TCP, error en ráfaga, Performance, ns-2

1 Introducción

El protocolo de comunicaciones TCP (Transmission Control Protocol) [1] ha sido ampliamente utilizado en las redes de datos desde su desarrollo hasta la actualidad. Introduciéndole modificaciones se trató de adaptarlo a distintas condiciones de trabajo, permitiéndole así acompañar a las innovaciones tecnológicas que se han desarrollado en el área de las telecomunicaciones y que utilizan tanto medios

cableados como inalámbricos. Este protocolo se caracteriza por ser confiable, realizar un control de flujo y poseer un mecanismo de control de congestión de datos. También regula la secuencia de entrega de los paquetes, mediante la verificación de la recepción ordenada de dichos paquetes numerados en origen, corroborado el orden de llegada en el receptor. Este protocolo ofrece un servicio orientado a conexión, que basa su entrega confiable en un procedimiento conocido como ARQ (Automatic Repeat reQuest), en sus distintas variantes [2], que garantiza la integridad de los datos. Mediante el procedimiento ARQ y con la utilización de ACKs (acknowledgments) positivos, se logra que, con menos de un ACK por paquete de datos, se pueda confirmar la recepción de información de todo un conjunto de paquetes. Esta técnica se conoce como delayed-ACK [3] y permite lograr un importante aumento de eficiencia en el funcionamiento de la red.

A nivel de control de congestión, el protocolo TCP realiza una regulación del tráfico sobre el flujo de datos. Para lograr esto, el protocolo verifica si existe una pérdida de segmentos o si se produce una recepción de ACKs duplicados. Al analizar el resultado de esta verificación, el protocolo determina la ocurrencia de pérdida de paquetes y por consiguiente si existe o no congestión en la red [4] [5]. Mediante el perfeccionamiento de este método, se han desarrollado dos variantes para atender problemas de control de congestión. Una de ellas se basa en un control reactivo del problema, suponiendo que existe congestión en los enlaces ante la pérdida de segmentos [6] [7]. Por su parte, la otra variante trata de realizar un control proactivo de la congestión, en donde lo que se busca es desarrollar una estrategia que permita evitar que el tráfico llegue a una situación de congestión [8] [9].

En la actualidad, las tecnologías de transmisión apuntan a la calidad del flujo de datos, lo que posibilita tener una baja tasa de errores. En este escenario, las técnicas de control de congestión de la red se han basado fundamentalmente en la detección de paquetes perdidos. Por ello, bajo estas condiciones, los protocolos reactivos entienden que hay congestión en la red y accionan sus algoritmos de control de congestión. Sin embargo, existen situaciones en las cuales la pérdida de paquetes puede tener otro origen que no es la congestión y, por tanto, no deberían disparar sus mecanismos de control de congestión. Pensando en esta situación, se propone mediante la configuración de un modelo simple y utilizando la herramienta de simulación ns-2 [10] [11], analizar la respuesta de 4 protocolos ante la introducción de errores en ráfaga en una transmisión de datos.

2 Diseño del modelo de prueba

Para modelar y generar los datos del presente trabajo, se utilizó el ns-2 (Network Simulator 2), simulador de redes de eventos discretos en su versión 2.35 (released Nov. 4 2011). Con este simulador se implementó la siguiente topología de 3 nodos:

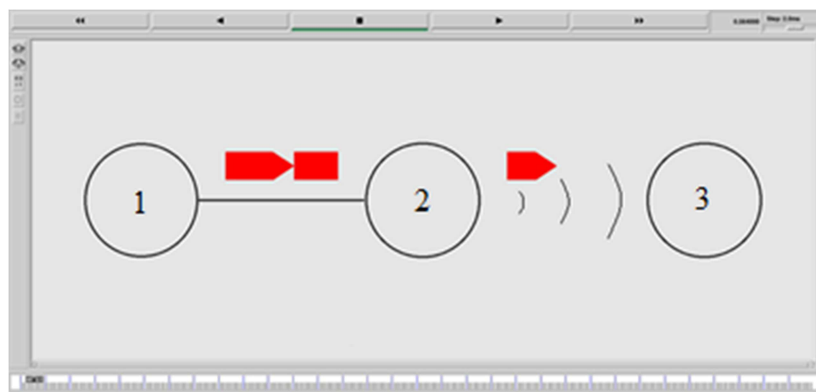


Grafico 1

Como se observa en el grafico 1, los nodos 1 y 2 están vinculados por un enlace cableado que se configuró como dúplex, con un ancho de banda 2 Mb/s, retardo de propagación 2 ms. y política de servicio de las colas DropTail. El enlace entre los nodos 2 y 3 es inalámbrico y se configuró como modo de propagación TwoRayGround, la capa física WirelessPhy, MAC 802.11, la antena OmniAntenna y el nodo inalámbrico sin movilidad.

La selección de este modelo es una aproximación a un escenario Wireless con un nodo fijo (nodo 1), una estación base (nodo 2) y un nodo móvil (nodo 3), con la simplificación práctica que el enlace inalámbrico no presenta desconexiones y solo tiene errores en forma de ráfagas.

En principio el modelo constaba de más nodos inalámbricos. Sin embargo, debido a que el presente trabajo se basó en el estudio de los errores en forma de ráfagas en medios híbridos no se consideró necesario, por el momento, aumentar la complejidad incorporando una mayor cantidad de nodos móviles.

El nodo 1 se configuró como emisor y en él un agente TCP, por otro lado el nodo 3 se configuró como receptor. A este enlace se asoció un tráfico FTP (file transfer protocol) como único tráfico.

Se realizaron simulaciones independientes sobre las implementaciones de las distintas variantes de TCP. Para cada una de ellas, se generaron simulaciones introduciendo distintas longitudes de los errores en ráfagas, con tamaños que fueron desde una prueba sin errores (0), pasando a pruebas con ráfagas de error de 5, 10, 15 y 20 paquetes. Los Agentes TCP que se utilizaron fueron Vegas, Cubic, Reno y Westwood, tal como están designados e implementados en esta versión de ns-2 (ver. 2.35), es decir sin modificación alguna. En el caso de TCP Vegas los valores de ALPHA=1 y BETA=3 son los que se utilizan por defecto.

La transmisión de datos comienza a los 5 segundos de iniciada la simulación y están condicionadas a la transmisión de 3.000 paquetes de TCP, de 1.000 bytes c/u, independientemente de la longitud de la ráfaga de error. Estas ráfagas siempre comenzaron después de transmitidos los primeros 999 paquetes, el ensayo concluyó al terminarse de transmitir los 3.000 paquetes TCP asociado al tráfico de FTP.

Se escribió el script TCL de manera que, al finalizar cada simulación, se obtuviera, además del archivo de traza, un archivo con el tamaño de la ventana de congestión del agente TCP en función del tiempo y el valor del número de secuencia en función del tiempo. A su vez, se utilizaron script AWK sobre el archivo de traza para obtener valores de throughput, retardo de extremo a extremo y ventana de congestión. Para procesar estos datos, se utilizó una planilla de cálculo a través de la cual se generaron los gráficos que aquí se presentan.

3 Resultados obtenidos

Para verificar el comportamiento de los protocolos ante la ocurrencia de ráfagas, se diseñaron pruebas sobre la topología definida en el punto anterior. Simulando una transferencia de datos, y sobre el envío de los 3.000 paquetes que conforman los datos, se genera un error en ráfaga, correspondiente a los casos de 5, 10, 15, 20 y 25 paquetes perdidos en ráfaga, que comienzan a perderse a partir del paquete 999. Bajo estas condiciones, tenemos los siguientes ensayos:

3.1 Pruebas para comportamiento del modelo sin existencias de errores

Se expone a continuación el comportamiento de los parámetros retardo extremo a extremo, throughput, número de secuencia y ventana de congestión para el caso de transmisión de datos sin existencia de errores.

3.1.1 Retardo extremo a extremo

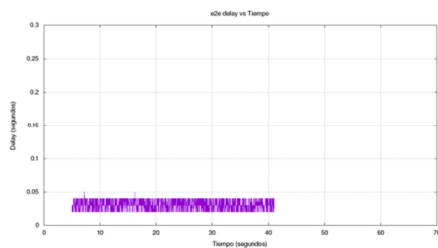


Fig. 1. VEGAS

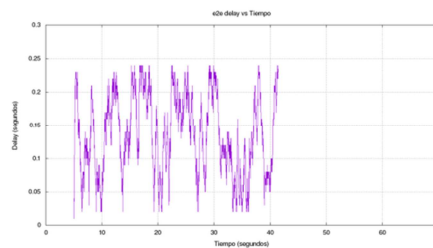


Fig. 2. CUBIC

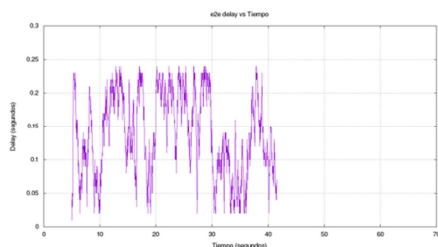


Fig. 3. RENO

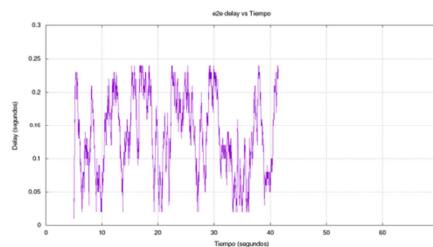


Fig. 4. WESTWOOD

3.1.2 Throughput

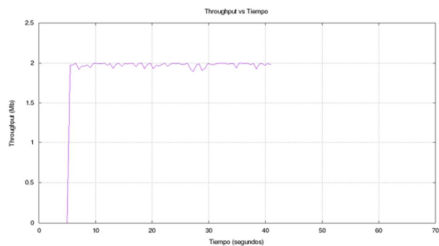


Fig. 5. VEGAS

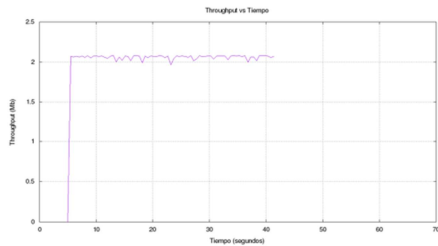


Fig. 6. CUBIC

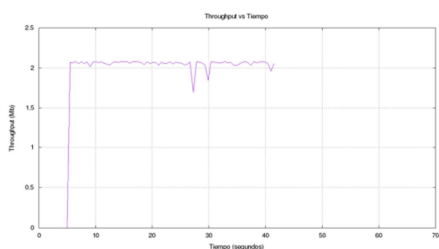


Fig. 7. RENO

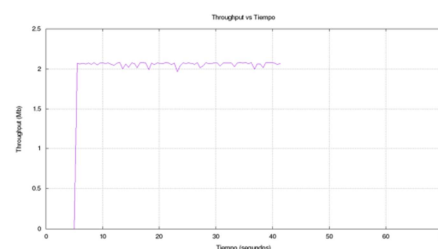


Fig. 8. WESTWOOD

3.1.3 Número de secuencia

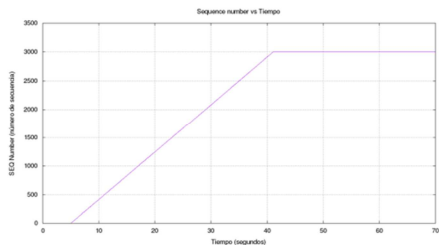


Fig. 9. VEGAS

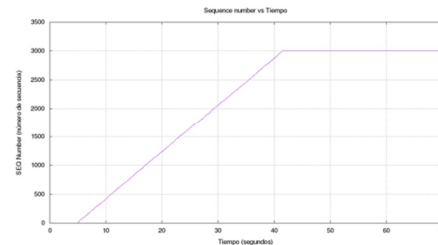


Fig. 10. CUBIC

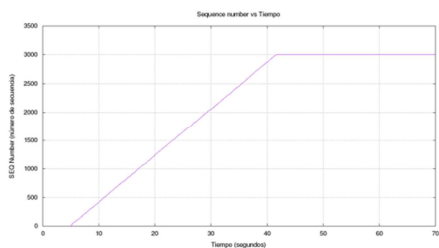


Fig. 11. RENO

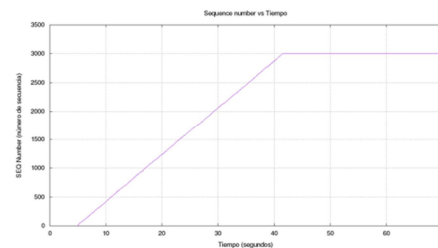


Fig. 12. WESTWOOD

3.1.4 Ventana de congestión (CW)

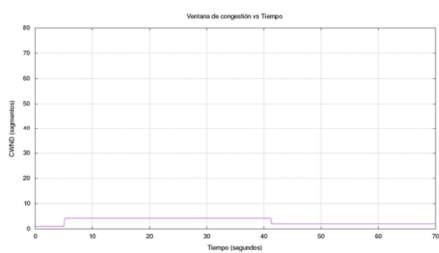


Fig. 13. VEGAS

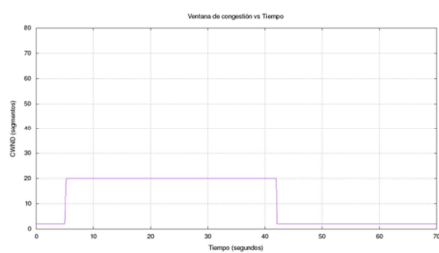


Fig. 14. CUBIC

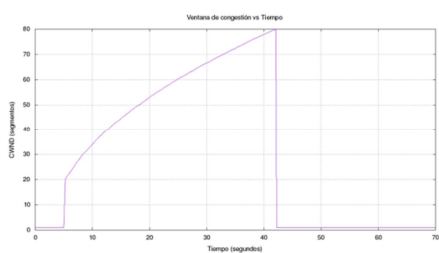


Fig. 15. RENO

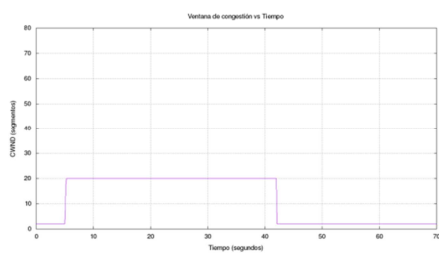


Fig. 16. WESTWOOD

3.2 Pruebas para comportamiento con una ráfaga de 15 paquetes

Se expone a continuación el comportamiento de los parámetros retardo extremo a extremo, throughput, número de secuencia y ventana de congestión introduciendo un error en ráfaga de 15 paquetes perdidos, a los efectos de comparar en comportamiento de los cuatro protocolos analizados.

3.2.1 Retardo extremo a extremo

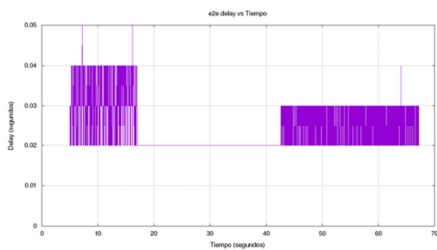


Fig. 17. VEGAS

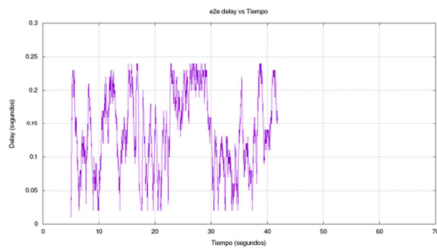


Fig. 18. CUBIC

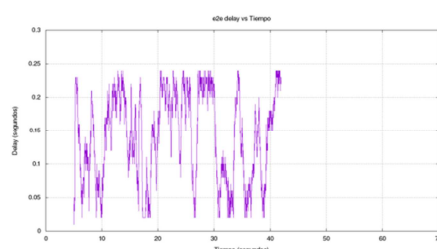


Fig. 19. RENO

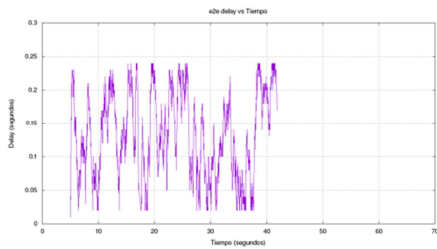


Fig. 20. WESTWOOD

3.2.2 Throughput

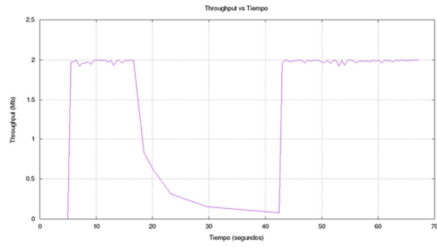


Fig. 21. VEGAS

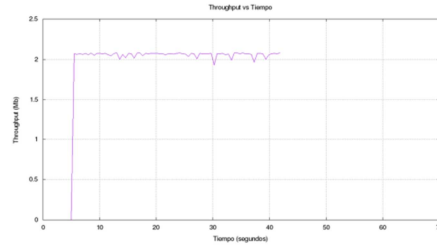


Fig. 22. CUBIC

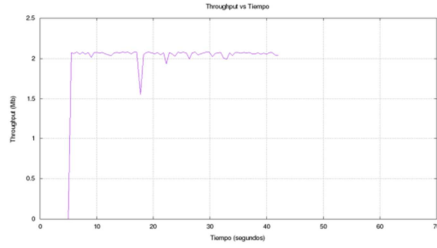


Fig. 23. RENO

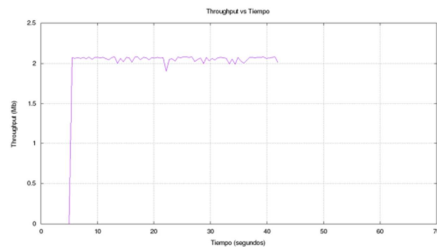


Fig. 24. WESTWOOD

3.2.3 Números de secuencia

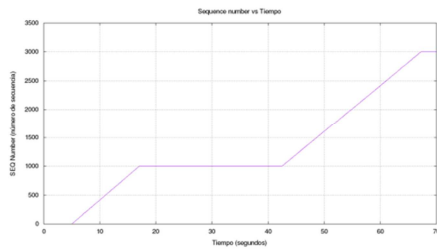


Fig. 25. VEGAS

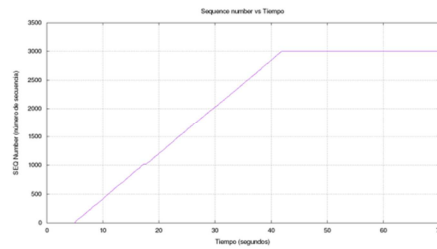


Fig. 26. CUBIC

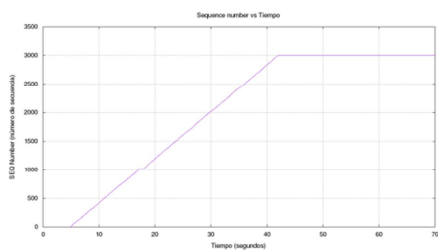


Fig. 27. RENO

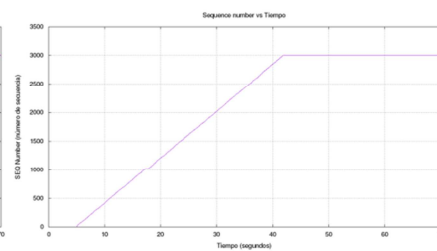


Fig. 28. WESTWOOD

3.2.4 Ventana de congestión (CW)

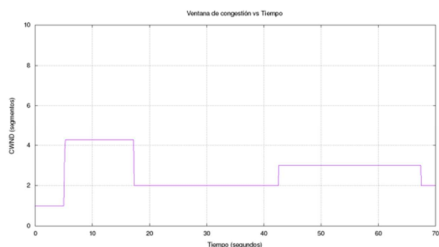


Fig. 29. VEGAS

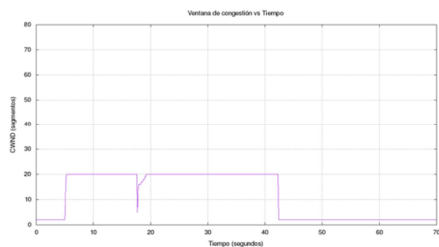


Fig. 30. CUBIC

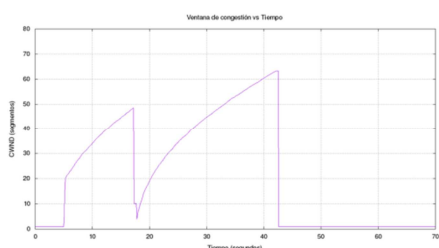


Fig. 31. RENO

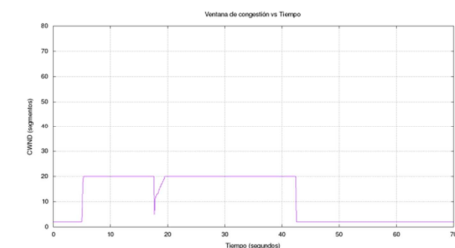


Fig. 32. WESTWOOD

3.3 Throughput de Vegas con distintos tamaños de ráfaga

Como se observa en la prueba de error en ráfaga de 15 paquetes, el agente TCP Vegas revela una sensibilidad mayor que los otros 3 protocolos, mostrando una demora en el reinicio de la transmisión. Es por ello que, a los efectos de apreciar cómo las ráfagas afectan de manera progresiva su comportamiento, recreamos pruebas de su parámetro throughput, para los valores de error en ráfaga correspondiente a 5, 10, 15, 20 y 25 paquetes y los mostramos a continuación:

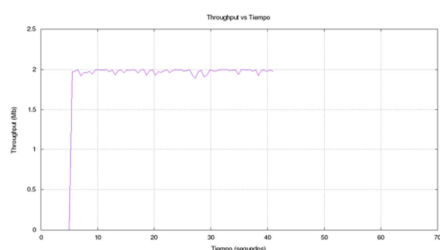


Fig. 33. Ráfaga de 0 paquetes

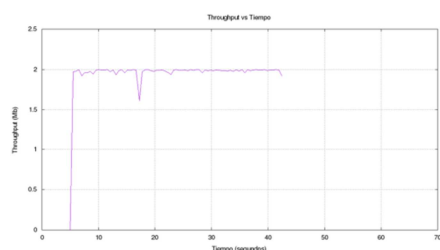


Fig. 34. Ráfaga de 5 paquetes

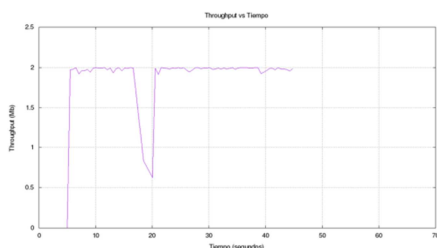


Fig. 35. Ráfaga de 10 paquetes

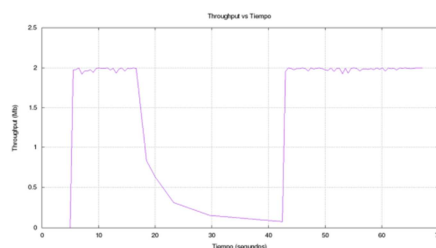


Fig. 36. Ráfaga de 15 paquetes

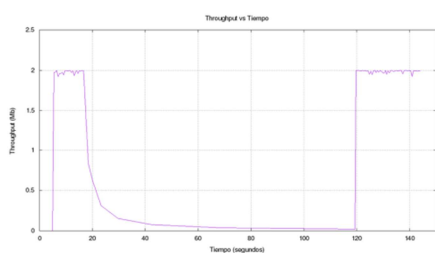


Fig. 37. Ráfaga de 20 paquetes

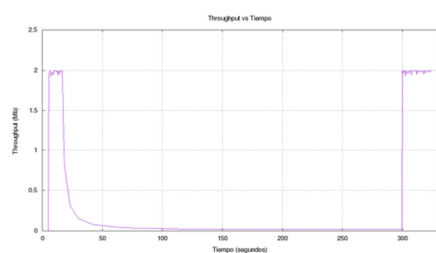


Fig. 38. Ráfaga de 25 paquetes

4 Conclusiones

Como comentábamos en la introducción, los distintos problemas que se plantean en las comunicaciones han sido solucionados en buena medida por los protocolos TCP, en sus diversas variantes, mediante el uso de distintas mejoras. Sin embargo, las variantes que solucionan determinados problemas terminan mostrando falencias bajo otras circunstancias. En este caso se ha analizado cómo los errores en ráfaga afectan a cuatro variantes características del protocolo TCP como son Vegas, Reno, Cubic y Westwood.

Observando los resultados obtenidos para todas las simulaciones, y comparando el comportamiento de los cuatro protocolos, es notoria la dificultad que presenta tempranamente el protocolo Vegas para recuperarse de un evento como es el error en ráfaga. Estas dificultades no se detectan en los demás protocolos, por lo menos en pruebas de bajos valores de errores en ráfaga (iguales o menores a 20 ráfagas). Puede observarse en las gráficas que, luego de producirse el evento, el protocolo reacciona disparando sus mecanismos de congestión, interpretando que una situación de este tipo se ha producido, cuando en realidad lo que se presenta es un problema de pérdida de paquetes por errores en ráfaga.

Debido a que se detectó este comportamiento por parte del protocolo Vegas en una ráfaga de 15 paquetes, se realizó un seriado de errores simulado distintos tamaños de ráfaga, comenzando con una simulación sin errores (0) y a continuación con 5, 10, 15, 20 y 25 paquetes perdidos en ráfagas, analizando su evolución a través del throughput (figuras 33 a 38). De esta manera se puede observar cómo el protocolo extiende progresivamente la demora en iniciar nuevamente la transmisión de los datos, resultando de esta manera sensible a este tipo de errores desde un primer momento.

5 Futuros Trabajos

En virtud de la reacción del agente Vegas ante este escenario, y para poder determinar más detalles en su comportamiento bajo condiciones similares, se propone realizar variaciones de los parámetros característicos alfa y beta de este protocolo a fin de verificar si es posible mejorar la respuesta ante los errores en ráfaga. Asimismo, se debe considerar un análisis más amplio que abarque una mayor cantidad de nodos inalámbricos y realizar ensayos con redes ad-hoc, utilizando para tal fin a la herramienta ns-3.

Referencias

- [1] Postel J., "RFC 793: Transmission Control Protocol", September 1981.
- [2] Stallings, William, "Comunicaciones y redes de computadoras". 6ta edición, Prentice Hall, 2000
- [3] David Clark, RFC 813: Window and Acknowledgment Strategy in TCP, Julio 1982
- [4] M. Allman, V. Paxson and W. Stevens, "RFC 2581: TCP Congestion Control", April 1999
- [5] M. Handley, J. Padhye and S. Floyd, "TCP Congestion Window Validation", RFC 2861, June 2000.
- [6] Allman M., Paxson V. and Blanton E., "TCP Congestion Control", IETF, Standards Track RFC 5681, Sept. 2009.
- [7] A. Afanasyev, N. Tilley, P. Reiher, and L. Kleinrock, "Host-to-Host Congestion Control for TCP," IEEE Communications Surveys Tutorials, vol. 12, no. 3, 3rd quarter 2010, pp. 304-340.
- [8] Stevens, W., "TCP slow start, congestion avoidance, fast retransmit, and fast re-recovery algorithms". RFC 2001, 1997.
- [9] V. Jacobson, "Congestion Avoidance and Control," ACM SIGCOMM Computer Communication Review, Vol. 25, No. 1, pp. 157-187, 1995.
- [10] Teerawat, Issariyakul & Ekram Hossain, "Introducción to Network Simulator NS2", Springer, 2009
- [11] The ns Manual, (formerly ns Notes and Documentation). A Collaboration between researchers at UC Berkeley, LBL, USC/ISI, and Xerox PARC. Kevin Fall & Kannan Varadhan, Editores, Nov 5, 2011 (ns-2.35)

Experimental study of performance and QoS of multicast and unicast video traffic in WAN links

Higinio Facchini¹, Santiago Pérez¹, Luis Marrone², Fabián Hidalgo¹, Adrian Cardenas¹

¹ CeReCoN – Departamento de Electrónica – Facultad Regional Mendoza – UTN
Rodríguez 273, Ciudad Mendoza
CP (M5502AJE) República Argentina
{higiniofac, santiagocp}@frm.utn.edu.ar

² LINTI – Facultad de Informática – Universidad Nacional de La Plata
50 y 120 – La Plata, Argentina
CP (B1900ASD) República Argentina

Abstract. In recent years there has been an exponential increase in the growth in multimedia applications, and in particular in video applications. In this work, a quantitative analysis is performed by experimentation, in order to evaluate the behavior and impact of video traffic on WAN networks. We propose a WAN test bed with four scenarios that allows to inject unicast or multicast video traffic, compressed with several codecs. From capturing video traffic, we identified several interesting performance and QoS metrics. We include detailed contributions on the impact produced by several factors, when there are restrictions on the bandwidth, as in a corporate real WAN link of some few Mbps. The work oriented on the configurations to be taken into account for an adequate management of similar networks and an efficient use of the resources available, without compromising the performance and the expected QoS.

Keywords: Multicast traffic, Video codec, WAN test bed M, QoS

1 Introduction

In recent years there has been an exponential increase in the growth of multimedia applications and, in particular, in video applications. Today you can receive digital TV with high definition and a greater supply of channels. There are Internet sites that offer movies on-line, and it is common for users to upload and download videos from sites web. It has also been widespread recording and share videos with mobile phones. The video call through the Internet is common. Large companies and organizations use video conferencing applications for face-to-face collaboration in different geographical regions. And even within the same company video traffic grows.

The growing number of multimedia users also involves a significant increase in bandwidth requirements. Consumers are becoming more demanding in terms of quality and the performance of the products based on video, and therefore, there is a strong incentive for continuous improvement in multimedia technologies. Understand

the structure of the frames of video and the requirements that they impose on the network, will help network administrators to improve the flow of traffic in the same. Aruba Networks [1] presents an interesting classification of the video traffic.

A number of standards have been developed for video encoding, such as H.261, H.263, MPEG1, MPEG2, MPEG4 and H.264. H.264/MPEG4 represents a great leap in video compression technology with, typically, a reduction of 50% of the average bit rate for a given video quality compared to MPEG-2, and approximately a reduction of 30% compared with MPEG-4 Part 2.

There are different jobs on the subjects considered, such as the analysis of multicast traffic and multicast routing protocols [2], analysis of video codecs [3], analysis of video traffic under different contexts [4], analysis of multicast video traffic [5], comparative analysis of multicast vs unicast [6], analysis of QoS for video traffic [7-9] and work on IPTV [10]. In this work was raised to obtain quantitative conclusions from experimental studies, on the best options of video traffic in laboratory networks WAN using actual equipment.

To do this, the main contributions of this paper are: (i) The demonstration, with detailed quantitative results, that: a) For multicast video traffic, all the metrics of performance and QoS obtained and analyzed respond to the expectations, and depend exclusively on the video codec used, and not of the topology of the network or of the number of clients receiving video streaming, and b) For the unicast video traffic, in addition, it was observed that the number of bytes transmitted and the final speed on each link if they depend on the topology of the network and the number of clients receiving video streaming; ii) The specification of a test bed WAN and four scenarios in experimental and of a comparative methodology to establish mechanisms between the different trials; y iii) The experimental determination of the behavior of each codec with regard to the statistical distributions of the interframes times.

The remainder of this paper is structured as follows: Section II provides an overview of the IP Multicast; section III presents the scenario and resources of experimentation; Section IV presents the results of the experimentation; and finally, in Section VII sets out the most important conclusions.

2 IP Multicast

For years, the suppliers of ISP services have been trying to use multicast technology to offer the service of high quality video through the interconnection of networks. They see the multicast capacity to scale easily and use the bandwidth effectively as an important tool, which could help them to control the loads of traffic, save bandwidth, offer new services, generate additional revenue and strengthen their long-term profitability.

Unfortunately, there are problems with the multicast deployment model in the interconnection of networks. In the first place, the multicast protocols need to be deployed in each router in the path between sources and receivers. Service providers have also found problems in providing reliable multicast services through the best efforts environment of the interconnected networks. Multicast solutions, in general, have lacked the bit rate required for adaptive streaming video transmission.

There is a widespread belief that you can't use the multicast technology to deliver video over interconnected networks. Certainly, the IP multicast promised to offer a number of benefits. Unlike unicast delivery, in which a server replicates a data stream for transmission to all client devices on the network that request it, the delivery of IP multicast invokes the server to transmit a single stream of data in the routing infrastructure. The routing infrastructure then replica effectively the data stream so that no router interface transport more than a single copy of the sequence, regardless of how many clients receive (Figure 1).

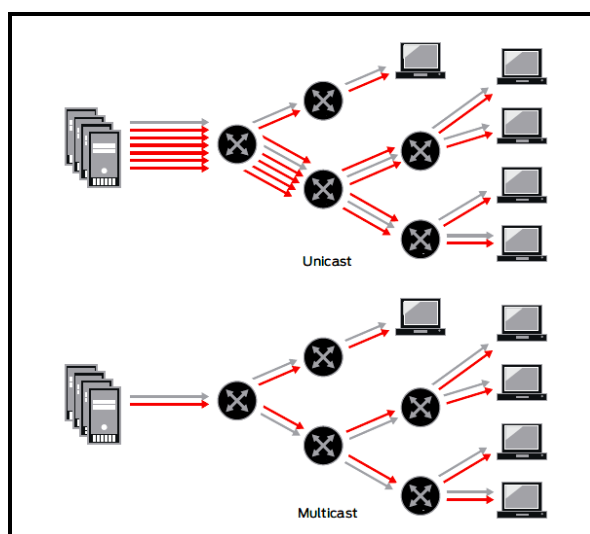


Fig. 1. Unicast and multicast

This difference is reflected in the preservation of bandwidth and data flows more efficient, which obviously adds up to a better network performance. But, in order to enjoy these benefits, you need to deploy IP multicast protocols on each router in the path. Therefore, the multicast IP address exists today mainly in LAN networks and in small areas of interconnected networks highly controlled. In fact, some companies have begun to offer customized solutions to try to resolve these limitations. The barriers that have so far prevented the widespread deployment of IP multicast are:

- Each routing device in the delivery route must support IP multicast, which means that if the IP multicast is omnipresent, each service provider must decide to offer multicast service.
- Is a great challenge to offer reliable multicast service in the best effort inherently environments that exist in the network today.
- Streaming video requires adaptive bit flow, what it lacks the multicast.
- The formats of video streaming of today, designed to use progressive download, mechanisms are not inherently capable of multicast.
- The interdomain multicast or between pairs of Internet providers ISP requires business agreements and complex network configurations.

In this experimental work has been quantified the impact and advantage on performance and QoS, of the use of the multicast video traffic, in environments of WAN interconnection, approaching the real situation of companies that link their LANs using WAN links of ISPs.

3 Scenario and experimental resources

In the present study we evaluated experimentally the performance of streaming video on a WAN test bed with the use of network computers, servers, and real customers.

The generic topology used includes a computer functioning as streaming server and up to 20 desktop computers (PCs) as clients connected at the ends of a network as if it were a WAN. This network is formed by WAN routers and switches with different types of links by linking to them. Figure 2 shows the general topology. In this topology, the links indicated with continuous line are of type FastEthernet with a transmission speed of 100 Mbps, while the links indicated by dashed line is serial interfaces connected in point-to-point format a transmission speed of 2Mbps. For the operation between routers R1 to R6 is configured the routing protocol OSPF v2. For the same routers are configured the routing protocol multicast PIM dense mode. The existing redundant links were raised as an option to simulate a WAN, but was used to configure the routing protocol for traffic between the server and each client always follow a single path.

For the topology is used the following hardware:

- Server: a desktop computer with the following features: CPU Intel Core i5 processor with 8GB of RAM and Windows 7 Professional 64 Bit.
- Clients: desktop computers with the following characteristics: AMD Athlon(tm) II X2 250 to 3GHz with 2GB of RAM and operating system Windows 7 Professional 32 Bit.
- Routers R1, R2, R3 and R4: Cisco model 2811, and Cisco routers R5 and R6 were resolved with switches Cisco WS-CS3750.
- Finally, for the connection of the routers with the PCs: Cisco switches Layer 2 Catalyst Model WS-2950-24.

And used the following software:

- Streaming server: The Unreal Media Server software, installed on the PC indicated how Video Server [11]. How to video player client software: the customer Unreal Media Player installed on all client computers. And for the capture Wireshark [12].

From the topology of experimentation, were derived 4 scenarios: is output from the server (a) is received in 3 clients (PC2-1, PC0-1 and PC3-1), (b) is received in 6 clients (PC1-1, PC2-1, PC0-1, PC0-2, PC3-1 and PC4-1), (c) is received in 9 clients (PC1-1, PC2-1, PC2-2, PC0-1, PC0-2, PC0-3, PC3-1, PC3-2 and PC4-1), and (d) is received on all clients. In all cases were injected with the video of [13].

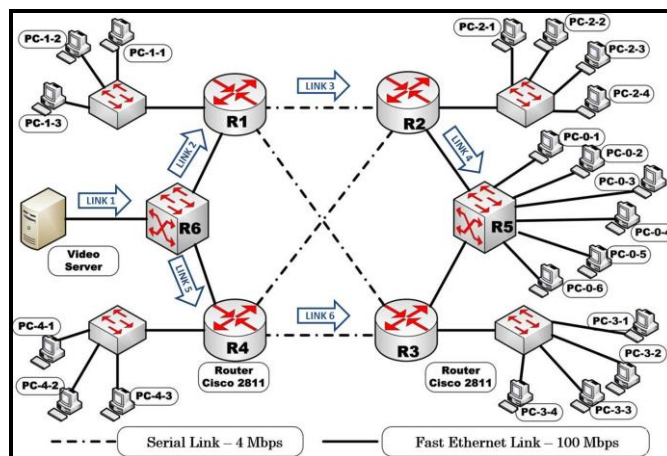


Fig. 2. Scenario of experimentation

4 Results of experimentation

4.1 Impact on bit rate of traffic

The PIM settings, routing and bandwidth based on the topology of work defined the paths for each of the traffic, shown in Figure 2, between the video server and remote stations.

This leaves in evidence a possible overload on the WAN links for the unicast traffic, which can compromise the bandwidth, performance and QoS, depending on the number of flows that pass through and the additional traffic to the video, typical of a business organization.

For the evaluation of the influence of the unicast traffic analyzes the bit rate, measured in each link individually for each scenario, and carried out a comparison between the same. In Figure 3 shows the bit rate for the scenario 4. This scenario implies an even greater overload on the WAN links for the unicast traffic, which can be compromised depending on the additional traffic to the video, typical of a business organization. In the link server output (1) there are 20 simultaneous flows, in the Link 2 there are 13 simultaneous flows, in the Link 3 there are 10 simultaneous flows, in the Link 4 there are 6 simultaneous flows, in the Link 5 there are 7 simultaneous flows and on the Link 6 there are 4 simultaneous flows. In this scenario it is clear that the overload, especially for the codecs MPEG-4 AVC and MPEG-2, is approaching the upper limit of capacity of the link 3.

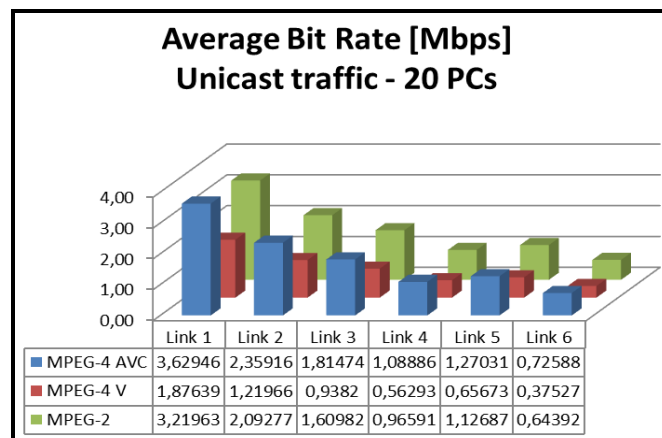


Fig. 3. Average rate per link for 20 PCs

4.2 Impact on the QoS

In addition, the transmission of video and multimedia over data networks facing specific problems in regard to the quality perceived by users. Several types of degradations usually occur in the video signals transmitted over packet networks.

The specific factors of IP networks that affect the quality of video, and on which we can draw conclusions from the tests carried out, are: packet loss, delay, jitter and transfer rate. In the tests carried out in the 4 scenarios, not detected packet loss, because we are using a network topology and configuration set to the application. According to the analyzed in Figure 4, the delays obtained in all cases do not exceed the 70 ms. In Figure 5, in the case of multicast the delay difference is not greater than 30 ms, while in the case of unicast, the jitter maximum is can generate some problems in video playback. In each video codec is identified the rate of transfer required, according to the resolution of the video. The bandwidth of each link used in the transfer must be sufficient for the transfer of all video streams participants, like all other traffic present. If this is not true, it may appear the above problems of packet loss and delays over the allowed. It was possible to observe the influence of the increase of occupation of the bandwidth of the links according to the amount of intervening for unicast traffic flows. For the multicast traffic, the increase of clients does not affect the bandwidth. In Figure 6 summarize this behavior, showing the total average transfer rate for each codec, in each of the scenarios. It is clear that the traffic is kept constant for multicast, regardless of the amount of clients; while, for the unicast traffic, at the exit of the server, the transfer rate increases proportionally to the number of clients connected, which can give an idea of the maximum amount of clients depending on the available bandwidth.

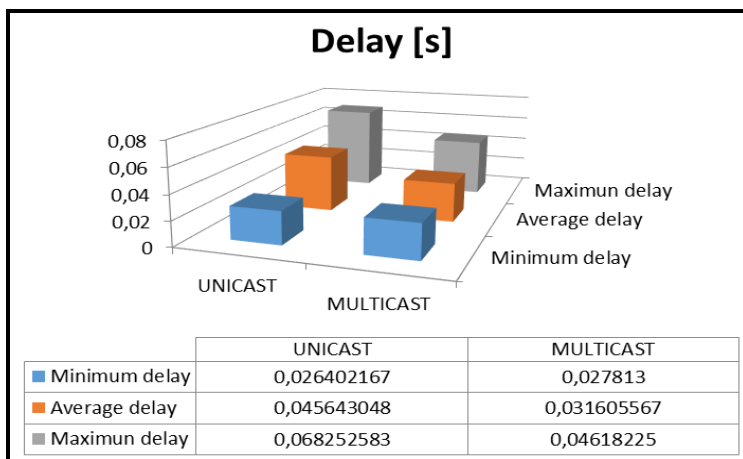


Fig. 4. Delay average, maximum, and minimum

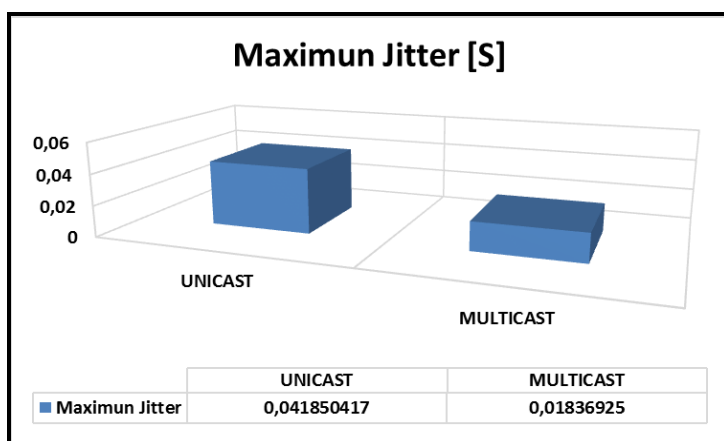


Fig. 5. Maximum jitter

4.3 Impact of the bursts of traffic

Most of the values of the metrics up to now have been averages. However, it must be considered that the video traffic introduces bursts of frames, that is to say, spikes in traffic. This may cause transient load commitments in the links and generate losses of frames. The best way to assess this behavior is to analyze the statistical distribution of the interframe time for the different codecs in one of the ends of the link.

In statistics, the Pareto distribution, formulated by the sociologist Vilfredo Pareto, is often used with said object. It is a continuous probability distribution with two parameters a and b whose density function for values $x \geq b$ is:

$$f(x) = \frac{ab^a}{x^{a+1}}$$

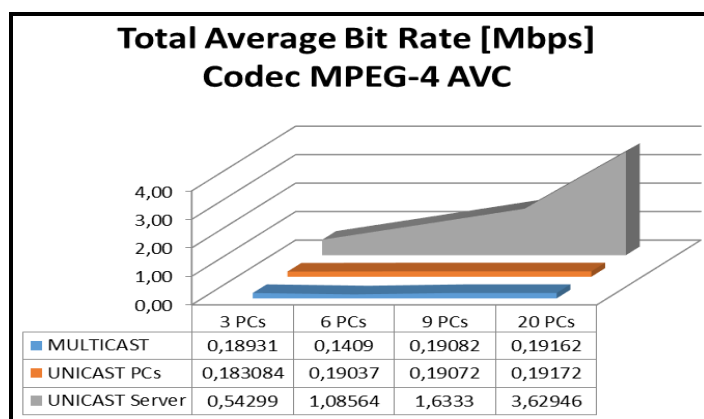


Fig. 6. Total average transfer rate – MPEG-4 AVC

- Behavior autosimilar codec MPEG-4 AVC, MPEG-4 VISUAL and MPEG-2

It analyzes the distribution of the 906, 913 and 608 frames of the video transfer, grouped by interframe spaces, for MPEG-4 AVC, MPEG-4 VISUAL and MPEG-3, respectively, compared with the expected distribution of Pareto. According to the previous point following vectors A comply with the Pareto distribution, and the type of traffic autosimilar has the parameters $a=0.75$ and $b=1.3$, $a=0.4$ and $b=1.75$, and $a=0.98$ and $b=1.15$, respectively.

	0.000	0.001	0.002	0.003	0.004	0.005	0.006	
0	451	200	56	43	30	24	23	...

	0.000	0.001	0.002	0.003	0.004	0.005	0.006	
0	274	124	101	87	43	36	28	...

	0.000	0.001	0.002	0.003	0.004	0.005	0.006	
0	344	98	57	25	22	17	12	...

- Comparison of the autosimilar distribution between codecs

In Figure 7 are analyzed comparatively standard distributions of the frames grouped by interframes spaces for the 3 codecs. It is noted that although all have a general behavior and typical traffic autosimilar have slight differences on the point of greatest concentration of burst of traffic (indicated in the various parameters a and b obtained from the Pareto function).

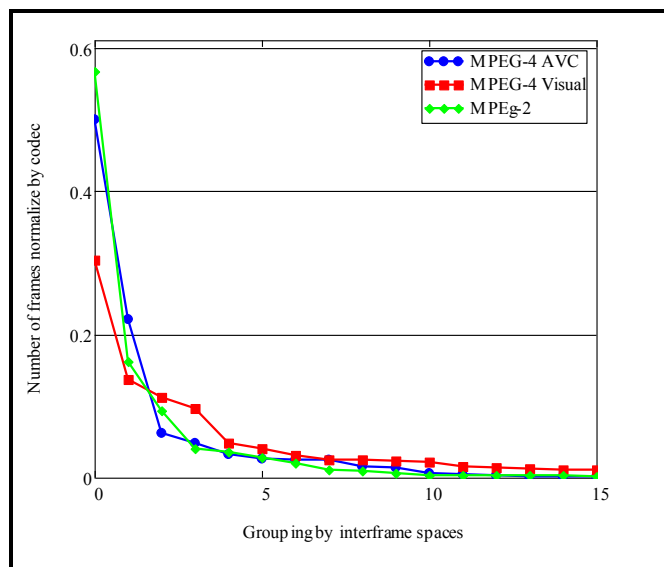


Fig. 7. Contrast of the number of frames 3 standard codecs

5 Conclusions and future work

This experimental work had as objective to make a contribution to the knowledge, updating and advancing the art state in the performance and QoS, multicast and unicast traffic for streamings of video in a real lab network as a testbed of a WAN.

It is presented in detail the experimental results obtained on various direct metric, average and statistics, according to a general topology and 4 scenarios, using a streaming video with 3 different encodings, with types of multicast and unicast traffic.

The pilot study shows that:

- The metrics collected on the laboratory pilot WAN show that multicast traffic provides the QoS and the expected return on each station, and on the server network segment, and the shared segments and links, using different types of codecs, to guide us about the behavior it can expect in a WAN real.
- Variations in the characteristics of the performance metrics and QoS observed in the multicast traffic video basically depend on the codec used. However, in the case of the unicast traffic of video it increases the dependency on the available bandwidth,

the interframe time, and the behavior of bursts on the various shared links (which vary directly according to the amount of connected clients simultaneously).

- The unicast traffic was more sensitive than the multicast for the number of end users, especially in the segment where is the unicast applications server and in the shared segments by the groups of customers unicast.
- The impact of the mixture of a global traffic link will depend on the topology of the codecs used, the number of customers to serve and the settings that are applied to the topology by multicast protocols of the network. And specifically, will depend on whether the video traffic is unicast or multicast.
- An adequate administrative management of the network can obtain an efficient use of the resources available, without compromise performance and the expected QoS.

Future work are oriented to the continuation and deepening of this study. In one of them follows a line of research by adding wireless clients to the topology and IPv6 and its new features for multicast traffic.

References

1. <http://www.arubanetworks.com/>, (Accessed March 16th, 2017)
2. Ian Brown, Jon Crowcroft, Mark Handley, Brad Cain "Internet Multicast Tomorrow" - Articles of interest Vol 1 N°6 - <http://www.isoc.org/pubs/int/cisco-1-6.html>
3. A. Abdalla, A. Mazhar, M. Salah and Sahar Khalaf, "Comparative Study of Compression Techniques for Synthetic Videos," *The International Journal of Multimedia & Its Applications (IJMA)*, vol. 6, no.2, April 2014.
4. D. Grois, D. Marpea, T. Nguyena and O. Hadarb, "Comparative Assessment of H.265/MPEG-HEVC, VP9, and H.264/MPEG-AVC Encoders for Low-Delay Video Applications," *SPIE Proceedings*, vol. 9217, *Applications of Digital Image Processing XXXVII*, California, USA, September, 2014.
5. V. Chandrasekar, K. Baskaran - Performance of Video Conferencing using Protocol Independent Multicast Routing with Core failure - *International Journal of Computer Applications* (0975 – 8887) Volume 41– No.10, March 2012
6. Sorin Cocorada - Multicast vs. Unicast Transmissions for wireless IP Camera Surveillance Suystems – *Acta Technica Napocensis – Vol 48 N°3* 2007
7. S. Tanwir, H. Perros and B. Anjum, "A QoS evaluation of video traffic models for H.264 AVC video," *Fifth International Conference on Next Generation Networks and Services (NGNS)*, pp. 313-320, Casablanca, Morocco, May, 2014.
8. S. Pérez, H. Facchini, A. Dantiacq, G. Cangemi and J. Campos, "An Evaluation of QoS for intensive video traffic over 802.11e WLANs-Quantitative analysis," *International Conference on Electronics, Communications and Computers (CONIELECOMP)*, pp. 8-15, Cholula, Mexico, february, 2015.
9. S. Pérez, H. Facchini, A. Dantiacq, G. Cangemi and J. Campos, "Behavior of Codecs for Multicast Video Traffic using WAN Test Bed - Experimental Study," *International Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON 2015)*, pp. 269-274, Santiago, Chile, October, 2015.
10. Acosta Escobar M, Treminio Henriquez J, Estudio de IPTV multicasting para la Universidad de Don Bosco – Trabajo final Ingeniería en Ciencias de la Computación.
11. Unreal Streaming Technologies, Unreal Media Server, Los Ángeles, USA, Available: <http://umediasever.net/>.
12. Available: <https://www.wireshark.org/>.
13. Available: https://www.youtube.com/watch?v=iLbnDWS3f_g

Modelado y simulación de arquitecturas de Cloud Computing con CloudSim: comunicación entre entidades.

Rosatto Daniel¹, Bond Román¹, Belizán Maximiliano¹, Morales Martín^{1,2}, Encinas Diego^{1,3}

¹Instituto de Ingeniería y Agronomía - Universidad Nacional Arturo Jauretche

²Unidad CodApli - Facultad Regional La Plata - UTN

³Instituto de Investigación en Informática LIDI (III-LIDI) - Facultad de Informática - UNLP

danielrosatto@gmail.com, roman.alejandro.b@gmail.com, maximiliano.h.belizan@gmail.com, martin.morales@unaj.edu.ar, dencinas@unaj.edu.ar

Resumen. Las diferentes herramientas de simulación para las distintas áreas del conocimiento son excelentes sistemas proveedores de posibles comportamientos en cuanto a obtener conclusiones previas durante el diseño. En el caso de los sistemas de Cloud Computing, la diversidad de herramientas de simulación es escasa e insuficiente, comparando con otros sistemas de simulación como por ejemplo redes de computadoras. En este trabajo se abordará el desarrollo obtenido como una alternativa válida para la implementación de entidades de simulación capaces de soportar intercambios de datos simulados en el contexto de un framework de simulación de arquitecturas de Cloud Computing.

Palabras clave: Cloud Computing, Redes, Simulación.

1 Introducción.

Las ventajas de los sistemas Cloud se combinan con la centralización de la autenticación, la seguridad en el acceso y el licenciamiento de los sistemas virtualizados. Los grandes proveedores de servicios Cloud, desarrollaron productos comerciales para cubrir las necesidades TI de pequeñas, medianas y grandes empresas. En este marco, es posible desarrollar conceptos académicos pero siempre hay restricciones que se vinculan con las características del despliegue, o restricciones relacionadas con los servicios que brinda el proveedor de Cloud. Algunas son: los códigos fuente de los servidores Cloud generalmente no se publican, no es posible acceder a un paquete compilado, o nunca se sabe fehacientemente en donde se encuentran los datos o información.

Al comenzar con el desarrollo de un gran despliegue de un sistema Cloud, los posibles riesgos a abordar son los costos, la incertidumbre de la implementación del paradigma, el nivel de confianza y seguridad que se le proveerá al cliente, según lo expresado en el acuerdo del nivel de servicio. Una buena forma de previsualizar los

resultados es mediante la utilización de sistemas de simulación, configurando los posibles escenarios que reflejen situaciones reales y analizando los datos de salida para determinar si es viable o no realizar tal despliegue.

La herramienta de simulación utilizada es el framework CloudSim, que mediante entidades (datacenters, máquinas virtuales, etc), permite reflejar diferentes escenarios de infraestructura Cloud. Debido a las necesidades de este trabajo y ya que el core de CloudSim permite el agregado de nuevas funcionalidades, se desarrolla una expansión de CloudSim. En el siguiente apartado se explican los criterios de elección del framework CloudSim para realizar el aporte.

1.1 Simulación de arquitecturas de Cloud Computing.

CloudSim [1] [2] es un framework de simulación generalizado y extensible que permite el modelado, la simulación y la experimentación de diferentes infraestructuras y servicios de aplicaciones de Cloud Computing. Un ejemplo de utilización es la simulación de muchos centros de datos.

Su arquitectura consiste en entidades específicas que se representan como clases Java que pueden ser heredadas o variadas para simular experimentos. Estas clases representan centros de datos, hosts físicos, máquinas virtuales, servicios a ejecutar en los centros de datos, servicios en la nube de usuarios, redes internas, centro de datos y consumo de energía de los hosts físicos y elementos de los centros de datos. Además, CloudSim soporta la inserción dinámica de los elementos de simulación y proporciona aplicaciones de paso de mensajes y la topología de la red del centro de datos.

Una definición importante de CloudSim es la de entidad. Una entidad es una instancia de un componente, que es una clase o un conjunto de clases que representan un modelo CloudSim (data center, host). El motor de simulación es capaz de simular el tiempo de ejecución de las aplicaciones ingresadas como Cloudlets con información básica [3].

La versatilidad de CloudSim es la principal ventaja del sistema. La integración de nuevos parámetros y conceptos de la simulación es implementada desde abstracciones preestablecidas convenientemente en este trabajo. Las abstracciones principales son: SimEvent [4], SimEntity [4], DataCenterCharacteristics y Virtual Machine (Vm). En cada una de estas instancias es posible establecer el detalle del comportamiento y el estado de la simulación teniendo en cuenta el alcance del modelado que se planifica.

Se propone una simulación que provea métricas de entrada/salida de servicios Cloud simulados tales como SAAS, PAAS o IAAS desarrollada con objetos que heredan de SimEvent y de SimEntity pero con sus métodos y estados únicos en cada simulación. Las denominaciones de los servicios SAAS, PAAS e IAAS hacen referencia al objetivo del sistema Cloud, es decir cuál es el servicio que provee.

- Un sistema SAAS, (Software as a service) brinda servicios de software a través del acceso Cloud, como por ejemplo el reproductor de videos online de Youtube.
- Un sistema PAAS (Platform as a service) brinda una plataforma de hardware virtualizado a través del acceso Cloud, como por ejemplo los storages de archivos como SugarSync.

- Un sistema IAAS (Infrastructure as a service) brinda una infraestructura de hardware y software completa accesible a través de acceso Cloud, como por ejemplo, los servidores virtuales de Iplan.

Los grandes proveedores de servicios Cloud brindan los tres paradigmas juntos, de tal manera que proporcionan servicios integrales.

1.2 Simulación de procesamiento de datos.

CloudSim establece que el entorno de simulación es un sistema de Data Centers con la administración de un broker que asigna los recursos disponibles. En el modelo de simulación de CloudSim, el objetivo es procesar todas las instancias de Cloudlets, que se definen al inicio de la simulación.

Las Cloudlets son definiciones de unidades de información que cualquier Dashboard de un sistema Cloud genera por la interacción de los distintos clientes. En el modelo de CloudSim, las Cloudlets contienen los requisitos de memoria necesaria, de forma que el broker asigne a una Vm capaz de procesarlo, según su algoritmo de asignación.

2 Desarrollo.

Se logró crear un desarrollo basado en las clases SimEntity y SimEvent de CloudSim, utilizando metodologías de expansión de trabajos previos [5] y [6]. Estas clases no están vinculadas con los Cloudlets porque el objetivo del modelado del ambiente de simulación de CloudSim no es procesar algoritmos de entrada. Sólo obtener la cantidad de Cloudlets procesados por cada data center y cuál fue su costo temporal. En éste caso se establecen las modificaciones en el ámbito de CloudSim para configurar a la memoria Ram de cada Vm (Virtual Machine) como una entidad de simulación que procesa eventos de simulación. Es la condición inicial de la simulación para observar y medir procesos distribuidos en cada una de las Vm del broker. Los datos en las celdas de cada SimEntity Ram, (que se ha denominado RamEntity) serían sincronizados por los eventos de simulación autogenerados por la lógica del nuevo modelo. La medición de los procesos se realizaría en el contexto de un Benchmark precargado en el modelo.

En lo referido a Cloudsim, los servicios Cloud que implementa son capaces de procesar las cloudlets que definen los creadores de CloudSim. En este caso la opción es evaluar los resultados obtenidos en el Log de CloudSim al finalizar el proceso sintético de Benchmark, que interactúa con el Sistema de Archivos sintético de CloudSim. Se explica en el siguiente apartado los llamados y las delegaciones del control entre cada uno de los actores de la simulación.

2.1 Expansión de entidades.

Para crear una nueva entidad cloudsim se debe proceder de manera similar a los cloudsim-examples pero adecuándose a las necesidades y sin modificar los archivos originales de cloudsim.

Se incorporaron las siguientes entidades: primero, una RamEntity como componente de las máquinas virtuales en reemplazo de la variable ram.

En segundo lugar, se cambió la clase Vm por SyntheticVm con una composición de RamEntity.

Los procesos sintéticos surgen de la necesidad de respuesta de la simulación ante factores internos. La delegación de la ejecución de las líneas de procesos sintéticos logra que la respuesta de su algoritmo sea obtenida por los recursos simulados de CloudSim independiente del hardware real que está ejecutándose. En la Figura 1 se observa que toda la simulación depende exclusivamente de los recursos que provee la capa de simulación, debiendo reportar cada paso a la capa de control.

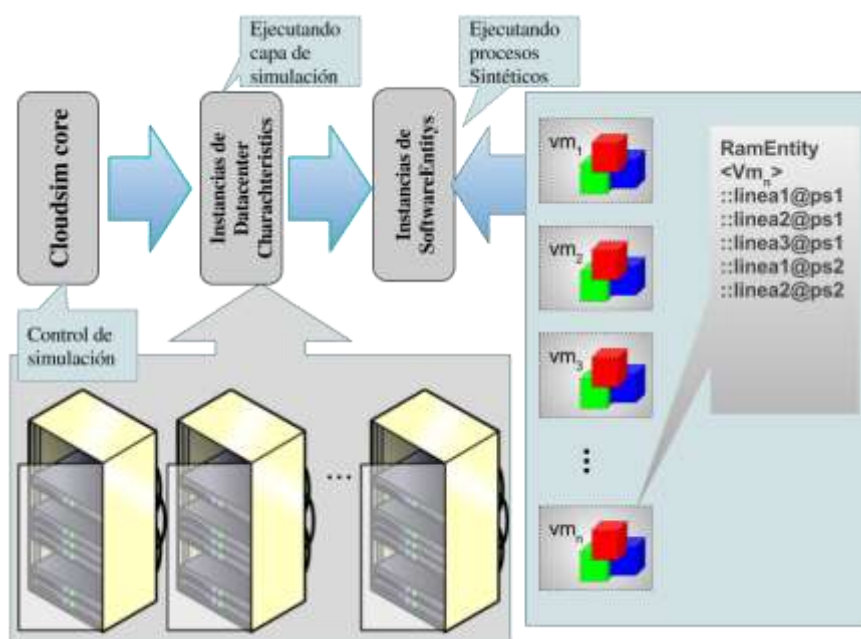


Figura 1: Diagrama de expansión de entidades de CloudSim

Se exponen a continuación el método de desarrollo de la entidad RamEntity, que proporciona el soporte para mantener la simulación de celdas accesibles mediante eventos de simulación.

Estudiando detalles de la clase abstracta SimEvent fue posible definir los requerimientos de software que la nueva entidad debe cumplir. Los requisitos son adecuados para la implementación de RamEntity a partir del cumplimiento de sus restricciones de software. Es por eso que se generan test JUnit RamEntity_test basado en cada caso de los requisitos de los métodos heredados de la superclase SimEntity.

Los Unit test producen una implementación de RamEntity que verifican el cumplimiento de los requerimientos de la superclase con el desarrollo aportado.

En la figura 2 se detallan las relaciones entre las nuevas entidades, el core del simulador y los eventos. Lo interesante en este modelo es que cada celda de RamEntity soporta la asignación de un Object de java. Es decir que se conserva la versatilidad original de CloudSim.

El evento inicial debe ser configurado antes de hacer un start simulation en el tiempo que se asigna en el método send del core de CloudSim. El evento al estar destinado al id (identificador) de la RamEntity2 y la fuente debe ser el id de RamEntity2. Será procesado por la simulación con la correspondiente etiqueta del evento, que en éste caso es una Ram_Access_Read. La versatilidad del simulador se refleja en la posibilidad de que RamEntity1 tiene la capacidad de generar un nuevo evento como respuesta.

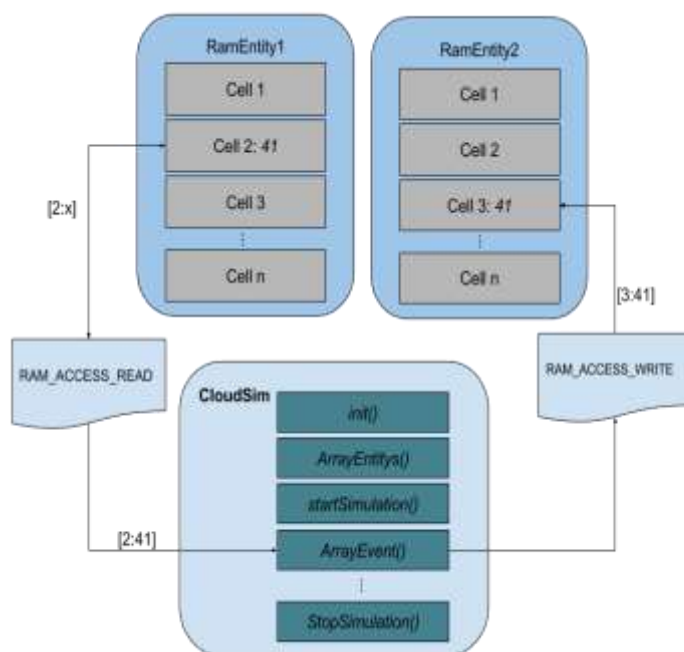


Figura 2: Diagrama de transferencia de información entre celdas RamEntity

Tabla 1: Salida LOG de la simulación de RamEntity.

```

Starting BenchmarkingCloudSimExample1...
Initialising...
Starting CloudSim version 3.0
Datacenter_0 is starting...
Broker is starting...
Entity >vm1&VRAM< : Cells init...
Entity >vm2&VRAM< : Cells init...
Entities started.
0.0: Broker: Cloud Resource List received with 1 resource(s)
0.0: Broker: Trying to Create VM #0 in Datacenter_0
0.0: Broker: Trying to Create VM #0 in Datacenter_0
0.1: Broker: VM #0 has been created in Datacenter #2, Host #0
0.1: Broker: Creation of VM #0 failed in Datacenter #2
0.1: Broker: Sending cloudlet 0 to VM #0
Entity >vm1&VRAM< :59.0 SUCCESS ACCESS READ: -1
Entity >vm1&VRAM< :79.0 SUCCESS ACCESS WRITE: -1 <- vm1&VRAM[23]=4
Entity >vm1&VRAM< :99.0 SUCCESS ACCESS READ: 4
Entity >vm2&VRAM< :101.0 SUCCESS ACCESS WRITE: 4 <- vm2&VRAM[23]=4
Entity >vm2&VRAM< :109.0 SUCCESS ACCESS READ: 4
Entity >vm1&VRAM< :: ENULL Ram event type, time: 118.91 ,data: -1
Entity >vm1&VRAM< :: ENULL Ram event type, time: 198.91 ,data: 4
Entity >vm2&VRAM< :: ENULL Ram event type, time: 218.91 ,data: 4
400.1: Broker: Cloudlet 0 received
400.1: Broker: All Cloudlets executed. Finishing...
400.1: Broker: Destroying VM #0
Broker is shutting down...
Simulation: No more future events
CloudInformationService: Notify all CloudSim entities for shutting down.
Datacenter_0 is shutting down...
Broker is shutting down...
Entity >vm1&VRAM< :Shutdown ...
Entity >vm2&VRAM< :Shutdown ...
Simulation completed.
Simulation completed.

===== OUTPUT =====
Cloudlet ID   STATUS   Data center ID   VM ID   Time   Start Time   Finish
Time
0            SUCCESS    2                0       400    0,1          400,1
Benchmarking CloudSim Example1 finished!

```

Luego, en el segundo resaltado cian, el evento que interactúa con vm1&VRAM produce una lectura de la celda 23, luego escribe 4 en la misma celda. Posteriormente se verifica en el tiempo 99 el valor 4 en la celda 23.

Seguidamente lo resaltado en amarillo verifica de manera similar la escritura de la celda 23 de vm2&VRAM. Los eventos ENULL posteriores representan a los eventos generados por las respuestas de las lecturas y escrituras de RamEntity. Los tiempos de las devoluciones ENULL son los sincronizados por el gestor de tiempos.



Figura 4: Comparación de los tiempos consumidos por cada respuesta de la RamEntity ante los eventos generados.

Se observa en la Figura 4 una comparación de los tiempos de respuesta obtenidos por las acciones de las SimEntity implementadas. En futuras versiones las respuestas de los tiempos se combinarán con un análisis estadístico para establecer las estimaciones de las variables temporales que se requieran.

Tabla 2: [9]

TIEMPO MEDIO DE INICIO DE MV		
Cloud	OS	Tiempo
EC2	Linux	96.9 segundos
EC2	Windows	810.2 segundos
Azure	WebRole	374.8 segundos
Azure	WorkerRole	406.2 segundos
Azure	VMRole	356.6 segundos

En la Tabla 2 se pueden observar los tiempos consumidos en ejecutar una máquina virtual, con distintos sistemas operativos, en los dos proveedores de Cloud Computing más conocidos. Con CloudSim el tiempo es mucho menor, para poder centrarse más en métricas de funcionamiento y de hardware simulado.

4 Conclusiones.

Se logró establecer un método de desarrollo adecuado para un framework de simulación de arquitecturas Cloud. El objetivo de lograr que las entidades de simulación soporten intercambios de datos fue cumplido exitosamente en la entidad RamEntity. Se destaca que el método es genérico para cualquier entidad que se quiera implementar a futuro de acuerdo a los objetivos del modelo.

La precisión del intercambio de los datos es susceptible al comportamiento implementado en el método que procesa los eventos. De manera que mientras más eventos se incluyan en la implementación de éste método, se incrementa la calidad de la transferencia de información sintética entre entidades.

Ampliar este framework ayuda a comprender no sólo los entornos Cloud Computing, sino que también se puede entender mejor este framework para poder seguir expandiéndolo y luego replicarlo en un espacio físico.

5 Trabajos futuros.

Esta línea de investigación es la base a una serie de extensiones de funcionalidades y características que continúan en la etapa de desarrollo. El modelo de simulación seguirá proporcionando la versatilidad original de CloudSim, y además se proporcionarán nuevos actores en el escenario de la simulación. Dichos actores serán entidades y eventos que se consideran relevantes según lo investigado en las experiencias actuales de desarrollo para complementar otras líneas de investigación [7] [8] obteniendo un modelo aplicable.

En los siguientes trabajos, los intercambios de datos en RamEntity darán la posibilidad de crear procesos simulados en una tabla de procesos, también simulada en cada SyntheticVm.

La tabla de procesos simulada estará incluida en la InitEntity que será una entidad administradora de procesos de cada SyntheticVm para dar una respuesta adecuada en cada VirtualMachine simulada.

6 Referencias

1. CloudSim - <http://www.cloudbus.org/cloudsim>. julio 2017
2. Rajkumar Buyya “CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms” Published online 24 August 2010 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/spe.995.
3. Kushang Parikh, Nagesh Hawanna, Haleema. P.K, Jayasubalakshmi.R and N.Ch.S.N.Iyengar. School of Computing Science and engineering Vellore Institute of Technology, Tamil Nadu, “Virtual Machine Allocation Policy in Cloud Computing Using CloudSim in Java.” 2015.
4. Fred Howell & Ross Mc Nab “SimJava: A Discrete Event Simulation Library For Java”. International Conference on Web-Based Modeling and Simulation. 1998.

5. Rodrigo N. Calheiros , Rajiv Ranjan, Anton Beloglazov, César A. F. De Rose and Rajkumar Weiwei Chen , Ewa Deelman - Information Sciences Institute-University of Southern California Marina del Rey, CA, USA - WorkflowSim: A Toolkit for Simulating Scientific Workflows in Distributed Environments. 2012
6. Hamza Ouarnoughi, Jalil Boukhobza, Frank Singhoff, Stephane Rubini, Erwann Kassis. “Considering I/O Processing in CloudSim for Performance and Energy Evaluation”. OpenStack Cloud Software: Open source software for building private and public clouds. © Springer International Publishing AG 2016 M. Tafer et al. (Eds.): ISC High Performance Workshops 2016, LNCS 9945, pp. 591–603, 2016. DOI: 10.1007/978-3-319-46079-6_40.
7. Zaccardi, G.; Galarza, B.; Encinas, D.; Morales, M. “Implementación de Cloud Computing utilizando OpenStack”. II Congreso Nacional de Ingeniería Informática y Sistemas de Información (CoNaIISI 2014). Ciudad de San Luis, Argentina.
8. Galarza, B.; Zaccardi, G.; Encinas, D.; Morales, M. “Análisis de despliegue de una IaaS utilizando Openstack”. XXI Congreso Argentino de Ciencias de la Computación (CACIC 2015). Ciudad de Junín, Argentina.
9. Ming Mao, Marty Humphrey “A Performance Study on the VM Startup Time in the Cloud”. IEEE 5th International Conference on Cloud Computing (CLOUD), 2012.

Procesador de Petri Modular para Sistemas Embebidos

Dr. Orlando Micolini, Ing. Emiliano N. Daniele, Ing. Luis O. Ventre, Geol. Marcelo Cebollada y Verdague, Ing. Maximiliano Eschoyez

Laboratorio de Arquitectura de Computadoras (LAC) FCEfyN
Universidad Nacional de Córdoba

{orlando.micolini, emiliano.daniele, luis.ventre,
marcelo.cebollada.y.verdague, maximiliano.eschoyez}@unc.edu.ar

Abstract. Los Sistemas embebidos, concurrentes y reactivos ejecutan algoritmos con restricciones según los requerimientos de la implementación. Para implementar estos sistemas mediante el uso de hardware y software es posible usar un Procesador de Petri. Desacoplando la lógica y la política de las acciones del sistema se mejora la mantenibilidad y se facilita la validación. Para lograr esto se integra el procesador de Petri con otros procesadores tradicionales, conformando un sistema heterogéneo, lo que agrega la capacidad de verificar el sistema con los formalismos matemáticos del modelo empleado en las redes de Petri. En este artículo se expone una arquitectura modular del Procesador de Petri y la incorporación de colas programables para mejorar la mantenibilidad, el reúso de los módulos y extender su semántica.

Keywords: Procesador de Petri, Redes de Petri, FPGA, IP Core, Procesadores Heterogéneos

1 INTRODUCCIÓN

Los procesadores heterogéneos agregan capacidades específicas no presentes en los procesadores homogéneos. Los sistemas embebidos, concurrentes y reactivos tienen que cumplir requerimientos no funcionales específicos [1]. Estos sistemas interactúan con el medio, del cual reciben eventos de entrada y generan reacciones de salida. El entorno impone las reacciones a cumplir. Durante la interacción del sistema con su entorno físico, hay situaciones en que éste debe reaccionar lo suficientemente rápido para satisfacer restricciones temporales. Esta respuesta depende tanto del algoritmo como de las características de la plataforma de ejecución; por lo que son determinantes los requerimientos del entorno y las prestaciones de la plataforma a utilizar, para estimar los tiempos de respuesta que el sistema tendrá [2].

El sistema propuesto en este trabajo es un sistema heterogéneo, con un “Procesador de Petri” (PP) y un procesador de propósito general (PG). El PP recibe los eventos y los procesa para calcular el próximo estado del sistema.; mientras que el PG calcula y ejecuta las acciones, desacoplando la lógica de las acciones [3].

La arquitectura propuesta implementa en forma directa e innovadora una relación entre el hardware y la lógica del sistema implementado por un monitor de

conurrencia, por lo que el programa prescinde de secciones críticas y sincronización, dado que es el PP quien realiza estas tareas. Existe una relación unívoca entre el programa ejecutado por el PP y el modelo del sistema realizado con la Red de Petri (RdP), lo que garantiza las propiedades que se verifican en el modelo.

Los requerimientos de los sistemas embebidos, reactivos y concurrentes son: precisión, fiabilidad y flexibilidad estructural. Para alcanzar estos requerimientos el PP se programa directamente con el modelo y esto resuelve la ejecución en relación a los eventos y estado del sistema[4], el PP, procesa y ordena eventos según las restricciones del sistema.

Las aplicaciones de este procesador no sólo se limitan al procesamiento de sistemas reactivos, sino que se puede emplear para sistemas paralelos. Se pueden encontrar numerosos casos resueltos con RdP en [5, 6].

La arquitectura del PP desarrollado en trabajos anteriores es de tipo monolítica [3, 4], en el presente trabajo se propone y desarrolla un PP modular que mantiene las ventajas de los anteriores y se agregan nuevas características. En el proceso se tuvo en cuenta ciertas convenciones de nombre, tanto en la implementación interna como en las interfaces, para así compatibilizar los componentes y estandarizarlos.

1.1 OBJETIVOS

Objetivo general.

Se propone separar en módulos cada función del PP para optimizar la mantenibilidad del sistema y la programación de las colas.

Lo que se busca en primera instancia es generar componentes de hardware reutilizables y estandarizados, que permitan reemplazar, quitar y agregar nuevos componentes de forma simple; esto permite implementar el procesador, en sistemas embebidos pequeños e instanciar los módulos necesarios.

La programación de las colas permite que el PP ejecute distintas RdP no autónomas, lo que proporciona mayor expresividad semántica.

2 Redes de Petri

2.1 Redes de Petri Ordinarias (PN)

Una RdP ordinaria (PN) [7] es una quintupla definida por $PN = (P, T, I, I^+, M_0)$ donde:

- $P = \{p_1, p_2, \dots, p_n\}$ es un conjunto finito, no vacío, de plazas.
- $T = \{t_1, t_2, \dots, t_m\}$ es un conjunto finito, no vacío, de transiciones.
- Dado que las plazas y las transiciones forman un conjunto bipartito, se cumple: $P \cap T = \{\emptyset\}$, $P \cup T \neq \{\emptyset\}$
- I, I^+ son las relaciones de incidencia de salida y entrada de las plazas, que indican las relaciones entre las Plazas y las Transiciones (I) y las relaciones entre las transiciones y las plazas (I^+). La Matriz de Incidencia es $I = I^+ - I$

- $M_0 = [m_0(p_1), m_0(p_2) \dots, m_0(p_n)]$ es el marcado inicial de la red, y representa la cantidad de tokens iniciales que contiene cada plaza.

Ahora una RdP queda definida por una 4-tupla: $PN = (P, T, I, M_0)$.

2.2 RdP sincronizadas o no autónomas.

Este tipo de RdP introducen en el modelado los eventos, y son una extensión de las RdP [3, 8]. Las RdP no autónomas modelan sistemas cuyos disparos son sincronizados con eventos discretos externos. Los eventos se asocian con las transiciones, en la Fig. 1 la transición esta sincronizada con el evento E^3 , y el disparo se produce cuando se cumplen las siguientes dos condiciones:

- La transición se encuentra sensibilizada
- Se produce el evento asociado a la transición

Los eventos externos se corresponden con cambios de estado del mundo exterior al sistema (incluyendo el tiempo) mientras que los eventos internos son cambios de estado del sistema en sí. Las RdP sincronizadas pueden definirse entonces como una 3-tupla: $PN_{sync} = (PN, E, sync)$ donde:

PN es una RdP marcada, E es un conjunto de eventos externos y $sync$ es la función que relaciona las transiciones T con $E \cup \{e\}$, donde $\{e\}$ es el evento nulo. Los disparos son atómicos e inmediatos.

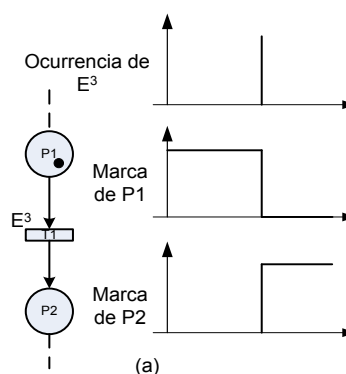


Fig. 1. Transición única asociada a un evento

2.3 Eventos perennes y no perennes.

Existen distintos tipos de eventos que pueden asociarse a una RdP no autónoma, estos son: eventos perennes, no perennes y automáticos o nulos [3].

Los **perennes** son aquellos que cuando ocurren, mantienen su solicitud de disparo hasta que las condiciones de sincronización se cumplan y la transición asociada se dispare. Estos eventos se encolan hasta que sus transiciones puedan dispararse.

Los eventos **no perennes** disparan la transición, si a su arribo esta se encuentra sensibilizada; en caso de que la transición no se encuentre sensibilizada, el evento se pierde. El periodo de tolerancia es de dos ciclos de reloj, por lo que, si luego de los dos ciclos la transición no se sensibiliza, el evento se descarta.

Los eventos **nulos o automáticos** $\{e\}$ están asociados a las transiciones automáticas. Estos son descriptos en el apartado siguiente.

2.4 Transiciones automáticas

Existe un tipo de evento no externo, que ocurre siempre, y que se denomina $\{e\}$.

En una RdP sincronizada, una o más transiciones pueden estar asociadas al evento $\{e\}$, es decir, pueden ser disparadas automáticamente cuando se sensibilizan. Una RdP simple con una transición T_i , que se encuentra asociada a $\{e\}$. Cuando T_i se sensibiliza, se dispara de inmediato, puesto que el evento $\{e\}$ “está ocurriendo siempre”.

2.5 Conflicto entre transiciones

Los conflictos entre transiciones en una RdP sincronizada se dan cuando dos o más transiciones se encuentran sensibilizadas, sus eventos asociados ocurren simultáneamente y el disparo de una de ellas desensibiliza la otra transición. Formalmente, definimos los conflictos como:

$$|\{t_s\}| > 1 \wedge |\{E^s\}| > 1 \text{ y } t_i, t_j \in \{t_s\}$$

Dónde:

- $\{t_s\}$ es un conjunto de transiciones sensibilizadas
- $\{E^s\}$ es un conjunto de eventos asociados a las transiciones $\{t_s\}$
- t_i y t_j son transiciones en conflicto, comparten al menos una entrada con la misma plaza, si se dispara una de ellas la otra transición puede quedar desensibilizada.

Estos conflictos son resueltos con una política de prioridades [6].

2.6 Interpretación de los eventos y la matriz de incidencia

De la semántica de disparo de una RdP no autónoma, se interpreta la matriz de incidencia como la evaluación conjuntiva de las columnas (transiciones) y las filas (plazas). Es decir, para una matriz de incidencia de dimensión $m \times n$ se evalúan n combinaciones de m variables lógicas y si consideramos que cada transición tiene un evento asociado, la expresión es la siguiente:

$$s_i = \left(\bigwedge_{h=0}^{m-1} (M(p_h) \geq i_{hi}) \text{ and } E^i \right),$$

Donde i_{hi} son los elementos de la matriz I y $M(p_h)$ es la marca en la plaza h . El resultado s_i son las componentes del vector de sensibilizado.

Esta ecuación es ejecutada por el PP y es la base de la programación directa de este, puesto que la matriz y los eventos son el programa. Esta arquitectura ejecuta una RdP extendida no autónoma, por lo cual su capacidad semántica es la de una máquina de Turing [9, 10].

3 Arquitectura del PP

El PP desarrollado en este trabajo presenta cambios en la arquitectura implementada en [11]. La diferencia radica en la programación de las colas, la división en módulos independientes de hardware y la optimización específica de cada módulo.

Los bloques principales del PP son: el núcleo, las colas, el módulo de prioridades y la interface con dispositivos externos.

3.1 Núcleo

La responsabilidad del núcleo es mantener el estado de la RdP que el procesador ejecuta. La Fig. 2 representa una versión sintetizada de la arquitectura del procesador donde los componentes que conforman el núcleo están marcados con un (*).

El núcleo está compuesto por los siguientes componentes: Matriz de Incidencia, Vector de Marcado y Vector con Peticiones de Disparo.

Donde las partes y responsabilidad de cada uno de estos componentes son:

- **Matriz de Incidencia I:** almacena la matriz I de la RdP. Su dimensión es de $|T| \times |P|$. Los valores que almacena son enteros.
- **Vector de marcado:** almacena el vector M de la RdP, es decir, el estado de la red. Los valores que guarda son enteros positivos.
- **Colas de entrada:** su interface expone un vector al PP, donde cada posición del vector se corresponde con una transición. El valor de una componente del vector es 1 si hay uno o más eventos en la cola, y es 0 si no hay eventos.
- **Vector de disparos automáticos:** cada posición del vector se corresponde con una transición. Si una transición se configura como automática, el valor correspondiente será siempre igual a 1.
- **Matriz con posibles próximos estados:** corresponde a la suma de cada columna de la matriz I con el vector de marcado. Cada i -ésima columna de esta matriz tiene el estado que se alcanzaría si se dispara la i -ésima transición.
- **Detector de signos:** es un vector en donde cada posición se corresponde con cada columna de la matriz de posibles próximos estados. Si algún valor de columna es

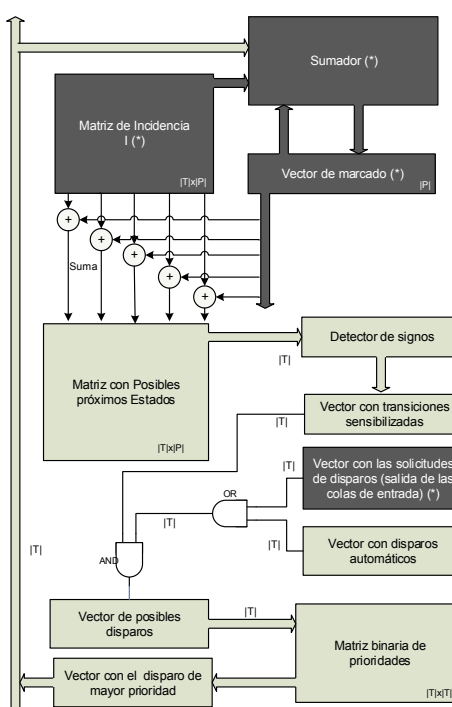


Fig. 2. Arquitectura del PP

negativo, entonces la posición se marca con un 1, indicando que el disparo de dicha columna corresponde a un marcado inalcanzable (valores negativos significa que se trataría de disparar una transición no sensibilizada).

- **Vector con transiciones sensibilizadas:** es la salida negada del vector de signos.
- **Vector de posibles disparos:** almacena los disparos posibles de la RdP, que se calcula a partir del vector de transiciones sensibilizadas y del vector de transiciones con solicitud de disparo (colas de entrada + transiciones automáticas).

3.2 Funcionamiento del procesador.

Siguiendo las características de procesadores anteriores [3], esta nueva versión del PP también ejecuta la RdP en dos ciclos. Se ha adoptado la semántica de servidor único.

Ciclo 1 – Cálculos. En este ciclo se calcula la transición a disparar. Para esto, el vector de marcado se suma a cada columna de la matriz de incidencia, con lo que se obtiene los signos de los posibles próximos estados para cada posible disparo. Esto se totaliza en una matriz, donde la columna i contiene los signos de los próximos marcados.

La salida de la matriz de signos de Posibles Próximos Estados se utiliza como entrada del módulo detector de signos. Este módulo realiza un operación *or* con los valores almacenados en cada columna, si alguno de los valores es negativo, significa que ese próximo estado no será un estado alcanzable por la RdP (la transición no está sensibilizada).

El vector de transiciones sensibilizadas es un vector que contiene un 1 en las posiciones de aquellas columnas de la Matriz de Posibles Próximos Estados que no tenían valores negativos.

Para calcular las transiciones que pueden ser disparadas, necesitamos la información de dos vectores adicionales: el vector de solicitudes de disparo (la salida de las colas de entrada) y el vector de transiciones automáticas. El primero indica cuáles fueron las transiciones que se solicitaron por medio de instrucciones explícitas al procesador (evento de entrada). En cambio, el vector de transiciones automáticas indica aquellas transiciones que no requieren de solicitudes de disparo explícitas $\{e\}$.

Por último, para determinar la transición a disparar, el vector con posibles disparos se utiliza como entrada al módulo de cálculo de prioridades. La salida de este módulo es la transición con mayor prioridad, con un único 1 en la posición de la transición seleccionada. Dicha transición se disparará en el próximo ciclo.

Ciclo 2 – Actualización. En este ciclo se realiza el disparo y se actualiza el valor del vector de marcado. Para lograr esto, la transición seleccionada funciona como un selector de columna. El sumador, que se encuentra en la parte superior de la Fig. 2 lleva a cabo la suma del vector de marcado, con la columna de la matriz I seleccionada por el vector de disparo. El resultado de esa suma se almacena como el nuevo vector de marcado.

Además, según corresponda, se actualizan las colas; incrementando el contador de la cola de salida y decrementando el contador de la cola de entrada.

3.3 Colas.

Las colas de entrada y salida del PP fueron rediseñadas, para ser configurables.

Existe una instancia de cola de entrada y de salida por cada transición que posea el PP. Cada instancia incluye un contador saturado en ambos extremos, que cuenta valores enteros positivos. Cada contador posee un detector de máximo que se utiliza para la señal de overflow, y un detector de cero para indicar que la cola está vacía.

Cola de entrada.

La cola de entrada almacena las peticiones de disparo para las transiciones de la RdP, en la Fig. 3 (a) se muestran las señales necesarias para implementar este módulo. Las colas poseen tres modos de funcionamiento.

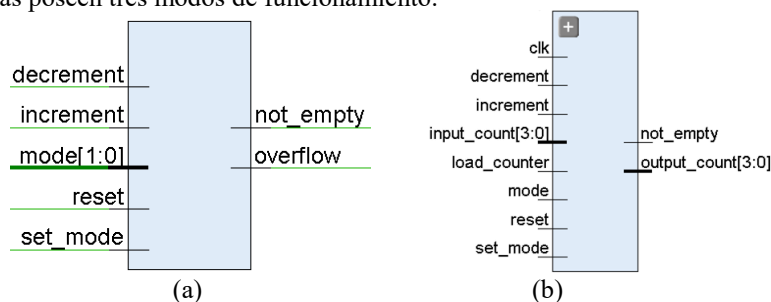


Fig. 3. (a) Cola de entrada, (b) Cola de salida

Modo 1 – Modo normal (evento perenne, modo por defecto): El contador binario se incrementa con cada solicitud de disparo (instrucción explícita que llega al procesador), y se decrementa por cada disparo realizado en el ciclo 2 del funcionamiento del PP. Al contabilizar eventos perennes, el contador se mantiene con su cuenta hasta que se incremente o decremente a través de las señales de entrada de la cola. En este caso, los eventos encolados se mantienen en ese estado hasta que se dispare su transición asociada. Este modo es el que se utiliza en los PP previos

Modo 2 – Transición automática: Este modo de funcionamiento se caracteriza por mantener la señal **not_empty** en alto, ver Fig. 3 (a). Esto se puede interpretar como una transición que siempre se encuentra con solicitud de disparo, o lo que es lo mismo, como una transición automática.

Modo 3 – Evento no perenne: Los eventos que se acumulan en la cola son no perennes. Esto significa que, pasado un determinado tiempo (dos ciclos del PP), el contador asociado vuelve su cuenta a cero. Esto indica que el evento encolado si no ha disparado la transición no deberá dispararla.

Cola de salida.

Como se observa en la Fig. 3 (b), las colas de salida almacenan la cuenta de los disparos realizados. Si se solicita disparar una transición, la cola de salida asociada incrementa en uno su contador y cuando la cola es leída se decrementa en uno. Las colas de salida pueden configurarse, en dos modos de funcionamiento.

Modo 1 – Modo informa: en este modo, el contador interno se incrementa cuando se dispara su transición asociada, y se decrementa cuando se lee para saber si la

transición se disparó. Si se activan las interrupciones del procesador, se genera una interrupción cuando el contador es distinto de cero.

Modo 2 – Modo no informa: en este modo, el contador interno no se utiliza y la señal **not_empty** es cero.

Prioridades y conflictos entre transiciones.

El PP no puede detectar un estado de conflicto, por esto todas las transiciones sensibilizadas como si estuvieran en conflicto. El vector de posibles disparos del PP representa a todas aquellas transiciones que se encuentran en condiciones.

El PP dispara solo una transición por ciclo de ejecución (servidor único). Esto soluciona el problema de los conflictos y mantiene al sistema determinístico; la decisión, de que transición disparar, la toma el módulo de prioridad. Este es implementado por una matriz binaria que establece una relación para determinar la transición de máxima prioridad. Este módulo es también programable en tiempo de ejecución.

Interfaz con Microblaze MCS.

Dado que el PP trabaja con un procesador asociado [3], para construir el sistema es necesario conectar al PP con un procesador tradicional, que ejecute las acciones requeridas por el sistema. Se elige el softcore microblaze MCS [12], puesto que tiene un bajo impacto en los recursos necesarios para su implementación y es soportado por la herramienta de desarrollo de Xilinx [13].

La arquitectura del software que se ejecuta es muy simple y consta de un programa principal y dos drivers. El primer driver es el encargado de exponer una interfaz de comunicación con el PP (**pp_driver**) y el segundo driver (**external_comm**) es el encargado de inicializar y controlar al módulo UART para enviar instrucciones, mediante conexión serial.

4 Resultados

Con el PP implementado se han ejecutado diferentes casos de aplicación para evaluar su desempeño, podemos destacar que se han corrido con éxito los problemas de líneas de producción planteados por Naiqi Wu y MengChu Zhou en [14], las comparaciones se han realizado con respecto resultados obtenidos en [3], arrojando resultados similares.

4.1 Consumo de recursos de la FPGA

Para determinar la cantidad de recursos utilizados y luego compararlas con las demás implementaciones, se instanció al procesador con diferentes cantidades de elementos de los vectores y matrices. Se realizaron múltiples síntesis del núcleo utilizando datos de 4 y de 8 bits de longitud.

La Fig. 4 muestra, para las configuraciones sintetizadas, la cantidad de recursos que se utilizaron de la FPGA. Los registros corresponden a los flip-flops consumidos,

mientras que las LUTs, están relacionadas directamente con la arquitectura de la FPGA Atlys, donde se implementó el procesador.

En la Fig. 4 se observa el aumento exponencial del consumo de recursos, a medida que se aumenta la cantidad de elementos de las matrices y vectores. Una característica importante es que las configuraciones de 4 bits ocupan aproximadamente la misma cantidad de registros que la configuración de 8 bits inmediatamente anterior (incluso en algunas configuraciones la cantidad de registros disminuye). Por ejemplo, la configuración 32x32x4 ocupa menos recursos que la configuración 24x24x8. Eso significa que se pueden procesar redes más grandes, disminuyendo la cantidad de bits utilizados.

El impacto Microblaze el Microblaze es del 12,72% de las LUTs, y el 21,09% de los registros. Estos recursos son los mismos para todas las arquitecturas del PP.

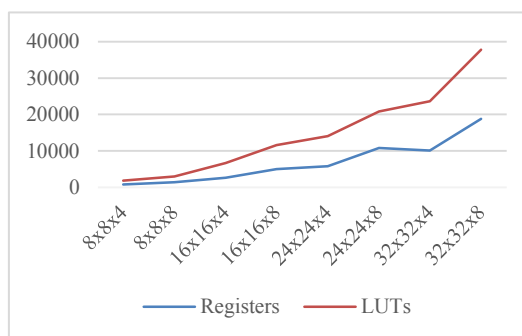


Fig. 4. Gráfico de consumo de recursos

4.2 Análisis de frecuencias

Para realizar un análisis de las frecuencias, se utilizó el mismo procedimiento de la sección anterior. Esta muestra las frecuencias máximas que se alcanzaron, en cada una de las instancias implementadas del PP.

Un dato importante a tener en cuenta es que el aumento de la longitud de datos almacenados en las matrices y vectores, no afecta de manera significativa a la frecuencia máxima alcanzable. Por ejemplo, si tomamos la configuración 8x8x4 y 8x8x8, la diferencia de frecuencia es solo de 2 MHz (80,073 MHz y 78,422 MHz respectivamente). Sin embargo, si aumentamos la cantidad de elementos de las matrices, es decir, el tamaño de la RdP, la frecuencia es significativamente más baja. Por ejemplo, entre las configuraciones 8x8x8 y 16x16x8 existe una diferencia de 23 MHz aproximadamente (78,422 MHz y 55,32 MHz respectivamente). La limitación en la frecuencia está dada principalmente por la FPGA utilizada.

5 Conclusiones

En el presente proyecto, se amplió y se hizo modular la arquitectura del PP implementado en [3], obteniendo un sistema conformado por componentes interconectados en reemplazo del bloque monolítico. El diseño de módulos y la inclusión de las colas programables en el PP no aumentaron los recursos necesarios. La programación de las colas soporta distintos tipos de eventos. Se ha simplificado la mantenibilidad del procesador, por lo que es posible agregar nuevas funcionalidades, como redes jerárquicas y temporales, extendiendo solo la etapa de control.

Los resultados obtenidos de las optimizaciones de hardware muestran que es apto para sistemas embebidos que requieran de hasta 32 condiciones (transiciones), 32 variables lógicas (plazas) y 32 eventos que deban ser evaluados simultáneamente.

La inclusión del módulo de comunicación serial ha facilitado las pruebas del PP, configurarlo y programarlo desde una consola.

La implementación modular del PP implica un avance para el mantenimiento, la escalabilidad y la futura autoconfiguración del PP. Aún más, la inclusión de colas programables ha ampliado la capacidad semántica del PP.

6 Bibliografía

- [1] A. Munir, A. Gordon-Ross, and S. Ranka, *Modeling and Optimization of Parallel and Distributed Embedded Systems*, 2016.
- [2] A. Gamatié, *Designing embedded systems with the Signal programming language: synchronous, reactive specification*: Springer Science & Business Media, 2009.
- [3] O. Micolini, "ARQUITECTURA ASIMÉTRICA MULTI CORE CON PROCESADOR DE PETRI," Doctor, Informatica, UNLaP, La Plata, Argentina, 2015.
- [4] E. Bainomugisha, A. L. Carreton, T. Van Cutsem, S. Mostinckx, and W. De Meuter, "A survey on reactive programming," in *ACM Computing Surveys*, 2012.
- [5] F. Moutinho and L. Gomes, *Distributed Embedded Controller Development with Petri Nets: Application to Globally-Asynchronous Locally-Synchronous Systems* vol. 150: Springer, 2015.
- [6] M. Haustermann. (2017). *Applications of Petri Nets*. Available: <https://www.informatik.uni-hamburg.de/TGI/PetriNets/applications/>
- [7] M. Diaz, *Petri Nets Fundamental Models, Verification and Applications*. NJ USA: John Wiley & Sons, Inc, 2009.
- [8] R. David and H. Alla, *Discrete, continuous, and hybrid Petri nets*. Springer Science & Business Media, 2010.
- [9] J. E. H. R. M. J. D. Ullman, "Introduction to Automata Theory, Languages, and Computation " *Prentice Hall*, 2006.
- [10] L. Popova-Zeugmann, "Time and Petri Nets," *Springer*, 2013.
- [11] M. Orlando, P. Martín, G. N. A., and A. M. A., "Procesador de Petri para la Sincronización de Sistemas Multi-Core Homogéneos," in *CASECongreso Argentino de Sistemas Embebidos*, 2012, pp. 3-8.
- [12] I. Xilinx, "Microblaze processor reference guide," *reference manual*, vol. 23, 2011.
- [13] XILINX. (2017). *ISE WebPACK Design Software*.
- [14] M. Z. Naiqi Wu, *System Modeling and Control with Resource-Oriented Petri Nets*. Boca Raton, FL, 2010.

Requerimientos de VoIP en Redes Malladas Inalámbricas con Relaying

Arístides Silvestris¹, Federico Aguirre¹, Noelia Veglia¹ y Ezequiel Tardivo¹,

¹Grupo de Investigación y Desarrollo Aplicado a las Telecomunicaciones (GIDAT) - UNRC
Ruta Nacional N° 36 – Km. 601 – Río Cuarto – Córdoba – Argentina
{asilvestris, faguirre, nveglia, etardivo}@ing.unrc.edu.ar

Resumen. Argentina es un país longitudinalmente extenso, por lo que existen muchas zonas/comunidades de entre 1000 y 2000 habitantes que carecen de infraestructura de telecomunicaciones, ya que no es fácil amortizar económicamente la inversión por parte de las empresas prestatarias. Ante esta realidad, surge la necesidad de plantear un sistema de comunicaciones simple, robusto, de fácil instalación y económico. La mayoría de las publicaciones vinculadas a la temática “cobertura rural” proponen despliegues de redes inalámbricas del tipo punto-multipunto, malladas, híbridas con *relaying*; sin embargo, en ninguna de ellas se toma consideración de los parámetros necesarios para establecer una llamada IP de una calidad aceptable y comprensible, como lo son ancho de banda por canal, latencia, jitter, etc. Este trabajo propone evaluar la factibilidad de un sistema comunicaciones de voz para el caso mencionado, implementando tecnologías de *relaying* sobre una topología mallada inalámbrica realizando una implementación del sistema.

Palabras claves: Redes malladas inalámbricas, Relay, VoIP, latencia, jitter.

1 Introducción

Argentina es un país longitudinalmente extenso, posee determinadas zonas donde la densidad poblacional es baja y los habitantes no tienen la posibilidad de acceder a algunos servicios, esto ocurre porque no se han desarrollado las obras de infraestructura en tiempo y forma, como gasoductos, cloacas, etc. En lo que respecta a las telecomunicaciones, ocurre algo similar, ya que no es fácil amortizar económicamente la inversión debido a la densidad de habitantes. Este problema se mantiene pese al tendido de la red de fibra óptica impulsado por el estado nacional.

En los últimos 10 años las investigaciones referidas a infraestructuras de rápido despliegue y bajo costo, basadas en redes malladas inalámbricas (*WMN*, por sus siglas en inglés “*Wireless Mesh Network*”) [1], ha sido un tópico de inmensa explotación en el dominio de las telecomunicaciones. El abordaje de dicha temática se realizó durante la primera década de este siglo focalizándose en dos aspectos: Análisis radioeléctrico y Algoritmos eficientes de ruteo [2], [3]. Sin embargo, en estos trabajos se exponen resultados de simulaciones realizadas para modelos acotados, sin realizar pruebas piloto en donde se pueda evaluar el desempeño real de este tipo de sistemas.

Por otro lado, la eficiencia energética, primordial para sistemas que deseen poder operar autónomamente, así como el bajo costo de implementación de dicha arquitectura, la cual se obtiene como producto de la utilización de WiFi SoCs (*por sus siglas en inglés "WiFi - System-on-Chip"*) no fue explotada hasta entrado el año 2010. Con respecto a este último punto (bajo costo de infraestructura) se puede citar un reciente artículo [4], en donde se presentan las ventajas de las WMN para aumentar la penetración en el mercado de aplicaciones de comunicación masivas, en particular el trabajo propone la implementación de protocolos de comunicación para internet de las cosas. Mientras que en lo referido a eficiencia energética en topologías WMN, utilizando técnicas de relaying, la bibliografía consultada para los últimos años [5], [6], [7] presentan resultados que en todos los casos mejoran o aumentan la vida útil de las baterías, a costa de disminución en el throughput de la red.

Sin embargo, al analizar trabajos ligados a la aplicación de VoIP en WMN con técnicas de relaying, se observó que la mayoría de los trabajos realizados fueron desarrollados entre de 2005 y 2008 [8], [9], [10]. Esta carencia percibida constituye la principal motivación para la realización del presente artículo.

El objetivo de este trabajo es evaluar la factibilidad de implementar un sistema simple para comunicaciones de voz en caso de que los sistemas de telefonía no sean plausibles económicamente, mediante la implementación de la tecnología inalámbrica de retransmisión de paquetes enfocados en una topología mallada [11]. En la sección 2 se presentan los conceptos sobre la convergencia hacia el mundo IP y evolución de las WMN [1], [12], en 2.1 se presentan las técnicas de relaying y en 2.2 las configuraciones posibles para cada nodo. En la sección 3, se presenta el Hardware utilizado para el despliegue y el escenario. En la sección 4, se presenta como afecta el uso de relaying en un enlace, en 4.1 se realiza un contraste entre el throughput en condiciones ideales contra el experimental para cada estándar utilizado, en 4.2 se presentan las latencias obtenidas y se mencionan las características necesarias para establecer llamadas aceptables de VoIP utilizando el codec G.711 [13].

2 WMN: Convergencia hacia el mundo IP

Los sistemas de telecomunicaciones se encuentran totalmente interconectados y cada vez es más notoria la convergencia hacia el mundo IP, en el que todos los dispositivos del hogar, oficina e industria, se conectan a la red de datos. Acompañado a esto, se han desarrollado y evolucionado las tecnologías inalámbricas a un precio relativamente bajo considerando las versiones b, g y n del estándar IEEE 802.11. Diversos fabricantes diseñan WiFi SoCs que integran cada vez más funcionalidades, como múltiples interfaces de radio, varios rangos de frecuencia de trabajo, etc. Además, el costo del Hardware es bajo en comparación con otros tipos de tecnologías inalámbricas. Por lo tanto, son atractivos y suelen ser aprovechados para el desarrollo de sistemas de bajo costo, consumo y tamaño. Acorde a las ventajas ya mencionadas, se plantea el despliegue de nodos conectados entre sí inalámbricamente con la finalidad de subsanar la falta de infraestructura de telecomunicaciones en zonas remotas. Se aprovecha, además, las ventajas de la topología de las WMN, como la redundancia de caminos, auto configuración, auto reparación, etc. [1], [14].

Las WMN poseen un funcionamiento peculiar, a diferencia de las redes inalámbricas punto-a-multipunto, las cuales se calculan con un radio de cobertura que se ve afectado principalmente por la geografía del terreno o por la distribución de edificaciones, en una WMN no existe un radio de cobertura como tal. En estas redes es posible clasificar a los nodos como cliente mesh, router mesh y gateways. Los clientes mesh pueden ser dispositivos tanto fijos como móviles. Los router mesh, son los encargados de reenviar el tráfico generado por los clientes hacia los gateways, estos últimos son los que proveen de conexión al exterior a la red mesh. Por lo tanto lo que se tiene en realidad es una “nube mesh” formada por los router mesh donde se les brinda servicios a los clientes mesh, ver Fig. 1. Debido a las características de las WMN se pueden superar obstáculos fácilmente sin necesidad de usar grandes torres o altas potencias de transmisión, solventando problemas de infraestructura [5], [15].



Fig. 1. Esquema de WMN

2.1 Técnicas de Relaying

Las técnicas de retransmisión son un elemento clave para mejorar la capacidad de las redes inalámbricas, debido a sus grandes ventajas, como por ejemplo: mejoras en la cobertura, throughput, capacidad del sistema, duración de las baterías, etc.

En los últimos años, la técnica de relaying para múltiples saltos ha llamado la atención de las universidades y la industria. Los usos más comunes de relaying son Decodificar y reenviar (*DF*, por sus siglas en inglés *Decode-and-Forward*) y Amplificar y reenviar (*AF*, por sus siglas en inglés *Amplify-and-Forward*) [6], [7]. El modo DF, la señal recibida es decodificada, luego vuelve a modularse y transmitirse; por otro lado, el modo AF, solo amplifica y transmite sin decodificar. Comparando ambas técnicas, la complejidad de DF es mayor debido a la capacidad de procesamiento requerida, ya que es necesario poder analizar la capa de control de acceso al medio, sin embargo, la interferencia que produce es menor que el modo AF [16]. En estos procedimientos, es crucial agregar encabezados adicionales al realizar las retransmisiones, ya que las respuestas deben conocer el camino de retorno.

2.2 Configuraciones de Nodos Inalámbricos

Como ya se mencionó anteriormente, cada nodo de la nube mesh posee un rol, y en lo que respecta a la configuración física, se pueden identificar por su funcionalidad.

En las WMN es necesario clasificar los tipos de tráfico, como acceso y troncales o backhaul; el primero se encuentra dedicado al transporte de información neta

dedicada a los clientes; por otro lado, los troncales vinculan los nodos y por ellos se transporta además señalización, como por ejemplo: topología de red, rutas para enviar cada paquete, etc. Es posible que un nodo integre varias interfaces de radio 1, 2 o 3 [17], pudiendo compartir la frecuencia central de trabajo y canales asociados.

En la configuración 1, se posee una interfaz de radio, que es utilizada para transmitir y recibir una frecuencia central “frec.1” con un canal asociado “ch1”, también es llamada configuración SISO (*por sus siglas en inglés, Single-Input-Single-Output*). Esto da como resultado que sea utilizada tanto para tráfico de acceso como troncal, trabajando en modo HD (*por sus siglas en inglés Half Duplex*).

En la configuración 2, se poseen dos interfaces, cada una trabaja en una frecuencia “frec.1” y “frec.2”, con sus respectivos canales. En este caso es posible separar físicamente los tráficos asociados a los clientes del backhaul, obteniendo canales exclusivos, mejorando el desempeño de la red, sin embargo, el modo de transmisión sigue siendo HD, ya que se utiliza la misma interfaz para transmitir y recibir.

Finalmente, la configuración 3, integra una interfaz adicional, en esta se utiliza un radio dedicado al tráfico de clientes operando en “frec.1” con su canal asociado “ch1” en modo HD. Las interfaces restantes se utilizan para la interconexión de nodos, donde cada una de ellas se dedica solo a una función transmitir o recibir, con frecuencia central “frec.2” y canales separados “ch2” y “ch3”, esto último da como resultado que la transmisión sea FD (*por sus siglas en inglés Full Duplex*).

En la tabla 1 puede apreciarse la cantidad de interfaces de radio, tráfico asociado por interfaz, frecuencias y canales asignados, como fue descripto anteriormente.

Tabla 1. Configuraciones posibles para cada nodo.

Interfaces de radio	Clientes	Troncal
1	1 Rx y Tx frec.1 - ch1	
	Half Duplex	
2	1 Rx y Tx frec.1 - ch1	1 Rx y Tx frec.2 - ch2
	Half Duplex	Half Duplex
3	1 Rx y Tx frec.1 - ch1	1 Rx frec.2 - ch2 1Tx frec.2 - ch3
	Half Duplex	Full Duplex

En este artículo se evalúa el desempeño de la configuración 1, que posee una única interfaz de radio 1x1 SISO, comparte la misma frecuencia para transmisión y recepción, y además sólo es posible utilizar un canal para ambos tipos de tráfico.

3 Implementación de una WMN aplicando Relaying

Para el despliegue se propone la utilización de plataformas de código abierto, donde la principal ventaja es la flexibilidad y el bajo costo. Se aprovecha el WiFi-SoC Atheros AR9331 que posee un CPU MIPS de 32-bit corriendo a 400MHZ, 64MB DDR2 RAM, además es posible utilizar diferentes tipos de periféricos mediante USB, como fue utilizado en [18]. Particularmente para este artículo se utilizó una memoria externa con la finalidad de correr aplicaciones y protocolos adicionales; para la

comunicación son soportados estándares IEEE 802.3, IEEE 802.11n, etc. Se posee además una antena omnidireccional del tipo IFA con ganancia de 2,5dBi [19].



Fig. 2. Plataforma de Código Abierto basada en Linux Embebido

Existen diferentes formas de desplegar WMN, por ejemplo, la utilización de redes del tipo Ad-Hoc y para ello el protocolo más evolucionado y estable es B.A.T.M.A.N Avanzado (*por sus siglas en inglés, Better Approach to Mobile Ad-hoc Networking*) o Batman-adv. Este es una implementación, en capa 2 del modelo OSI del protocolo de enrutamiento BATMAN original que trabaja en capa 3 [20]. Este último nació a causa de las deficiencias del protocolo OLSR (*por sus siglas en inglés, Optimized Link State Routing*), también orientado a redes Ad-Hoc, cuya desventaja era el consumo de recursos en los procesadores; estos fueron relevados en los artículos [2], [3], [21].

El *hardware* posee un consumo energético muy bajo, lo que lo hace ideal para que sea alimentado por un pequeño panel solar, respaldado por una batería con capacidad moderada. Además, al plantear la instalación en exteriores, puede ser colocado en una caja estanca de PVC con protección para rayos ultra violeta.

En el escenario propuesto se sitúan los equipos al aire libre, equidistantes entre sí, en donde la distancia se encuentra relacionada a la potencia de transmisión, atenuación del medio y MCS (*por sus siglas en inglés, Modulation and Coding Scheme*) acorde a la sensibilidad del receptor. A modo de clarificar la idea, los nodos son representados en la Fig. 3; cabe mencionar que de esta forma se aproxima la mejor performance posible para las zonas rurales.

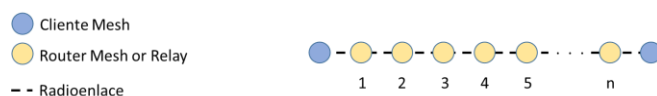


Fig. 3. Escenario propuesto

En dicho escenario, los nodos se encuentran montados sobre una varilla de 1,60mts de altura, son alimentados mediante una batería durante el transcurso del ensayo; fue necesario regular los voltajes de alimentación acorde a los indicados por el fabricante, como se aprecia en la Fig. 4.



Fig. 4. Nodo alimentado por una batería externa

4 Evaluación de WMN-Relaying

El hecho de compartir los canales para los diferentes tipos de tráfico, tiene como consecuencia que el throughput se vea degradado entre saltos [22], ya que es necesario utilizar por lo menos 2 *time-slot* para retransmitir la información, considerando tiempos de computo bajos. La técnica de relay permite la transmisión indirecta de datos entre dos puntos utilizando nodos intermedios, como fue señalado en secciones previas. Al utilizar la misma interfaz de radio para ambos tráficos, el throughput se ve afectado por la ecuación 1, en donde N es el número de saltos.

$$K = 1 / N . \quad (1)$$

La interfaz de radio de cada nodo operará en modo HD, será la encargada de transmitir la información a los clientes que se encuentren en su zona de cobertura y también de re-transmitir la información a los demás nodos junto a la señalización.

4.1 Análisis de Throughput para cada estándar

Al mencionar throughput es posible apreciar que los fabricantes indican las tasas a nivel de capa física. En esta, existe mucha información propia de la especificación como: preámbulos, técnicas de corrección de errores, etc., por lo tanto, al llegar a la capa 3 en donde se realizarán las mediciones, las tasas rondan el 50 y 60% [23], [24], [25], [26]. Si, además, se agrega la técnica de relay, multiplicando por K de la ecuación 1, se enfatizarán aún más las magnitudes; ya que es necesario agregar datos que permitan conocer el origen para poder enviar las respuestas correspondientes.

En la tabla 3 es posible apreciar las magnitudes del throughput de forma ideal, afectado por el número de saltos, por medio de la ecuación 1. Además, al comparar contra la Tabla 4 que posee el throughput medido desde cada nodo por estándar, bajo la configuración ya mencionada. Es notable como este desciende de forma abrupta mientras aumenta la cantidad de saltos en la Fig. 5; similar a lo ocurrido en [27], [28].

Claramente, los valores obtenidos de forma empírica y los calculados idealmente son muy similares.

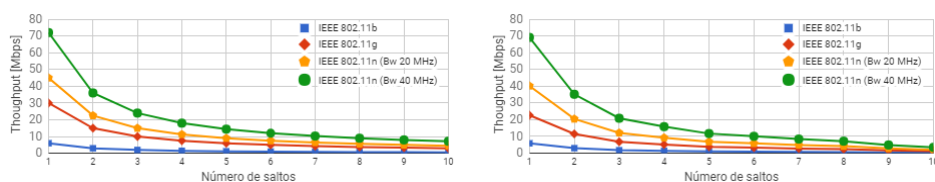
Tabla 3. Throughput ideal para cada especificación contra el número de saltos.

Especificación	Throughput [Mbps] vs Número de saltos									
	1	2	3	4	5	6	7	8	9	10
IEEE 802.11b	6,0000	3,0000	2,0000	1,5000	1,2000	1,0000	0,8571	0,7500	0,6667	0,6000
IEEE 802.11g	30,0000	15,0000	10,0000	7,5000	6,0000	5,0000	4,2857	3,7500	3,3333	3,0000
IEEE 802.11n (Bw 20MHz)	45,0000	22,5000	15,0000	11,2500	9,0000	7,5000	6,4286	5,6250	5,0000	4,5000
IEEE 802.11n (Bw 40MHz)	72,0000	36,0000	24,0000	18,0000	14,4000	12,0000	10,2857	9,0000	8,0000	7,2000

Tabla 4. Throughput medido para cada especificación contra el número de saltos.

Especificación	Throughput [Mbps] vs Número de saltos									
	1	2	3	4	5	6	7	8	9	10
IEEE 802.11b	5,8690	2,9857	1,7694	1,3440	0,9938	0,8536	0,7141	0,6082	0,4154	0,2950
IEEE 802.11g	22,6350	11,5150	6,8240	5,1835	3,8327	3,2923	2,7541	2,3455	1,6022	1,1378
IEEE 802.11n (Bw 20MHz)	40,1650	20,4330	12,1090	9,1980	6,8010	5,8420	4,8870	4,1620	2,8430	2,0190
IEEE 802.11n (Bw 40MHz)	69,1030	35,1620	20,8377	15,8283	11,7035	10,0532	8,4098	7,1622	4,8924	3,4744

Adicionalmente, los nodos generan una demora por el procesamiento previo a la retransmisión del paquete, lo que degrada aún más el enlace como en los ya mencionados [2], [3], [21], [22].

**Fig. 5.** Throughput ideal (a la izquierda) contra el throughput medido (a la derecha) para cada especificación contra el número de saltos

Al utilizar la misma frecuencia y canal para re-transmitir, se produce interferencia co-canal entre los nodos, sin embargo, debido a la disposición de estos, puede despreciarse. No obstante, las interferencias co-canal son consideradas ya que las mediciones no fueron realizadas en una cámara anecoica, por lo que las mismas no han podido ser discriminadas.

Este prototipo se propuso para solventar los problemas de infraestructura de telecomunicaciones, entonces para que funcione fielmente, es necesario indicar que para establecer una llamada VoIP de calidad decente, deben cumplirse ciertos requisitos: ancho de banda, dependerá de la compresión o codec utilizado; latencia, vinculada a la distancia entre los puntos finales de la comunicación; jitter que acatará el comportamiento la red, como también la pérdida de paquetes. Todos los parámetros mencionados, se ven comprometidos en la solución propuesta.

4.2 Análisis de Latencia

Según los criterios de calidad vinculados a la UIT (Unión Internacional de Telecomunicaciones) [29] y la IETF (*Internet Engineering Task Force*), es posible tolerar entre los dispositivos finales una latencia menor a 150 ms, junto a un jitter

menor a 100 ms [30]. Por otro lado, el ancho de banda requerido y la pérdida de paquetes, dependen totalmente del codec. Para el modelo propuesto, se adopta el codec G.711 [13] que posee 2 tipos de compresión, ninguna de ellas posee cargos en VoIP. Este comprime logarítmicamente, obteniendo una tasa de bits de 64 Kbps. Este codec no necesita mucha capacidad de procesamiento para entregar una buena calidad de audio con puntaje de opinión media (*MOS*, por sus siglas en inglés “*Mean Opinion Score*”) de 4.2, las muestras se toman cada 20 o 30 ms para garantizar que la señal sea débilmente estacionaria. Sin embargo, el ancho de banda requerido es superior a otros codec licenciados, rondando los 83 Kbps en cada sentido.

Entonces, según los criterios de calidad, es posible establecer una llamada entre 2 dispositivos atravesando saltos mayores a 10 para cualquiera de los estándares, respetando los criterios de ancho de banda. Sin embargo, existen otros parámetros críticos como la latencia y jitter, que indican que no sería recomendable sobrepasar los 5 o 6 saltos, se aprecia en la tabla 5 y en la Fig. 6 de tiempo de ida y vuelta (*RTT*, por sus siglas en inglés, “*Round-Trip Time*”), similares a lo ocurrido en [4], [8].

Tabla 5. RTT para diferente número de saltos [ms]

Tiempo	RTT [ms] vs Número de saltos									
	1	2	3	4	5	6	7	8	9	10
Mínimo	1,33	2,60	3,65	4,89	6,05	7,25	8,59	9,64	10,44	11,32
Promedio	2,86	5,38	8,98	14,03	24,30	26,65	31,01	39,02	41,38	42,96
Máximo	25,15	27,08	62,89	73,17	193,43	203,34	217,79	211,80	213,86	232,25

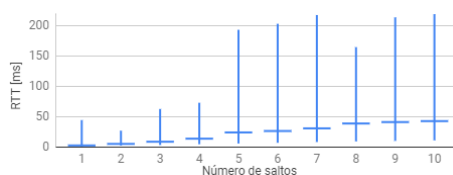


Fig. 6. RTT [ms] para diferente número de saltos

Bajo el mismo escenario para la misma calidad, es posible realizar 15 llamadas concurrentes utilizando 1 salto, 12 llamadas para 2 saltos, 7 llamadas para 3 saltos, 4 llamadas para 4 saltos, 2 llamadas para 5 saltos y 1 llamada para 6 saltos. En caso de que la cantidad máxima de saltos sea superada, se obtiene un audio de mala calidad e incomprensible debido al descarte de paquetes, similar a lo ya indicado en [27], [28].

5 Conclusión

En este trabajo se presentó una solución basada en WMN para brindar conectividad a pequeñas comunidades; fue posible desplegar la arquitectura de forma simple y rápida. Además, con la finalidad de simular distancias mayores y fenómenos de propagación, se ajustó la potencia de transmisión y se fijó los nodos.

Planteado el escenario, se procedió a tomar mediciones acordes a los criterios establecidos. Una vez recabada toda la información, fue posible analizar los valores,

demostrando la factibilidad de realizar una plataforma de comunicación de voz para zonas de cobertura poco extensas debido a la limitación en latencia, consecuencia del uso de más de 6 saltos, como fue expuesto en la sección 4.2.

En este trabajo se encontró un sistema simple que cumple con los objetivos propuestos. Como líneas futuras de investigación, se propone dotar a los nodos de mayor cantidad de interfaces de radio, separando los canales de acceso y troncal. Esto es posible mediante la incorporación de un *Dongle* USB, aprovechando así el mismo *hardware* y escenario. Logrando que la interferencia co-canal sea reducida, ya que no se compartirán los canales, como fue mencionado en la sección 4.1

Referencias

1. Ian.F. Akyildiz, X. Wang, W. Wang, "Wireless mesh networks: a survey," vol. 47, Issue 4, pp. 445-487, March 2005
2. M. Abolhasan, B. Hagelstein and J. C. P. Wang, "Real-world performance of current proactive multi-hop mesh protocols," 2009 15th Asia-Pacific Conference on Communications, Shanghai, 2009, pp. 44-47
3. M. S. Singh and V. Talasila, "A practical evaluation for routing performance of BATMAN-ADV and HWMN in a Wireless Mesh Network test-bed," 2015 International Conference on Smart Sensors and Systems (IC-SSS), Bangalore, 2015, pp. 1-6
4. A. A. dela Cruz, M. L. A. Parabuac and N. M. C. Tiglao, "Design and implementation of a low-cost and reliable wireless mesh network for first-response communications," 2016 IEEE Global Humanitarian Technology Conference (GHTC), Seattle, WA, 2016, pp. 40-46.
5. Ajayi, A.; Roedig, U.; Edwards, C.; Race, N., "A survey of rural Wireless Mesh Network (WMN) deployments," *Wireless and Mobile, 2014 IEEE Asia Pacific Conference on*, vol., no., pp.119,125, 28-30 Aug. 2014
6. Suchitra Shrestha and KyungHi Chang, "Analysis of outage capacity performance for cooperative DF and AF relaying in dissimilar Rayleigh fading channels," 2008 IEEE International Symposium on Information Theory, Toronto, ON, 2008, pp. 494-498
7. A. Ali et al., "Performance analysis of AF, DF and DtF relaying techniques for enhanced cooperative communication," 2016 Sixth International Conference on Innovative Computing Technology (INTECH), Dublin, 2016, pp. 594-599
8. S. Armenia, L. Galluccio, A. Leonardi and S. Palazzo, "Transmission of VoIP traffic in multihop ad hoc IEEE 802.11b networks: experimental results," First International Conference on Wireless Internet (WICON'05), 2005, pp. 148-155
9. Kee Ngoh Ting, Yin Fern Ko and Moh Lim Sim, "Voice performance study on single radio multihop IEEE 802.11b systems with chain topology," 2005 13th IEEE International Conference on Networks Jointly held with the 2005 IEEE 7th Malaysia International Conf on Communic, 2005, pp. 5 pp
10. S. A. Ramprasad, D. Li, U. C. Kozat and C. Pepin, "An Analysis of Joint Aggregation, Bursting, Routing, and Rate Adaptation for Increasing VoIP Capacity in Multi-Hop 802.11 Networks," in IEEE Transactions on Wireless Communications, vol. 7, no. 8, pp. 3128-3139, August 2008.
11. L. Zhao, A. Y. Al-Dubai, X. Li and G. Chen, "A New Relay Node Selection Model for Wireless Community Mesh Networks," 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing, Liverpool, 2015, pp. 1178-1185.

12. Rabbi, M.F.; Rahman, M.T.; Uddin, M.A.; Salehin, G.M.A., "An Efficient Wireless Mesh Network: A New Architecture," *Communication Technology*, 2006. ICCT '06. International Conference on , vol., no., pp.1,5, 27-30 Nov. 2006
13. Recommendation ITU-T G.711, Pulse code modulation (PCM) of voice frequencies, Int'l Telecommunication Union, Geneva, 1988
14. Akyildiz, I. F., Wang, X. & Kiyon (2005), "A Survey on Wireless Mesh Networks", *IEEE Radio Communications*, Vol.0263-6804, pp23-30
15. Kobel, C.; Baluja Garcia, W.; Habermann, J., "A survey on Wireless Mesh Network applications in rural areas and emerging countries," *Global Humanitarian Technology Conference (GHTC)*, 2013 IEEE , vol., no., pp.389,394, 20-23 Oct. 2013
16. Levin, G., & Loyka, S. (2012, February). Amplify-and-forward versus decode-and-forward relaying: Which is better?. In *International Zurich Seminar on Communications* (p. 123)
17. N. Kaur and J. S. Saini, "Performance enhancement of 802.11 based wireless mesh network by using Multi- Radio Multi-Channel," *2013 International Conference on Green Computing, Communication and Conservation of Energy (ICGCE)*, Chennai, 2013, pp. 71-76
18. F. Aguirre, A. Silvestris, J. Alemany and F. Magnago, "A low cost smart metering infrastructure with centralized demand response management," *2015 IEEE PES Innovative Smart Grid Technologies Latin America (ISGT LATAM)*, Montevideo, 2015, pp. 520-524
19. C. Soras, M. Karaboikis, G. Tsachtsiris and V. Makios, "Analysis and design of an inverted-F antenna printed on a PCMCIA card for the 2.4 GHz ISM band," in *IEEE Antennas and Propagation Magazine*, vol. 44, no. 1, pp. 37-44, Feb 2002
20. R. Sanchez-Iborra, M. D. Cano and J. Garcia-Haro, "Performance Evaluation of BATMAN Routing Protocol for VoIP Services: A QoE Perspective," in *IEEE Transactions on Wireless Communications*, vol. 13, no. 9, pp. 4947-4958, Sept. 2014
21. Pramod J, Sahana K S, Akshay A and V. Talasila, "Characterization of Wireless Mesh Network performance in an experimental test bed," *2015 IEEE International Advance Computing Conference (IACC)*, Bangalore, 2015, pp. 910-914
22. Holland, G., & Vaidya, N. (2002). Analysis of TCP performance over mobile ad hoc networks. *Wireless Networks*, 8(2/3), 275-288
23. V. Visoottiviset, T. Piroonsith and S. Siwamogsatham, "An empirical study on achievable throughputs of IEEE 802.11n devices," *2009 7th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks*, Seoul, 2009, pp. 1-6
24. O. O. Olusegun and C. E. Tan, "TCP throughput efficiency enhancement in IEEE 802.11n network," *2013 8th International Conference on Information Technology in Asia (CITA)*, Kota Samarahan, 2013, pp. 1-5
25. P. Mahasukhon, M. Hempel, S. Ci and H. Sharif, "Comparison of Throughput Performance for the IEEE 802.11a and 802.11g Networks," *21st International Conference on Advanced Information Networking and Applications (AINA '07)*, Niagara Falls, ON, 2007, pp.792-799
26. Youngsoo Kim, Sunghyun Choi, Kyunghun Jang and Hyosun Hwang, "Throughput enhancement of IEEE 802.11 WLAN via frame aggregation," *IEEE 60th Vehicular Technology Conference*, 2004. VTC2004-Fall. 2004, 2004, pp. 3030-3034 Vol. 4
27. T. Vanhatupa, M. Hainnikainen and T. D. Hdmalainen, "Multihop IEEE 802.11b WLAN Performance for VoIP," *2005 IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications*, Berlin, 2005, pp. 1925-1930
28. F. Y. Li, A. Hafslund, M. Hauge, P. Engelstad, Ø. Kure and P. Spilling, "Does higher datarate perform better in IEEE 802.11-based multihop ad hoc networks?," in *Journal of Communications and Networks*, vol. 9, no. 3, pp. 282-295, Sept. 2007
29. Recommendation ITU-T G.114, One-Way Transmission Time, Int'l Telecommunication Union, Geneva, 1996
30. Schulzrinne, H., Casner, S., Frederick, R., and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications", STD 64, RFC 3550, July 2003

Un Modelo de Detección de Anomalías en una LAN usando K-NN y Técnicas de Computación de Alto Desempeño.

Mercedes Barrionuevo, Mariela Lopresti, Natalia Miranda, Fabiana Piccoli
LIDIC. Universidad Nacional de San Luis,
Ejército de los Andes 950 - 5700 - San Luis - Argentina
{mbarrio, omlopres, ncmiran, mpiccoli}@unsl.edu.ar

Abstract. Detectar valores anormales a partir de grandes volúmenes de información producido por el tráfico de red ha adquirido un interés considerable en el área de seguridad de redes. Es de relevancia para todo sistema de computadoras conectadas a una red contar con un sistema de detección de eventos anómalos y un tiempo de obtención de tales eventos lo más cercano posible a su ocurrencia. Detectar valores anómalos puede conducir a los administradores de red a identificar fallas del sistema y, por lo tanto, tomar medidas preventivas antes de una masiva propagación. La detección de anomalías es un punto de partida para evitar nuevos ataques. En este artículo, presentamos una forma de pre-procesar datos para identificar anomalías mediante un algoritmo de clasificación K-NN con técnicas de computación paralelas usando Unidades de Procesamiento Gráfico.

Keywords: Tráfico de red. Anomalías. K-NN. Computación de Alto Desempeño. GPU.

1 Introducción

La variedad y complejidad del tráfico actual en Internet superan todo lo imaginado por los diseñadores pioneros de la arquitectura subyacente de Internet. Actualmente estamos inmersos en una sociedad dependiente del uso de sistemas computarizados presentes en diversas ramas como: las finanzas, industria, medicina y varios aspectos de la vida cotidiana. Con el fin de proteger o prevenir estos sistemas y la información importante o de interés para la empresa u organización, es necesario implementar tecnologías o modelos para evitar accesos no autorizados o maliciosos. Esto puede ser posible si se determinan patrones de acceso válidos.

Las amenazas a una red de datos están conformadas por un conjunto de tramas con características específicas que buscan detectar vulnerabilidades en un sistema. Estas representan riesgos, los cuales son usados para realizar ataques.

Al detectar situaciones fuera de lo común, es decir aquellas que se desvían del perfil normal de la red, los administradores se hacen preguntas tales como ¿Qué significan las desviaciones? ¿Se puede considerar tal situación como un ataque? ¿Dicha desviación pertenece a tráfico generado por nuevas aplicaciones? En base a ellos

surgen los sistemas de detección basados en anomalías, los cuales reportan toda actividad no habitual hasta ese momento, la cual puede ser normal o no.

En los sistemas cuyo objetivo es detectar ataques generalmente el margen de error suele ser alto, por ello es necesario contar con una interpretación semántica de los resultados. Esto unido a la multiplicidad de tráfico de red generado por las aplicaciones y las características tales como ancho de banda, duración de las conexiones, entre otras, hacen de esto un trabajo de alto costo computacional.

Generalmente, las investigaciones en esta área tienden a limitar la evaluación de los sistemas de detección de anomalías al cálculo de la desviación de las nuevas instancias respecto al perfil normal, constituyendo un reto convertir los resultados en reportes semánticos para los administradores de redes. En base a esto se propone un Sistema Paralelo-Supervisado de Detección de Anomalías de Red, P-SADS.

En [1] se presentó un primer modelo no supervisado de P-SADS para detectar anomalías a través de comparación de imágenes por SIFT (Scale Invariant Feature Transform) [5]. En este trabajo, se propone combinar técnicas de clasificación de tráfico y computación de alto desempeño para obtener en el menor tiempo posible buenos resultados trabajando sobre grandes volúmenes de datos. Aquí nos enfocamos en la etapa de pre-procesamiento de los datos obtenidos de la red. Se propone usar el algoritmo paralelo de clasificación supervisada K-NN (K vecinos más cercanos) para obtener un conjunto más pequeño de datos a procesar en la segunda etapa.

Este documento está organizado como sigue: la próxima sección describe los conceptos teóricos involucrados en el desarrollo. La sección 3 detalla las características de la primera etapa de P-SADS y en la sección 4 se muestran los resultados experimentales de su desarrollo. Finalmente, se detallan las conclusiones y trabajos futuros.

2 Marco Teórico

En este trabajo se involucran diferentes conceptos, entre los cuales se destacan el tráfico en redes de computadoras y sus anomalías, métodos de extracción de características de paquetes y algoritmos de clasificación supervisada. En esta sección describimos cada uno de ellos.

2.1 Tráfico de Datos en Redes de Computadoras

El tráfico de red proporciona información sobre qué viaja en la red. Los tipos de datos más comunes son los datos de *log* o registro, tales como registros del protocolo de Internet (TCP/IP), registros de eventos, datos de visitas a Internet, datos de informes del protocolo de gestión de red (SNMP), entre otros [10]. Esta información es de interés para la seguridad de la red, específicamente para la detección de eventos anómalos. La figura 1 ilustra un ejemplo de tráfico TCP/IP, donde las filas detallan el tráfico de red individual y las columnas son características específicas de cada tráfico. En el ejemplo, la primera columna es un índice de sesión de cada conexión y la segunda es la fecha en que se produjo la conexión [10].

```

1 06/24/1998 08:12:58 00:00:01 ntp/u 123 123 172.016.112.020 192.168.001.010 0 -
2 06/24/1998 08:12:58 00:00:01 ntp/u 123 123 172.016.112.020 192.168.001.010 0 -
3 06/24/1998 08:15:52 00:00:04 smtp 1024 25 172.016.114.169 195.115.218.108 0 -
4 06/24/1998 08:15:55 00:00:01 domain/u 53 53 192.168.001.010 172.016.112.020 0 -
5 06/24/1998 08:15:55 00:00:01 domain/u 53 53 192.168.001.010 172.016.112.020 0 -
6 06/24/1998 08:15:55 00:00:02 smtp 1025 25 172.016.114.169 196.227.033.189 0 -
7 06/24/1998 08:17:08 00:00:04 smtp 1026 25 172.016.113.084 195.115.218.108 0 -
8 06/24/1998 08:17:11 00:00:02 smtp 1027 25 172.016.113.084 196.227.033.189 0 -
9 06/24/1998 08:17:18 00:00:02 smtp 1028 25 172.016.112.149 195.115.218.108 0 -
10 06/24/1998 08:17:36 00:00:01 domain/u 53 53 192.168.001.010 192.168.001.020 0 -
11 06/24/1998 08:17:36 00:00:01 domain/u 53 53 192.168.001.010 192.168.001.020 0 -
12 06/24/1998 08:17:37 00:00:02 smtp 1029 25 172.016.114.169 194.027.251.021 0 -
13 06/24/1998 08:17:38 00:00:02 smtp 1048 25 172.016.114.169 194.007.248.153 0 -
14 06/24/1998 08:17:39 00:00:02 smtp 1049 25 172.016.114.169 197.182.091.233 0 -
15 06/24/1998 08:17:40 00:00:02 smtp 1051 25 172.016.114.169 195.115.218.108 0 -
16 06/24/1998 08:17:41 00:00:02 smtp 1052 25 172.016.114.169 196.227.033.189 0 -
17 06/24/1998 08:17:45 00:00:01 smtp 1104 25 172.016.114.169 135.008.060.182 0 -
19 06/24/1998 08:18:07 00:00:01 eco/i - - 192.168.001.005 192.168.001.001 0 -
20 06/24/1998 08:18:07 00:00:01 eco/i - - 192.168.001.005 192.168.001.001 0 -

```

Figura 1: Ejemplo de tráfico TCP/IP

Los datos viajando por la red pueden proporcionar información importante en referencia a los comportamientos del usuario y del sistema. Estos tipos de datos se pueden recopilar con productos comerciales o programas específicos. Los datos TCP/IP pueden ser capturados mediante distintas herramientas, denominados sniffers¹.

El tráfico de red está compuesto de paquetes, flujos y sesiones. Un paquete es una unidad de datos enviada entre una fuente y un destino en Internet u otra red basada en TCP/IP; un flujo de red es una secuencia unidireccional de paquetes entre dos puntos finales; y los datos de sesión representan la comunicación entre computadoras. Una comunicación implica el intercambio de varios flujos. Tradicionalmente, un flujo IP contiene un conjunto de atributos, los de mayor interés son: dirección IP origen y destino, puerto origen y destino, y tipo de protocolo. Este último puede ser de capa 2 (TCP o UDP) o 3 (ICMP) si se tiene en cuenta el modelo de 4 capas de TCP/IP. Esta información permite establecer una línea base de comportamiento o patrón normal del tráfico de red haciendo más fácil identificar un comportamiento inesperado o no deseado, denominado tráfico anómalo. Por lo tanto, una estrategia de análisis por anomalías, se basa en la descripción del tráfico de comportamiento normal, clasificando como anomalía a todos los patrones que se alejen de él.

Para obtener el conjunto de datos antes mencionados existen diversas técnicas, algunas son mencionadas en la siguiente subsección.

2.1.1 Análisis y Extracción Específica de Datos de un Paquete

Cuando se estudian los aspectos particulares del tráfico de red, es necesario extraer sólo la información de los paquetes de datos para luego procesarlos. Existen diferentes técnicas de extracción y procesamiento, algunas de ellas son:

- *Representación gráfica de los datos en bruto:* Las representaciones son generalmente en forma de gráficos de dispersión 2D y 3D, gráficos basados en tiempo, histogramas, gráficos circulares o diagramas.

¹ Un **sniffer** es un programa para monitorear y analizar el tráfico en una red de computadoras, detectando los cuellos de botellas y problemas que existan en ella.

- *Información estadística y extracción de patrones*: Basados en cálculos de promedios, distribuciones de tiempo y funciones de distribución de probabilidad.
- *Análisis basado en reglas (firmas), detección de anomalías y políticas*: son todos los análisis de inspección de tráfico en busca de coincidencias con una determinada regla o firma. Las reglas se definen como valores de ciertos campos del encabezado o de una combinación de varios de ellos. Estas técnicas se utilizan en sistemas de detección de intrusos (IDS), como Snort.
- *Análisis basado en el flujo*: se centran en el tratamiento de tráfico de la red como flujos. Como la mayoría de la información intercambiada en una red es orientada a conexión y no orientado a paquetes, el análisis puede tomar ventaja de ello. Un claro ejemplo de flujo de red típico es una conexión TCP, donde los datos intercambiados se rigen por la máquina de estado TCP [8].

Cada una de estas técnicas es adecuada para determinadas situaciones, también es posible realizar una combinación de ellas. Este trabajo se basa en el análisis de flujos.

2.1.2 Ataques Comunes

Uno de los mayores desafíos de los administradores de red es detectar ataques a redes de computadoras. Un ataque consiste en aprovechar una vulnerabilidad de un sistema informático (sistema operativo, software de aplicación o sistema del usuario) con propósitos desconocidos y que, por lo general, causan daño. Por ello es imposible hacer una clasificación completa de todos los ataques reales y posibles debilidades de las redes cuando estas se conectan a Internet. Sin embargo, los de denegación de servicio (DoS) tienen gran interés.

Un ataque DoS proviene de una sola entidad y está destinado a hacer que los recursos o servicios de una computadora no estén disponibles para los usuarios. Existen distintos tipos, en particular este trabajo se ha enfocado en los ataques DoS *Smurf*, *Fraggle* y *Land* [3, 4], cada uno tiene las siguientes características:

- *Smurf*: Este ataque utiliza el protocolo ICMP (Control Management Protocol) para enviar un ping de difusión con una dirección de origen falsificada. Existen diferentes formas de hacer un ping, ellas son:
 - *Ping Normal*: se envía una o más peticiones de echo ICMP a un sistema, el cual responde con una o más respuestas de echo ICMP, verificando así la operatividad del sistema remoto.
 - *Ping de Difusión*: envía la solicitud de tipo echo ICMP a una dirección de difusión. Cada sistema responderá a quien le envió, inundándolo con respuestas de tipo echo ICMP.
 - *Ping de difusión con fuente falsificada*: se envía un ping de difusión con la dirección de origen falsificada con la dirección de la víctima. Cada sistema de la red responderá e inundará a la víctima con las respuestas. Esta operación es una combinación de los anteriores.

El patrón considerado para reconocer este tipo de ataque consiste en analizar para el protocolo ICMP, si las direcciones IP origen y destino pertenecen a la misma red, y si la dirección destino es un mensaje de difusión.

- *Fraggle*: Para comprobar si un sistema está funcionando se pueden utilizar herramientas basadas en UDP en vez de ICMP, inspeccionando si el sistema

está escuchando en un puerto específico o no. Esto se hace comúnmente con diferentes tipos de escaneos de vulnerabilidad usados tanto por atacantes como por administradores de seguridad. Por ejemplo, si un sistema escucha en el puerto TCP o UDP 19, cuando se establece una conexión a éste, el sistema respondería con un flujo constante de caracteres. Normalmente, el sistema origen utilizará el puerto TCP o UDP 7. Cuando el sistema origen comienza a recibir los caracteres, sabe que el sistema de destino está operativo y cierra la conexión. En un ataque *fraggle*, se envía un paquete de difusión con la dirección falsificada al puerto 19 de la víctima, si tienen el puerto 19 abierto responderá un flujo constante de caracteres a la víctima.

El patrón es semejante al de *Smurfpero* para el protocolo es UDP.

- *Land*: es un ataque utilizando el protocolo TCP. Consiste en crear un “bucle infinito”, provocado por el envío de una solicitud SYN con la misma dirección IP de origen y destino. Esto hace que la computadora iterativamente se responda a sí misma logrando el bloqueo del equipo y no aceptando nuevas solicitudes. Como, además, agota todos los recursos del procesador, produce finalmente una denegación de servicio. El patrón considerado para reconocer este tipo de ataque consiste en analizar la coincidencia entre las direcciones IP origen y destino como así también la igualdad de los puertos.

Como se mencionó anteriormente existen muchos otros ataques de denegación de servicios pero la detección de patrones en ellos requiere un análisis más profundo y detallado, lo cual escapa a este trabajo.

2.2 Algoritmo de Clasificación Supervisada K-NN

El proceso de clasificación construye modelos capaces de determinar la pertenencia de un objeto a una categoría sobre la base de sus características. La clasificación es supervisada si de antemano existe un conjunto de observaciones ya clasificadas, y se conoce a cual pertenece cada observación. Los algoritmos dedicados al problema de la clasificación supervisada operan usualmente sobre la información suministrada por un conjunto de muestras, patrones, ejemplos o prototipos de entrenamiento que son asumidos como representantes de las clases [9].

En particular en este trabajo se utiliza el algoritmo de clasificación supervisada basada en criterios de vecindad llamado vecino más cercano (K-NN). El método del vecino más cercano y sus variantes está basado en la idea intuitiva de que objetos similares pertenecen a la misma clase, la clase a la que pertenece un objeto puede ser inferida a partir de la clase a la que pertenecen los objetos de la muestra de aprendizaje más parecidos. La idea de similitud es reflejada formalmente en el concepto de distancia, normalmente se utiliza la distancia euclidiana [6]. Si bien el cálculo del vecino más cercano puede resolverse usando técnicas paralelas, en [6] se ha demostrado que la implementación en paralelo usando GPU [7], obtiene muy buenos tiempos de respuesta.

3 Sistema Paralelo-Supervisado de Detección de Anomalías en una Red: P-SADS

La tarea de detectar posibles paquetes de datos anómalos en una red de computadoras es muy costosa, combinar técnicas de clasificación y Computación de Alto Desempeño (HPC) en su solución parece una buena alternativa. P-SADS es un sistema donde se combina HPC, clasificación supervisada y procesamiento de imágenes para detectar posibles ataques. La arquitectura del sistema planteado distingue dos etapas. En la primera se encuentran todos los pasos necesarios para la captura y pre-clasificación de tráfico, mientras que la segunda es la encargada de detectar posibles ataques en base a la comparación de imágenes generadas con los datos obtenidos en la etapa anterior. En este trabajo nos enfocamos en la primera etapa del sistema, la segunda fue presentada en [1].

La etapa 1 se encarga de capturar el tráfico de la red, pre-procesarlo y organizarlo en flujos de datos determinando si es normal o puede ser anómalo. En la figura 2 se muestra una representación gráfica de P-SADS.

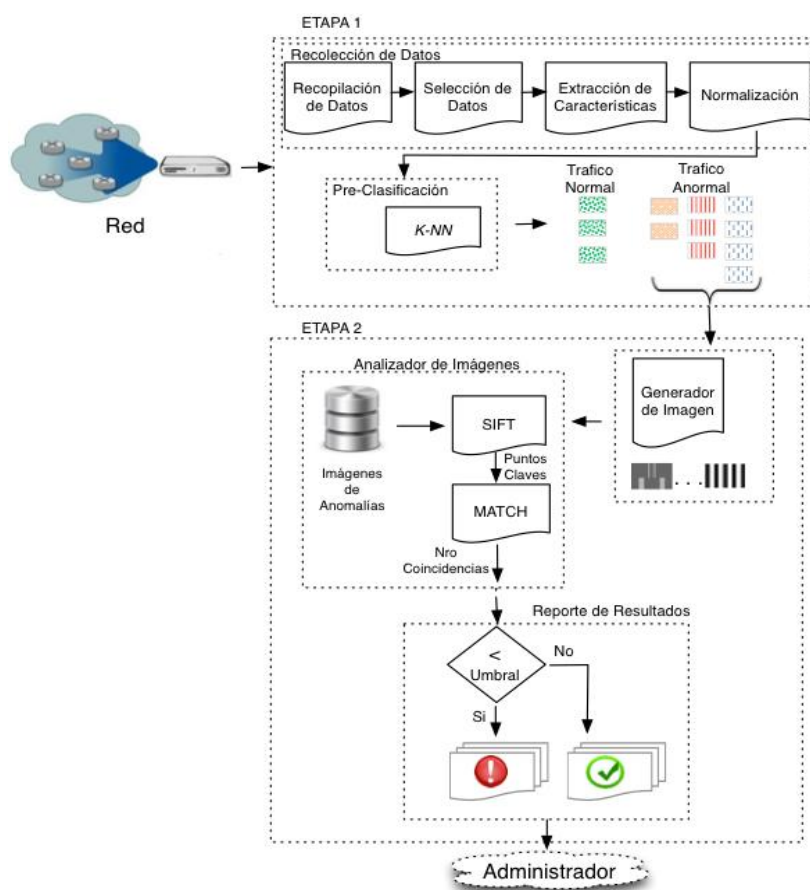


Figura 2. Esquema del modelo propuesto

Las tareas son agrupadas en la *Recolección de Datos* y su *Pre-Clasificación*, cada una con las siguientes características y funciones:

- *Recolección de Datos*: El objetivo es obtener los datos con los cuales se va trabajar, clasificarlos y reducir su volumen para la siguiente etapa. Para lograrlo se realizan en esta tarea los siguientes pasos:
 1. *Recopilación de datos*: Es el proceso de captura de tráfico de red con un sniffer y en momentos específicos donde se asume mayor tráfico de red.
 2. *Selección de datos*: En este paso se seleccionan las tramas a analizar, según los ataques considerados, son tramas TCP, UDP e ICMP.
 3. *Extracción de características*: Extrae de cada trama los campos de interés a analizar, ellos son: dirección IP origen y destino, puerto origen y destino, y tipo de protocolo de comunicación (TCP, UDP o ICMP). La figura 3 muestra ejemplos de tuplas de patrones de tráfico normal (a) y tráfico anómalo (b).

$$f_{normal1} = \{ 10.0.0.1, 212.48.72.19, 31215, 80, UDP \}$$

$$f_{normal2} = \{ 10.0.0.1, 13.29.10.199, 2233, 25, TCP \}$$

$$f_{normal3} = \{ 13.29.10.199, 10.0.0.1, , , ICMP \}$$

(a) Patrones de tráfico normal

$$f_{Smurf} = \{ 10.0.0.1, 10.255.255.255, 31245, 80, ICMP \}$$

$$f_{Land} = \{ 10.0.0.1, 10.0.0.1, 80, 80, TCP \}$$

$$f_{Fraggle} = \{ 10.0.0.1, 10.255.255.255, 31245, 80, UDP \}$$

(b) Patrones de ataques

Figura 3: Ejemplos de Tráfico Normal y Anómalo

4. *Normalización*: Este paso es fundamental, las tuplas se convierten en un vector de valores enteros. La dirección IP $a.b.c.d$ es transformada según la ecuación (1).

$$(a \times 256^3) + (b \times 256^2) + (c \times 256^1) + (d \times 256^0) \quad (1)$$

- **Pre-clasificación de Datos**: Las amenazas no son el único tráfico presente en la red, los paquetes generados en un ataque van acompañados de paquetes de tráfico normal. Este módulo clasifica las tramas obtenidas en el punto anterior como tráfico normal o tráfico anómalo. Para ello, utilizamos el algoritmo de clasificación K-NN, el cual fue implementado en paralelo usando GPU. Su trabajo consiste en comparar cada vector (flujo pre-procesado), con los flujos de datos registrados y representantes de anomalías. Para ello se calcula una distancia a fin de determinar si cada nuevo flujo de datos puede considerarse una anomalía o no.

En la siguiente sección se analiza el desempeño del P-SADS en relación a su efectividad en la detección de anomalías.

4 Experimentación y Análisis de Resultados

En esta sección se presentan los experimentos realizados y el análisis de los resultados de la primera etapa de P-SADS.

La captura del tráfico de datos se hace mediante la herramienta T-Shark. Se trabajó con varias muestras de aproximadamente 2000 tramas cada una (*Recopilación de Datos*). La captura se realizó en la red LAN del Laboratorio de Redes del Departamento de Informática de la Universidad Nacional de San Luis. Se simularon los tres ataques a un servidor, buscando realizar una denegación al servicio HTTP.

Una vez obtenidas las tramas normalizadas, se construyen las bases de datos tanto con tráfico normal como anómalo. El módulo *Pre-clasificación* usando K-NN, recibe como datos de entrada estas bases de datos y realiza la evaluación para diferentes valores de $K = 5, 7, 10$ y 12 .

Para el cálculo de K-NN se usó una computadora con una GPU Tesla K20c (2496 procesadores), 4.6 GB de Memoria y 706 MHz de Frecuencia de reloj y 2600 MHz de memoria del procesador.

Una vez obtenidos los K-NN con los distintos valores, se realizó la evaluación de los resultados en base a las siguientes métricas: *Valor Predictivo Positivo (Precision-PPV)*, la *Tasa de Verdaderos Positivos (Recall-TPR)* y *F-measure (F)* [2]. Todas estas medidas pueden ser calculadas en base a los siguientes parámetros:

- *Verdadero positivo (TP)*: es el número de tramas anómalas detectadas correctamente.
- *Falso positivo (FP)*: número de tramas normales consideradas como tramas anómalas.
- *Falso negativo (FN)*: cantidad de tramas anómalas no detectadas.
- *Verdadero negativo (TN)*: número de tramas normales detectadas correctamente.

A partir de estos parámetros, las métricas se definen como:

- *Valor Predictivo Positivo*: Representa la posibilidad de que una muestra etiquetada como positiva sea efectivamente un verdadero positivo, es el porcentaje de tramas detectadas anómalas. Matemáticamente se define como:

$$PPV = \frac{TP}{(TP + FP)}$$

- *Tasa de Verdaderos Positivos o Sensibilidad*: Es el porcentaje de tramas anómalas detectadas correctamente entre todas las analizadas por el clasificador. Se calcula según:

$$TPR = \frac{TP}{(TP + FN)}$$

- *F-measure (F)*: es una métrica típica de recuperación de la información, se define como la media armónica entre las dos métricas explicadas anteriormente: *Precision* y *Recall*. Se utiliza el valor de medida F como una métrica del desempeño de K-NN, es decir el desempeño en la detección de anomalías. Su fórmula es:

$$F = 2 \times \frac{(PPV \times TPR)}{(PPV + TPR)}$$

Los resultados obtenidos con cada métrica se visualizan en la figura 4, para cada uno de los ataques considerados *Smurf* (a), *Land* (b) y *Fraggle* (c).

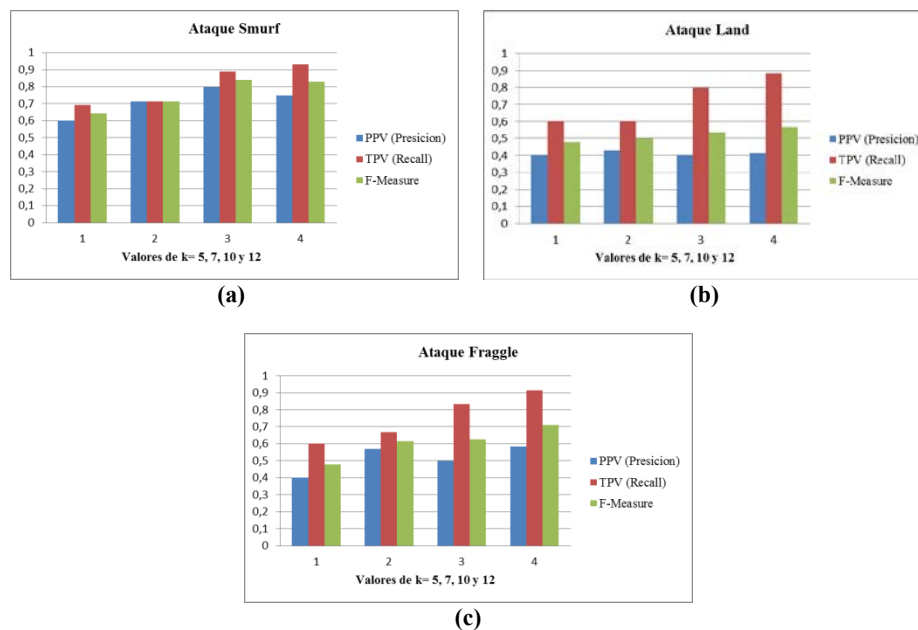


Figura 4: Resultados obtenidos en los ataques *Smurf*, *Land*, *Fraggle*.

De las gráficas anteriores se puede deducir que el ataque *Smurf* tiene una precisión aproximada del 72%, el ataque *Land* un 41% y el ataque *Fraggle* el 52%.

Con respecto a la métrica *recall* se obtuvieron resultados significativos, en promedio en *Smurf* es del 80%, en *Land* es un 72% y en *Fraggle* es un 75%. Esta métrica permitió inferir que la tasa de detección de tráfico anómalo es alta, particularmente cuando $k=12$.

Por último, la métrica *F-measure* también retornó resultados representativos, en particular para $K=12$ en los tres tipos de ataque, obteniéndose un valor de 0,83 en *Smurf*; 0,56 en *Land* y 0,71 en *Fraggle*, significando que nuestro modelo tiene un buen desempeño respecto a la detección de ataques.

Por lo tanto, en base a las métricas analizadas podemos concluir que P-SADS cumple satisfactoriamente con los objetivos planteados en este trabajo.

5 Conclusiones y Trabajos Futuros

La detección de tráfico anómalo es una tarea de gran interés en la actualidad y si bien existen diversas herramientas para su detección, existe mucho trabajo por realizar

debido a la gran cantidad de datos circulantes en Internet y el constante cambio de perfil de tráfico.

Este trabajo se enfocó en la etapa de clasificación del tráfico; se propuso un modelo y se analizó su factibilidad para tres ataques conocidos del tipo de denegación de servicios: *Smurf*, *Land* y *Fraggle*. En su implementación se aplicaron técnicas HPC en GPU a fin de acelerar el proceso y obtener resultados en menos tiempo.

Se evaluó la propuesta según diferentes métricas, observándose que el modelo propuesto muestra una precisión entre el 40% y 70%, y sensibilidad entre el 60% y 83% dependiendo del tipo de ataque. Además, se evaluó el desempeño del sistema según *F-measure* obteniendo valores entre 0,5 y 0,83.

Si bien los resultados fueron satisfactorios, es necesario analizar los factores que influyen para obtener diferentes desempeños según el ataque. Otro trabajo futuro es buscar y analizar otros patrones de ataque. También, se pretende comparar nuestro desarrollo con aquellos que usan técnicas de aprendizaje de máquina o herramientas inteligentes, y en caso de ser necesario mejorar la existente.

Referencias

1. Barrionuevo, M., Lopresti, M., Miranda, N., Piccoli, M.: Un enfoque para la detección de anomalías en el tráfico de red usando imágenes y técnicas de computación de alto desempeño. XXII Congreso Argentino de Ciencias de la Computación. CACIC 2016. p. 1166-1175 (2016)
2. Davis J., Goadrich, M.: The relationship between precision-recall and roc curves, in ICML '06: Proceedings of the 23rd international conference on Machine learning. New York, NY, USA: ACM, 2006, pp. 233–240 (2006)
3. Gibson, D.: CompTIA Security+: Get Certified Get Ahead: SY0-201 Study Guide Createspace Independent Pub (2009). ISBN 9781439236369.
4. Henao Ríos J. L.: Definición De Un Modelo De Seguridad En Redes De Cómputo, Mediante El Uso De Técnicas De Inteligencia Artificial. Tesis presentada como requisito parcial para optar al título de Magíster en Ingeniería – Automatización Industrial. Universidad Nacional de Colombia. (2012)
5. Lowe, D.: Distinctive image features from scale-invariant keypoints. International journal of computer vision. Pp 91-110, (2004)
6. Miranda, N.: Cálculo en Tiempo Real de Identificadores Robustos para Objetos Multimedia Mediante una Arquitectura Paralela GPU-CPU. Tesis de Doctorado en Ciencias de la Computación. UNSL (2014).
7. Piccoli María F.: Computación de alto desempeño de GPU. 1era edic. ISBN: 9789503407592. La Plata Edulp, (2011)
8. S. Institute, Transmission Control Protocol: DARPA Internet Program Protocol Specification. Defense Advanced Research Projects Agency, Information Processing Techniques Office, (1981).
9. Tribak Hind. Análisis Estadístico de Distintas Técnicas de Inteligencia Artificial en Detección de Intrusos. Tesis Doctoral. Universidad de Granada. (2012).
10. Wang Y.: Statistical Techniques for Network Security: Modern Statistically-Based Intrusion Detection and Protection, Chapter III Network Traffic and Data, Information Science Reference - Imprint of: IGI Publishing, (2008).

IX Workshop Innovación en Sistemas de Software (WISS)

Coordinadores

Fillottrani, Pablo – Zanarini, Dante – Finocchietto, Jorge –
Estayno, Marcelo

Comité Científico

Bursztyn, Andrés (Argentina)	Montejano, Germán (Argentina)
Capobianco, Marcela (Argentina)	Riesco, Daniel (Argentina)
Estayno, Marcelo (Argentina)	Sposito, Osvaldo (Argentina)
Fillottrani, Pablo (Argentina)	Vitturini, Mercedes (Argentina)
Flores, Andrés (Argentina)	Vivas, Luis (Argentina)
Meles, Judith (Argentina)	Zanarini, Dante (Argentina)

"Informática UNLP" la App de la Facultad de Informática

Juan Fernández Sosa¹, Alfonso Cuitiño¹, Pablo Thomas¹, Lisandro Delía¹, Germán Cáseres¹, Leonardo Corbalán¹, Patricia Pesado^{1,2}

¹ Instituto de Investigación en Informática LIDI
Universidad Nacional de La Plata,

² Comisión de Investigaciones Científicas de la Pcia. de Buenos Aires (CIC)
La Plata Buenos Aires, Argentina

{jfernandez, acuitino, pthomas, ldelia, gcaseres, corbalan, ppesado}@lidi.info.unlp.edu.ar

Resumen. Las aplicaciones móviles facilitan las actividades de las personas. Diariamente aparecen nuevas aplicaciones con funcionalidades novedosas, imposibles de implementar en sistemas de escritorio, pero factibles y de gran utilidad en dispositivos móviles. Particularmente en este trabajo se presenta una aplicación móvil (*app*) desarrollada para la comunidad de la Facultad de Informática de la Universidad Nacional de La Plata. La *app* "Informática UNLP" es una herramienta de *software* en continuo crecimiento, que mejora la comunicación entre alumnos y docentes, facilitando además el acceso a la información relevante sobre la Facultad y sus carreras.

Keywords: Aplicaciones móviles multiplataforma, dispositivos móviles, cartelera virtual.

1 Introducción

El uso masivo de dispositivos móviles ha generado un nuevo mercado para desarrolladores de *software* y al mismo tiempo nuevos desafíos para mejorar la vida cotidiana de las personas.

Actualmente cada nuevo emprendimiento de desarrollo de *software*, debe contemplar la necesidad o conveniencia de generar la correspondiente versión para dispositivos móviles.

El panorama ha cambiado para la Ingeniería de Software. El desafío de generar aplicaciones móviles ha provocado la necesidad de utilizar nuevas herramientas de desarrollo, pensar en nuevas funcionalidades, integrar aplicaciones existentes con nuevos productos; todo esto en tiempos acordes a las necesidades de clientes y usuarios.

La tecnología, en resumen, brinda nuevas posibilidades que conducen a pensar continuamente en innovar.

La República Argentina está entre los primeros 25 países del mundo con mayor cantidad de usuarios de *smartphones* [1]. Esta particularidad debe ser aprovechada.

La portabilidad y la ubicuidad de una aplicación móvil constituyen dos características insoslayables.

En ese sentido, la Facultad de Informática de la Universidad Nacional de La Plata intenta ser pionera al ofrecer nuevos servicios a su comunidad educativa a través de una aplicación móvil institucional. Así es como surge "Informática UNLP", una *app* con servicios novedosos, muy útiles para estudiantes y docentes.

En la sección 2 se presentan las características de la *app* "Informática UNLP" y sus principales funcionalidades. En la sección 3 se describe la tecnología utilizada para su desarrollo. Luego, en las siguientes secciones, se presentan una breve estadística de uso, las conclusiones y trabajo futuro.

2 Características de "Informática UNLP"

Considerando que los alumnos son naturalmente usuarios experimentados de aplicaciones móviles, el propósito de la *app* "Informática UNLP" es constituir un canal de comunicación Institucional efectivo, adicional a los ya existentes (sitio *web*, Facebook, Twitter, entre otros), allanando el paso de los alumnos por la Institución durante el transcurso de su carrera universitaria.

De acuerdo a su condición de *app* institucional, "Informática UNLP" refleja la identidad visual de la Facultad de Informática. En la Figura 1 se presenta la interface principal de la *app*.

Cabe destacar que la *app* accede a la información provista por un sistema *web* denominado Gestión Docente, en el cual se gestionan las aulas y horarios de la Facultad, junto a las actividades docentes.

A continuación se detallan las principales funcionalidades que ofrece la *app*.

2.1 Planes de Estudio

A través de esta función se pueden consultar los planes de estudio de las cuatro carreras que ofrece la Unidad Académica: Licenciatura en Informática, Licenciatura en Sistemas, Ingeniería en Computación y Analista Programador Universitario.

2.2 Calendario Académico

Esta función permite tener una rápida visualización de las fechas de inicio y fin de cursadas, fechas de inscripciones, fechas de finales, entre otros datos de interés para los alumnos.

2.3 Finales

Con esta operación se pueden consultar los días y horarios habituales de los exámenes finales de cada asignatura.

2.4 Noticias

Esta función muestra las noticias relevantes para los alumnos en el ámbito de la Facultad.

Cabe destacar que la Dirección de Difusión de la Facultad posee un sistema de *software* especialmente diseñado para publicar las noticias que posteriormente serán accesibles desde la *app*.

Al publicarse una nueva noticia, se envía una notificación a todos los usuarios que tengan la *app* instalada.

2.5 Ahora en la Facultad

Esta operación permite consultar en tiempo real lo que está sucediendo en cada una de las aulas de la Facultad. La información es obtenida a través de la API del Sistema Gestión Docente. En la figura 2 se presenta el resultado generado por esta consulta en un día y hora determinados.



Fig. 1. Informática UNLP



Fig. 2. Ahora en la Facultad

2.6 Aulas

Con esta función se habilita la cámara del dispositivo móvil para escanear un código QR.

La puerta de cada aula de la Facultad de Informática posee un código QR que permite identificarla. Al enfocar la cámara del dispositivo sobre este código, luego del correspondiente reconocimiento, la *app* responde en tiempo real informando el nombre de la actividad que se está llevando a cabo en ese momento en el Aula. Además se muestran las actividades previas y posteriores, de acuerdo al horario en que se haya realizado la consulta.

La figura 3 muestra el resultado de escanear el código QR del aula 4, el día 16 de Mayo de 2017 a las 10:04 hs. Es de notar que las actividades posteriores se presentan sombreadas para que se distinga claramente la actividad actual, y que por la hora en que se realizó la consulta, no hubo actividades previas en ese día.



Fig. 3. Resultado de escanear el código QR frente al aula 4

2.7 Cartelera

Los docentes a través de un sistema de *software* específico de Gestión, actualizan una cartelera virtual que posee cada asignatura. Esta cartelera virtual está disponible de forma pública en el mismo sistema.

Por otra parte, el usuario de la *app* puede configurar las asignaturas que formarán parte de su cartelera personal, indicando en cada caso si se desea recibir una notificación cada vez que la asignatura publique una novedad.

Cuando un docente publica una novedad sobre alguna asignatura, esta información es visible desde la cartelera en la *app*. Aquellos usuarios que hayan configurado la recepción de notificaciones de dicha asignatura, serán notificados en tiempo real sobre el nuevo aviso añadido a su cartelera, pudiendo acceder entonces, de manera inmediata, al contenido de la novedad generada.

La figura 4 presenta la edición de un mensaje desde el sistema *software* que utilizan los docentes, mientras que la figura 5 muestra la cartelera virtual de una asignatura en la *app*.

2.8 Horarios Materias

A través de esta operación se pueden consultar los horarios de las asignaturas de cada una de las carreras de la facultad. La Figura 6 reproduce un ejemplo de los horarios de la asignatura Bases de Datos 2.

Facultad de Informática
Gestión Docente

UNIVERSIDAD NACIONAL DE LA PLATA

/ Modificar

Editar Mensaje de cartelera

Datos

Materia
Fundamentos de Organización de Datos (Primero)

Titulo
Teorias

Cuerpo
Se publicaron todos los slides de las teorías 2017 de FOD en la Plataforma Ideas.

Al dar de alta o modificar un mensaje se generará una notificación a los dispositivos móviles registrados con la materia.

Secretaría de Modernización de la Gestión
Dirección de Sistemas Facultad de Informática - UNLP

Fig. 4. Edición de un mensaje que será enviado a la *app* Informática UNLP



Fig. 5. Cartelera virtual de una asignatura



Fig. 6. Horarios de la asignatura Bases de Datos 2

2.9 Sugerencias

Desde el menú lateral se puede enviar sugerencias al equipo de desarrollo de la *app*.

3 Tecnología Utilizada

La *app* "Informática UNLP" consta de dos capas, una *front-end* y una *back-end*. La primera de ellas, es la capa de presentación con la cual el usuario interactúa: la aplicación móvil. Desde el *back-end* se mantiene el contenido que es consumido por el usuario desde la aplicación móvil. Por otro lado, la aplicación también se alimenta de información suministrada por el sistema de Gestión Docente de la Facultad de Informática [2], en donde se administra todo lo referido a las reservas de aulas y se generan mensajes en una cartelera virtual para cada asignatura.

Para el desarrollo de la *app* "Informática UNLP" se utilizó el *framework* de código abierto Ionic [3], el cual utiliza un enfoque de desarrollo multiplataforma del tipo híbrido [4].

Ionic se basa en el *framework* de desarrollo de aplicaciones móviles Apache Cordova [5], que permite usar tecnologías *web* estándar como HTML5, CSS3 y Javascript. Además, Ionic incorpora AngularJS [6], un *framework* para facilitar el desarrollo en capas.

La decisión de utilizar el *framework* Ionic se basó esencialmente en dos características: por un lado es multiplataforma, y por el otro al ser un enfoque híbrido, las aplicaciones generadas tienen muy buena performance en comparación con las generadas por otros *frameworks* multiplataforma [7].

Para proveer algunas funcionalidades como el acceso al *hardware* de los dispositivos, las notificaciones o la descarga de información desde Internet, se utilizaron una serie de *plugins* (CordovaHTTP [8], Barcode Scanner [9], entre otros).

El *back-end* fue desarrollado utilizando Symfony 2 [10], diseñado para optimizar el desarrollo de las aplicaciones *web* basado en el patrón Modelo-Vista-Controlador. También se utilizó MySQL como motor de base de datos.

Desde el *back-end* se cargan las novedades de la Facultad que quieren ser comunicadas a los usuarios de la aplicación por medio de notificaciones *push*. Éstas se implementan usando el servicio de Google GCM (Google Cloud Messaging).

GCM es un servicio para notificaciones móviles desarrollado por Google. Este servicio permite enviar mensajes desde el servidor de la aplicación a cada uno de los dispositivos que tengan la aplicación instalada y hayan activado la opción de recibir novedades.

Cuando el usuario instala la aplicación en su dispositivo móvil, se realiza una solicitud al servidor de GCM y éste devuelve un código identificador. Luego, este código se almacena en el servidor *back-end* que fue desarrollado para la aplicación. Cuando se quiere enviar una notificación, se genera un mensaje desde el servidor *back-end*, y se lo envía al servidor de GCM, el cual se encarga de distribuir la notificación a cada uno de los dispositivos móviles registrados.

Para la sección de cartelera de la aplicación móvil, se usa otra funcionalidad propia de GCM, los "*topics*" o "temas". Cada asignatura dictada en la Facultad es considerada un "tema".

Cuando un usuario selecciona la opción de recibir notificaciones en la cartelera, se envía una solicitud al servidor con el nombre de la asignatura y el identificador del usuario. Esta solicitud es reenviada al servidor de GCM. Cuando desde el sistema de gestión docente de la Facultad de Informática se genera un aviso en la cartelera virtual de una asignatura, se da aviso al servidor *back-end* de la aplicación. Este último se encarga de propagar el aviso al sistema de GCM, el cual envía una notificación a cada uno de los usuarios que se hayan registrado para esa asignatura. La arquitectura de este procedimiento se reproduce en la figura 7.

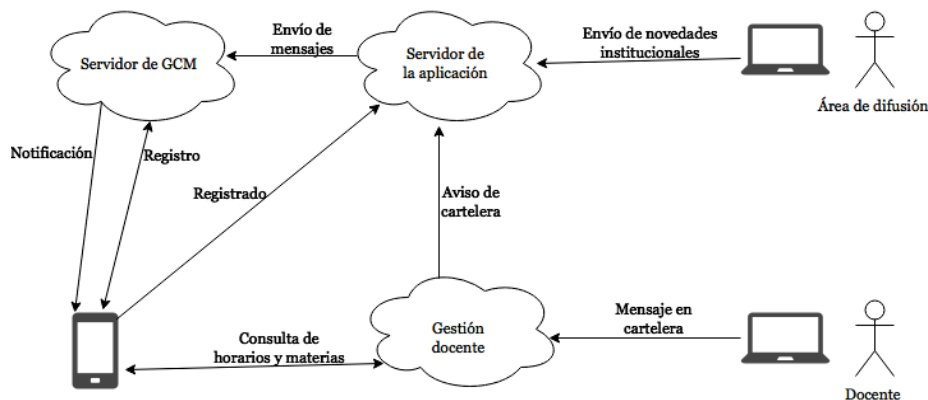


Fig. 7: Arquitectura del procedimiento de notificaciones

4 Estadística de uso

A continuación se presentan algunos valores que describen el uso de la *app*. Desde su inicio a mediados del año 2016 hasta Mayo 2017 se registraron 2056 instalaciones.

En relación a su cartelera virtual, 33 asignaturas la han utilizado, registrándose 21 publicaciones en la cartelera virtual durante Mayo 2017.

Por otra parte, se han recibido sugerencias de los usuarios relacionadas con mejoras posibles, algunas de las cuales se encuentran en desarrollo.

5 Conclusiones y trabajo futuro

Se ha desarrollado una aplicación móvil que brinda servicios a la comunidad de la Facultad de Informática de la UNLP, especialmente a sus alumnos.

Esta aplicación no intenta reemplazar a los otros canales de comunicación y servicios ya existentes, sino que actúa como un complemento, ofreciendo algunas ventajas adicionales. Por ejemplo, es normal suponer que si un alumno desea informarse sobre alguna asignatura, debe acceder al sitio *web* de la misma, a su plataforma virtual, o de manera presencial a la cartelera física disponible en la Facultad. A partir del año 2017, con la *app* Informática UNLP, la cartelera de una asignatura "busca" al alumno a través de las notificaciones correspondientes. Es decir, los alumnos reciben las novedades, no necesitan ir a buscarlas.

Claramente esta funcionalidad más otras descritas en este trabajo, fortalecen el vínculo entre los estudiantes y la Institución, por lo que la *app* Informática UNLP solo ofrece innovación que la tecnología móvil permite.

Como trabajo futuro se prevé agregar más funcionalidades a la *app*. Una de las más destacadas, en vías de desarrollo, es la que permitirá al alumno armar su cronograma de horarios de cursada (a partir de la consulta de horarios por asignatura ya disponible), para que la *app* le notifique unos minutos antes, sobre la asignatura que está por comenzar.

Referencias

1. <https://www.emarketer.com/Article/2-Billion-Consumers-Worldwide-Smartphones-by-2016/1011694> consultado el 15/05/2017
2. <http://gestiondocente.info.unlp.edu.ar>
3. <https://ionicframework.com/>
4. Lisandro Delía; Nicolás Galdámez; Pablo Thomas; Leonardo Corbalán; Patricia Pesado. *Multi-platform mobile application development analysis*. 2015 IEEE 9th International Conference on Research Challenges in Information Science (RCIS), ISBN 978-1-4673-6630-4.
5. <https://cordova.apache.org>
6. <https://angularjs.org/>
7. Lisandro Delía, Nicolás Galdamez, Leonardo Corbalan, Pablo Thomas, Patricia Pesado, *Un Análisis comparativo de rendimiento en Aplicaciones Móviles Multiplataforma*, CACIC 2015, UNNOBA, Junín. ISBN: 978-987-3724-37-4
8. <https://www.npmjs.com/package/cordova-plugin-http>
9. <https://www.npmjs.com/package/cordova-plugin-barcode-scanner>
10. <http://symfony.es/documentacion/>

Desarrollo de un Planificador Embebido para la Gestión de Riego Automático aplicado a invernáculos. Implementación en INTA El Sombrero

José Manuel Alonso¹, Hugo Ferrari², Iván Sambrana¹, Emanuel Irrazábal¹,

¹ Departamento de Informática,
Facultad de Ciencias Exactas y Naturales y Agrimensura,
Universidad Nacional del Nordeste

² Departamento de Ingeniería,
Facultad de Ciencias Exactas y Naturales y Agrimensura,
Universidad Nacional del Nordeste
eirrazabal@exa.unne.edu.ar

Abstract. Las nuevas tecnologías son ampliamente utilizadas en la producción agrícola. Entre ellas, los sistemas de riego automatizados permiten mejorar el rendimiento de los cultivos y es un insumo para la investigación agropecuaria. En este trabajo se describe el desarrollo de un Planificador Embebido para la Gestión de Riego Automático en el marco de las necesidades del grupo de investigación de INTA El Sombrero - Corrientes. Se construyó un sistema basado en tecnología Arduino y una plataforma web de configuración. La solución tuvo en cuenta el riego manual y automático; y una arquitectura software-hardware ampliable. Se trabajó con una metodología iterativa-incremental y un equipo interdisciplinario de expertos en las áreas de sistemas de información, electrónica y la comunicación constante con el grupo de investigación. Esto facilitó el desarrollo del prototipo y el cumplimiento de los requisitos del sistema, el cual está siendo utilizado en un invernáculo dedicado a investigación.

Keywords: planificador automático de riego, INTA, Arduino.

1 Introducción

La actividad agrícola es una de las principales fuentes de producción de alimentos en el ámbito mundial. Y en particular, para la economía Argentina esta actividad aporta un 9% de su Producto Bruto Interno [1]. En este sentido, la agricultura necesita recursos básicos para su desarrollo: agua y energía eléctrica [2]. Por otro lado el incremento de la población conlleva una mayor demanda alimenticia. Esto plantea la necesidad de hacer más eficientes los procesos de producción para maximizar el rendimiento de los cultivos [2].

Las nuevas tecnologías de la computación y la comunicación ofrecen soluciones que pueden adaptarse a las necesidades de la agricultura para una producción más eficiente en el uso de recursos y por lo tanto más sustentable económica, ambiental

y socialmente. Entre ellas se encuentran los desarrollos relacionados con la automatización del riego. Desde hace varios años se han implementado sistemas de riego que permiten un uso más racional del agua [2][3][4][5][6][7]. La eficiencia en la aplicación y la productividad del agua de riego son temas prioritarios, teniendo en cuenta que la agricultura bajo riego es la mayor demandante, en la competencia con los otros usos (urbanos, recreativos, industriales y energéticos), utilizando aproximadamente el 70 % agua dulce que se consume en todas las actividades [8].

Para los diferentes cultivos existen requerimientos particulares para lograr un óptimo desarrollo y productividad. Cada fase del desarrollo del cultivo requiere un mínimo de acumulación de temperatura, así como necesidades hídricas y nutricionales que varían a lo largo de cada fase [9]. Esto hace indispensable tener un control asistido de estos parámetros. En diversas zonas de Argentina existe una gran variabilidad climatológica, como por ejemplo en el Nordeste (NEA) y, particularmente en la provincia de Corrientes. Esta es una región mayormente subtropical, con veranos cálidos y húmedos e inviernos templados y secos [10]. Además, las previsiones sobre los efectos del cambio climático para regiones subtropicales como el NEA, indican que las temperaturas serán cada vez más altas en verano, llegando a ser extremas y más bajas en invierno, con heladas y ocasionalmente caídas de granizo [11]. Esto afecta negativamente a los cultivos en el campo, promoviendo el uso de tendaleros con riego artificial controlado [12]. Pero no solo tiene utilidad en el marco de las actividades productivas, sino que también es necesario para las actividades de investigación en el ámbito agropecuario.

El Instituto Nacional de Tecnología Agropecuaria (INTA) desarrolla diferentes líneas de investigación para mejorar el desarrollo de los cultivos [13]. Y uno de los insumos necesarios es asegurar el control de las variables ambientales en los laboratorios e invernáculos. Por lo tanto, la inclusión de los sistemas de riego automatizados es un insumo para el desarrollo de actividades de investigación agropecuaria.

Ahora bien, los esquemas de riego estáticos tradicionales no se ajustan adecuadamente a las necesidades hídricas de los cultivos en producción o de los ensayos en el ámbito de investigación [14].

Todo esto ha motivado el siguiente trabajo, en el cual se describe el proyecto y posterior desarrollo de un Planificador Embebido para la Gestión de Riego Automático (PEGRA) para uso en investigación y como prototipo para uso bajo tendalero; en el marco del grupo de forrajes de INTA El Sombrero ubicado en la provincia de Corrientes.

El artículo está compuesto por 5 secciones. Además de la sección introductoria, en la segunda sección se detalla el análisis del problema y el diseño de la solución teniendo en cuenta otros desarrollos de sistemas de riego. En la sección 3 se describe la solución obtenida. En la sección 4 se describe el uso de PEGRA en un invernáculo de investigación en INTA. Finalmente en la sección 5 se describen las conclusiones y los trabajos futuros.

2 Diseño de la solución

El presente trabajo se ha desarrollado a partir de la construcción de un planificador de riego automatizado para uso académico de acuerdo con las necesidades de la Estación Experimental INTA – El Sombrero situada a 30 Km. al sur de la ciudad de Corrientes. Este prototipo se utilizará en el marco de los proyectos “Generación y evaluación de poblaciones de híbridos de *Setaria sphacelata* y *Acroceras macrum*, dos especies forrajeras para el nordeste argentino” y “Desarrollo de cultivares superiores de especies forrajeras para sistemas ganaderos y agrícola-ganaderos de la Argentina” en los que participa la Dra. Silvana Ferrari, haciendo un uso intensivo de invernáculos para el desarrollo de sus experimentos. De acuerdo a lo expuesto por la Dra. Ferrari existe una baja utilización de sensores y herramientas informáticas lo que dificulta conocer las necesidades hídricas de los cultivos y plantear una correcta estrategia de riego.

A partir de una serie de reuniones mantenidas con los integrantes del grupo de investigación se consensuaron los requisitos del PEGRA (ver Tabla 1). En la primera columna de la tabla se incluye un identificador del requisito.

Tabla 1. Requisitos Funcionales del PEGRA.

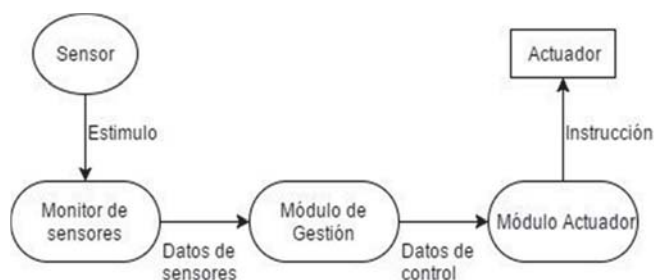
Ref.	Título	Descripción
RF_1	Riego estático	El sistema regará por tiempo de forma periódica.
RF_2	Riego Automático	El sistema mantendrá automáticamente la humedad del suelo dentro de un rango definido.
RF_3	Gestión remota	El sistema se configurará de forma remota.
RF_4	Sincronización con servidor Web	El sistema se sincronizará con el servidor web para obtener las configuraciones de riego y enviar datos del ambiente y logs del sistema.
RF_5	Sectorización	El sistema gestionará diferentes parcelas de un invernáculo.
RF_6	Perdurabilidad de los datos	El sistema guardará los datos de configuración, logs y censado en un sistema gestor de base de datos.
RF_7	Redundancia de datos	El sistema guardará los datos en una tarjeta SD para que no se pierda información si la conexión falla.
RF_8	Control de acceso	El sistema permitirá el acceso al servicio web solo a los usuarios registrados.
RF_9	Visualizador web del estado de los sectores	El usuario visualizará el último estado guardado de las variables ambientales del sector.
RF_10	Visualizador web del historial del sector	El usuario visualizará un gráfico de los cambios en las variables climatológicas del sector entre dos intervalos de tiempo en un rango de horas y días.

Teniendo en cuenta estos requisitos se continuó con el diseño de la solución. Para ello se analizaron trabajos similares a nivel regional e internacional. En la Tabla 2 se resumen estos trabajos teniendo en cuenta el tipo de riego desarrollado y el equipamiento seleccionado para construir la solución.

Tabla 2. Resumen de soluciones de riego.

Referencia	Tipo de Riego	Controlador	Sensor	Actuador
[6]	Goteo	Arduino UNO	SEN92355P	Bomba de agua
[5]	Humidificación	Arduino UNO	DHT11	Humidificador
[3]	Goteo	Arduino UNO y Raspberry PI	YL-69, DHT11	Electroválvula y bomba de agua
[4]	Aspersión	Arduino UNO y Raspberry PI	DHT22, FC-37	Electroválvula
[7]	Goteo	pic16f877a	RM1500, YL-69	Electroválvula

De acuerdo a lo analizado en otras soluciones, los sistemas de riego automatizado cuentan, en general, con 3 subsistemas. Por un lado el Módulo de monitoreo de sensores con los que diferentes sensores (especialmente de humedad) obtienen datos del medio. Por otro lado el Módulo de Gestión, donde se procesan las mediciones y, opcionalmente, se obtienen los datos de control, como por ejemplo, la humedad deseada. Y finalmente el Módulo actuador que controla el riego. En la Fig. 1 se puede ver un esquema sencillo de los tres Subsistemas. Este tiene un diseño análogo al modelo general de un sistema embebido de tiempo real presentado [15].

**Fig. 1.** Modelo general de un sistema embebido.

A modo de resumen, en las soluciones descritas en la Tabla 2, Arduino UNO es el controlador más utilizado para el desarrollo de módulos de monitoreo de sensores y módulos actuadores. El método utilizado para determinar el contenido de humedad en el suelo es la medición de la resistividad eléctrica [16]. Y el tipo de riego utilizado en la mayoría de las soluciones es el riego localizado, siendo el riego por goteo el que presenta mayor eficiencia respecto del uso del agua.

3 Desarrollo de la solución

De acuerdo con lo descrito anteriormente, se ha seleccionado una solución del tipo arquitectura cliente-servidor que estará compuesta por diferentes módulos (ver Fig. 2). Por un lado se encuentra el Módulo de Gestión. Por otro lado los Módulos de Servicios Web (Interfaz de usuario e Interfaz con el controlador) estarán alojados

en un servidor de tecnología Linux y procesarán peticiones de los clientes y comunicarse con el Módulo de base de datos. Finalmente el Módulo de base de datos será el que guarde la información obtenida por los sensores y las configuraciones generadas por los usuarios.

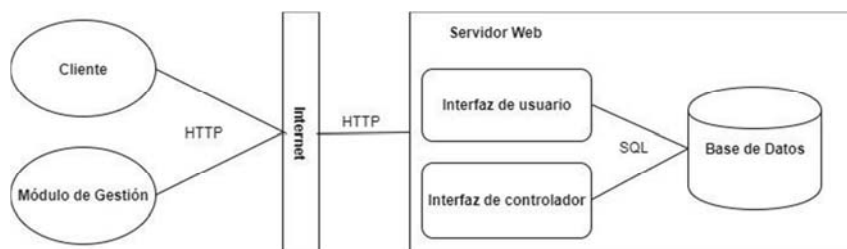


Fig. 2. Arquitectura cliente-servidor del PEGRA.

En la Fig. 3 se indican los procesos que componen el Módulo de Gestión. El Monitor de humedad del suelo será el encargado de tomar mediciones periódicas de los sensores YL-69, los valores obtenidos serán comparados en el proceso de control de humedad con los parámetros de humedad máxima y mínima establecidos por el usuario. El proceso de control de válvula es el encargado de encender o apagar la válvula de agua. Por otro lado el proceso de comunicación se encarga de enviar de forma periódica los datos procesados por el control de humedad y los eventos registrados en el monitor de estado de válvula. Además este proceso establece una comunicación periódica con la interfaz de controlador del servidor Web para obtener los ajustes de riego de la base de datos.



Fig. 3. Arquitectura del módulo de gestión del PEGRA.

Para el Módulo de Gestión se ha seleccionado el microcontrolador Arduino UNO. De acuerdo con las referencias estudiadas en la Tabla 2, este microcontrolador ha demostrado ser confiable. Además se utilizarán sensores YL-69 para realizar la medición de humedad del suelo por resistividad eléctrica. Para las mediciones de temperatura y humedad relativa del ambiente se utiliza el sensor DHT11, este sensor es apto para las condiciones ambientales de la región ya que el rango de temperaturas que puede obtener oscila entre los 0 y 50 grados Celsius con

una precisión de ± 2 grados [17]. También se utilizará una electroválvula Rain Bird 100 DVF accionada por un módulo Relay.

3.1 Metodología de desarrollo

Para la construcción de la solución se ha seguido una metodología Iterativa e Incremental. En un desarrollo iterativo e incremental el proyecto se planifica en diversos bloques temporales llamados iteraciones. Un aspecto fundamental para guiar el desarrollo iterativo e incremental es la priorización de los objetivos/requisitos en función del valor que aportan al cliente [18].

Teniendo esto en cuenta se han planificado cinco fases incrementales, tal y como se describe a continuación:

En la primera fase se ha realizado la planificación, el desarrollo y testeo del prototipo del Módulo de gestión de riego estático. Esta funcionalidad permite al usuario configurar la gestión de riego por tiempo.

En una segunda fase se ha diseñado la estructura y se ha creado la base de datos relacional con la cual mantener los datos de gestión. En la tercera fase se diseñó y desarrolló la interfaz entre la base de datos y el Módulo de Gestión con la tecnología API Rest [19].

En la cuarta fase se ha diseñado, desarrollado y probado un nuevo prototipo del Módulo de Gestión que incorpora las funcionalidades de gestión automática de riego. En una quinta y última fase se ha desarrollado la interfaz web de usuario. Esta interfaz permite que el usuario pueda gestionar la configuración de riego y ver los valores actuales de las variables del microclima dentro del invernáculo, así como también los datos históricos del mismo.

4 Implementación del planificador en INTA – El Sombrero

Una vez construido el prototipo funcional se planificó su uso en el ámbito académico de la Estación Experimental INTA – El Sombrero. Como se ha descrito en la sección 2, este prototipo se utilizará en el marco de los proyectos de investigación del grupo de forrajes, en sus invernáculos experimentales.

La implantación del prototipo fue realizada en 4 pasos. En el primer paso se instalaron las tuberías, válvulas y aspersores de agua. Se diseñó una estructura formada por 3 secciones de caños de polipropileno, válvulas y aspersores conectados a una fuente de agua suministrada por un tanque de 40 metros de altura. Uno de los puntos a tener en cuenta aquí fue asegurar el caudal y la presión del agua para lograr un riego eficaz. Para ello se realizaron los cálculos correspondientes a la demanda de presión y caudal [20] para cada sección como se muestra en la Tabla 3.

Tabla 3. Cálculo de la demanda de presión y caudal de las tuberías.

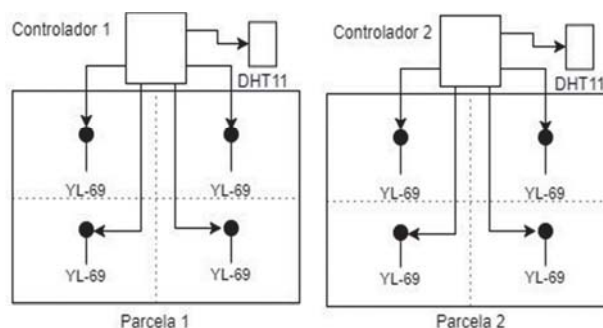
Tubería porta emisores	Sección 1	Sección 2	Sección 3
Presión de operación	2 bares	2 bares	2 bares
Diámetro de la tubería	½"	1"	1"
Caudal	0,88 m ³ /h	2,51 m ³ /h	2,51 m ³ /h
Velocidad	1,25 m/s	1,25 m/s	1,25 m/s

En el segundo paso de implantación del prototipo se instalaron los elementos de comunicación para el proceso de gestión de la comunicación descrito en la Fig. 3. Para asegurar la correcta comunicación entre el módulo interfaz de controlador y la placa controladora se utilizó una conexión cableada Ethernet con UTP a un Router conectado a Internet por Banda Ancha.

El paso 3 fue el montaje de la placa Arduino junto la fuente de alimentación dentro de un panel protector, para evitar la exposición a la humedad y la radiación solar que reducen la vida útil de los componentes.

Finalmente, en el cuarto paso se instaló el panel con la placa Arduino en el invernáculo junto con el sensor DHT11 y los sensores de humedad del suelo. La elección del esquema (Ver Fig. 4) sigue las siguientes premisas:

- Obtener medidas representativas de la humedad del suelo. Para esto se dividió la parcela en cuadrantes y se ubicó un sensor YL-69 en el centro de cada cuadrante.
- Obtener temperatura y humedad del ambiente dentro del invernáculo: El sensor DHT11 se instaló en la parte más alta del invernáculo.

**Fig. 4.** Distribución de sensores

La gestión web se realizó configurando las parcelas del invernáculo. En este caso fueron dos parcelas. En la primera se probó la configuración de riego automático con los siguientes parámetros: 40% de humedad del suelo como mínima y 60% de humedad del suelo como máxima. El sistema logró estabilizar la humedad del suelo en 50%. En la Fig. 5 –izquierda se pueden ver los resultados. Asimismo, en la segunda parcela se probó la configuración de riego manual (estático) a un ritmo de 30 minutos de riego cada 6 horas. Los resultados pueden verse en la Fig. 5 –derecha. El ritmo de riego manual fue aumentando la humedad relativa hasta el 75% aproximadamente, donde tendió a estabilizarse.

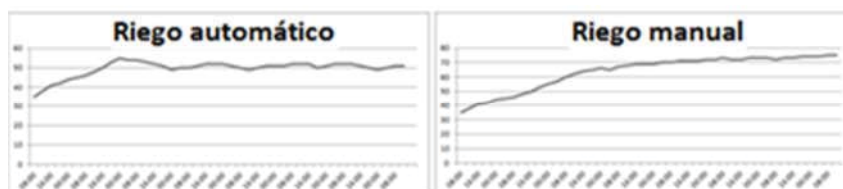


Fig. 5. Resultados de la gestión de riego automático y manual

En la Tabla 4 se analiza el grado de cumplimiento de cada requisito establecido previamente en la Tabla 1 y se incluyen oportunidades de mejora como posibles trabajos futuros.

Tabla 4. Cumplimiento de los requisitos del sistema.

Ref.	Título	Cumplimiento	Oportunidades de mejora
RF_1	Riego estático	Completado	No detectadas.
RF_2	Riego Automático	Completado	No detectadas.
RF_3	Gestión remota	Completado	No detectadas.
RF_4	Sincronización con servidor Web	Completado	No detectadas.
RF_5	Sectorización	Completado	No detectadas.
RF_6	Perdurabilidad de los datos	Completado	No detectadas.
RF_7	Redundancia de datos	No Completado	Incorporar un módulo SD al microprocesador.
RF_8	Control de acceso	No Completado	Mejorar el control de acceso al servicio web. Mejorar la usabilidad.
RF_9	Visualizador web del estado de los sectores	Completado	Posibilidad de graficar los sectores.
RF_10	Visualizador web del historial del sector	Completado	No detectadas.

5 Conclusiones y Trabajos Futuros

A lo largo de este trabajo se construyó el Planificador Embebido para la Gestión de Riego Automático (PEGRA) que tuvo en cuenta dos tipos de riego: manual y automático; y una arquitectura capaz de ser ampliada a mayor cantidad parcelas controladas.

Asimismo, el uso de una metodología Iterativa e Incremental mejoró la comunicación con el usuario final logrando una solución más acorde a sus necesidades. Además el cliente obtuvo prototipos funcionales desde los primeros incrementos del desarrollo.

Para la construcción del prototipo se trabajó con un equipo interdisciplinario, con expertos en las áreas de sistemas de información, electrónica y con una experta en el área de aplicación. Esto facilitó el desarrollo de un prototipo que cuenta con

características adecuadas en cuanto al desarrollo del software, el uso adecuado de los sensores y actuadores, y la configuración adecuada de una interfaz útil.

Los resultados se ajustan a las dos siguientes líneas:

A través del trabajo en equipo interdisciplinario se logró contribuir al conocimiento integral y las relaciones con la naturaleza del problema. Además de contribuir con la autoformación de los miembros del equipo y la participación individual, colectiva e integral en el proyecto [21].

La aplicación de una metodología iterativa e incremental permitió obtener prototipos funcionales desde etapas tempranas del proyecto [18]. Esto además permitió gestionar de manera natural los cambios de requerimientos que surgieron durante el desarrollo de las funcionalidades.

Como trabajo futuro se abordarán las mejoras propuestas por los requerimientos que no se cumplieron conforme se habían propuesto en el inicio del proyecto. Para ello, al finalizar el proyecto se realizó una reunión del tipo Planning Poker [22] para priorizar las oportunidades de mejora de manera grupal [21]. Las dos actividades prioritarias fueron las siguientes: se incorporarán al Módulo de Gestión una tarjeta SD para asegurar la perdurabilidad de los datos y se implementará los roles de usuario para la aplicación web.

Agradecimientos

A la Dra. Silvana Ferrari de la Estación Experimental INTA – El Sombrero y al Proyecto de Investigación "Métodos y herramientas para la calidad del software". PI "F010" SGCyT FaCENA – UNNE.

Referencias

1. Food and Agriculture Organization of the United Nations. www.fao.org. [En línea] 2010. [Citado el: 30 de Mayo de 2017.] <http://www.fao.org/3/a-bp551s.pdf>.
2. Modelo a escala de un sistema de riego automatizado, alimentado con energía solar fotovoltaica: nueva perspectiva para el desarrollo agroindustrial colombiano. MARTÍNEZ, Germán Arturo López, MORENO, Flor García y FIERRO, Juan Bedoya. 2014, Revista Tecnura, Vol. 17, págs. 33-47.
3. Castro Silva, Juan Antonio. Sistema de riego autónomo basado en la internet de las cosas. Neiva, Colombia: s.n., 2016.
4. Escalas Rodríguez, Gabriel. Diseño y desarrollo de un prototipo de riego automático controlado con Raspberry Pi y Arduino. Barcelona, España: s.n., 2014.
5. Escamilla Martínez, Fernando. Automatización y telecontrol de sistema de riego. Gandía, España: s.n., 2016.
6. Hernandez Garcia, C. Itzel. Diseño de un sistema automatizado de riego para el cultivo de tomate y pimiento bajo invernadero. Xalapa Enríquez, México: Diciembre de 2014.
7. An Intelligent Smart Irrigation System Using WSN and GPRS Module. Manimaran , P. . 6, 2016, International Journal of Applied Engineering Research, Vol. 11, págs. 3987-3992.
8. Instituto Nacional de Tecnología Agropecuaria. INTA. [En línea] 2013. [Citado el: 12 de Junio de 2017.] <http://inta.gob.ar/proyectos/PNAGUA-1133043>.

9. Rawson, Howard M. y Gómez Macpherson, Helena. Descripción de los problemas y soluciones - Factores ambientales. Trigo regado. Manejo del cultivo. Roma : Food and Agriculture Organization of the united Nations, 2001.
10. Ministerio del Interior, Obras Publicas y Viviendas. www.mininterior.gov.ar. [En línea] 2016. [Citado el: 02 de 05 de 2017.]
11. INTA. Instituto Nacional de Tecnología Agropecuaria. [En línea] [Citado el: 28 de Junio de 2017.] http://climayagua.inta.gov.ar/que_es_el_fenomeno_el_niño.
12. Magrin, Graciela, y otros, y otros. Instituto Nacional de Tecnología Agropecuaria. [En línea] Abril de 2003. [Citado el: 28 de Junio de 2017.] http://climayagua.inta.gov.ar/sites/default/files/cambiocli/Vulnerabilidad_Produccion_Agricola_Region_Pampeana.pdf.
13. Instituto Nacional de Tecnología Agropecuaria. [Inta.gov.ar](http://inta.gov.ar). [En línea] 2016. [Citado el: 30 de Mayo de 2017.] <http://inta.gov.ar/sites/default/files/intapresentacionespanol.pdf>.
14. Pereira, Luis Santos, y otros, y otros. Riego localizado. El riego y sus tecnologías. Albacete, España : Europa-América, 2010, págs. 229 - 231.
15. Sommerville, Ian. Software Embebido. Ingeniería de Software. Naucalpan de Juárez, Estado de México : Pearson Educación, 2011, págs. 540 - 554.
16. Martin, Edward C. The University of Arizona College of Agriculture and Life Sciences. [En línea] Enero de 2017. [Citado el: 26 de Junio de 2017.] https://extension.arizona.edu/sites/extension.arizona.edu/files/pubs/az1220s-2017_0.pdf.
17. D-Robotics UK. www.droboticonline.com. [En línea] 30 de Julio de 2010. [Citado el: 26 de Junio de 2017.] <http://www.micropik.com/PDF/dht11.pdf>.
18. Proyectos Agiles. [Proyectos Agiles.org](http://ProyectosAgiles.org). [En línea] Creative Commons, 2017. [Citado el: 19 de Abril de 2017.] <https://proyectosagiles.org/desarrollo-iterativo-incremental/>.
19. Marqués, Asier. Asiermarques.com. [En línea] 11 de Abril de 2013. [Citado el: 27 de Junio de 2017.] <http://asiermarques.com/2013/conceptos-sobre-apis-rest/>.
20. Comeval Valve System, S.L. y CIA. www.comeval.es. [En línea] Agosto de 2012. [Citado el: 27 de Junio de 2017.]
21. Pizarro Aguilar, Roxana. BINASSS. [En línea] [Citado el: 29 de Junio de 2017.] <http://www.binasss.sa.cr/bibliotecas/bhp/cupula/v8n17/art3.pdf>.
22. Romeu, Alberto. Albertoromeu.com. [En línea] 9 de Octubre de 2014. [Citado el: 30 de Junio de 2017.] <http://albertoromeu.com/scrump-planning-poker/>.

Desarrollo de un Prototipo detector de caídas utilizando la placa Intel Galileo Generación I y el sensor MPU6050

Esteban Carnuccio¹, Waldo Valiente¹, Mariano Volker¹, Graciela De Luca¹,
Gerardo Garcia¹, Daniel Giulianelli¹, Sebastián Barillaro¹

¹Universidad Nacional de La Matanza,
Departamento de Ingeniería e Investigaciones Tecnológicas,
Florencio Varela 1903 - San Justo, Argentina

{ecarnuccio, wvaliente, mvolker, gdeluca, ggarcia,
dgiunlian, sbarillaro }@unlam.edu.ar

Resumen: El objetivo de este estudio es exponer los avances conseguidos durante el desarrollo de un sistema prototipo que permite detectar la ocurrencia de una caída en una persona. El cual formará parte de un sistema integral orientado a Internet de las Cosas para controlar el estado de salud y vida diaria de las personas mayores de edad utilizando la placa Intel Galileo Generación I. Se describen los criterios tenidos en cuenta para elaborar el algoritmo de caídas, mencionando las fases por la que pasa el cuerpo en dicha situación. Además se describen qué tipos de actividades de la vida diaria han sido tenidos en cuenta para descartar falsos positivos, a través del análisis de mediciones, cómo se compone estructuralmente el prototipo y las dificultades encontradas durante su construcción, tanto de hardware como de software.

Palabras Clave: Acelerómetro, Giroscopio, detección de caída, Sensor MPU6050, Fases de caída.

1 Introducción

Las caídas no asistidas en forma oportuna en una persona anciana pueden provocar deterioro funcional, requiriendo de cuidadores o personas de apoyo. La prevención de caídas es tan importante que el ministerio de Desarrollo Social de la Nación a través la Dirección Nacional de Políticas para Adultos Mayores confeccionó una guía [1], donde el tema de caídas es parte principal de las recomendaciones antes situaciones de riesgo. Señalando que los efectos de las caídas pueden ser irreversibles sin una intervención adecuada y precoz. También la Organización Mundial de la Salud [2] posee un informe indicando que cada año un tercio de la población de mayores de 64 años sufre algún tipo de caída, siendo éstas unas de las más importantes causas de lesiones en adultos mayores y una de las principales causas de muerte por accidente en los mayores de 75 años. Una oportuna atención por parte de

servicios médicos o familiares, luego de una caída, puede brindar una recuperación y contención más rápida, mejorando la confianza, o hasta incluso salvándole la vida.

Desde la década de 1980 una parte de la investigación tecnológica se ha enfocado en crear una interfaz de humano a humano. Lo que dio lugar a la creación de la disciplina computación ubicua, cuyo objetivo es integrar la tecnología en el contexto de la vida cotidiana. Mediante Internet es posible que comunicar a los dispositivos individuales con cualquier otro dispositivo en el mundo y de esta forma diseñar sistemas de detección y aviso de caídas, que pueden ser clasificados en dos tipologías: los sensibles al contexto y los portátiles.

Los sistemas sensibles al contexto requieren de la instalación de sensores diseminados por la vivienda. Estas soluciones suele ser las más precisas ya que identifican el lugar de la caída y no necesita la persona cargar todo el día con un equipo (no intrusivo), pero es limitado por el alto costo de instalación y su rango de acción solo aplica en ámbitos cerrados. Los sistemas portátiles se basan en la incorporación de sensores en un aparato portable o vestible que supervisa las actividades de los adultos mayores en tiempo real, detectando caídas sobre la base de los cambios en algunos parámetros de movimiento. Los sensores utilizados para reconocer movimientos se componen de un acelerómetro y un giroscopio. Procesando la información sobre la orientación, la velocidad angular y la aceleración permitiendo describir el movimiento de una persona. Este es un enfoque es de bajo costo y fácil de emplear, no requiere de instalación, funciona tanto dentro como fuera de la casa. En contra partida al ser portable, puede ser olvidado o no activado por la persona mayor. Si es integrado a un celular puede tener reticencia a utilizarlo. Además un golpe involuntario en el sensor, o movimientos bruscos en determinadas condiciones pueden provocar una falsa alarma. Estas pueden ser reducidas mediante el uso combinado de sensores y consiguiendo con esto además una detección más precisa. Este tipo de dispositivo, permite además alcanzar el beneficio a una población de bajos recursos, ya que con un costo reducido puede comunicar al adulto mayor con un familiar o una persona cercana para conseguir pronta asistencia.

2 Desarrollo

2.1 Análisis de una caída.

La aceleración y orientación del cuerpo de una persona que se encuentra de pie y se cae, pasa por distintos estados desde el momento en que pierde el equilibrio hasta que se encuentra en el piso [3] [4]. En la Figura 1 se pueden observar las cuatro etapas por las que pasa una persona cuando experimenta una caída. Inicialmente una persona se encuentra en Estado de reposo inicial (A), dicha posición se caracteriza porque la persona se encuentra en un ángulo inicial de la caída. Luego cuando la persona pierde el equilibrio, entra en una caída libre (B), hasta impactar con el suelo (C). Una vez que se produjo el impacto la persona se encuentra un tiempo determinado en reposo, hasta que se pone de pie nuevamente. Durante ese intervalo de tiempo la persona adquiere un determinado ángulo final (D), con respecto al

ángulo inicial. En la última etapa del algoritmo, se buscó un cálculo sencillo que convalide la caída real, desechando los falsos positivos, que puedan surgir.

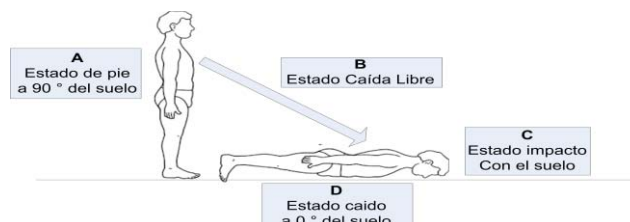


Fig. 1 Etapas de la Caída

Para poder calcular el ángulo que produce el movimiento del acelerómetro y el giroscopio dentro de un intervalo de tiempo, el método más empleado es el cálculo de los ángulos Pitch¹, Roll² y Yaw³, muy utilizado en aeronáutica. Oporto Díaz [3] calcula la posición de la persona al caerse utilizando fórmulas de seno y coseno para determinar la inclinación del individuo respecto a la vertical de los ejes X, Y y Z. Para ello en todo momento se ubica al sensor a la altura del pecho del individuo. En cambio Ramiro Blanco en “*El sistema de detección de caídas en un Geriátrico*” [4] determina si el sujeto ha sufrido un cambio de orientación, teniendo en cuenta qué eje se encuentra paralelo a la gravedad en el momento en que la persona finaliza su caída.

En esta investigación para determinar la inclinación que sufre la persona durante la caída, se comparan el ángulo inicial en que se encuentra la persona al ingresar en la fase de Reposo Inicial (A), con el que adquiere durante la fase de Reposo Final (D). El cálculo de los diferentes ángulos se realiza aplicando producto escalar de los vectores de aceleración, correspondientes al estado (A) y con el vector de aceleración que produce el movimiento en un instante determinado. Para ello se utiliza la fórmula (1).

$$\cos \alpha = \frac{(Acelx_{t0})^2 \cdot (Acelx_{tn})^2 + (Acely_{t0})^2 \cdot (Acely_{tn})^2 + (Acelz_{t0})^2 \cdot (Acelz_{tn})^2}{\sqrt{(Acelx_{t0})^2 + (Acely_{t0})^2 + (Acelz_{t0})^2} \cdot \sqrt{(Acelx_{tn})^2 + (Acely_{tn})^2 + (Acelz_{tn})^2}} \quad (1)$$

El resultado del ángulo calculado con la fórmula (1), es posteriormente verificado con el valor obtenido del Pitch y Roll. El Yaw no se está utilizando debido a que para saber si una persona se cayó, no es necesario conocer el ángulo que obtiene la persona entorno al eje Z.

¹ Angulo de rotación en eje Y

² Angulo de rotación en eje X

³ Angulo de rotación en eje Z

2.2 Prototipo

El prototipo realizado en el proyecto consta de una placa Intel Galileo Gen 1, con sistema operativo Linux Yoctos X.X.

Para detectar caídas se utiliza un acelerómetro con giroscopio MPU6050 y para lograr la autonomía se utiliza una batería especial de 8v Lithium, dado que la placa Galileo necesita sólo 5 volts de alimentación, se construye un circuito reductor con transistor LM2576. Además se le adicionó un botón pulsador como reset para poder reiniciar el sistema operativo en casos de necesidad. También consta de una conexión por wifi para organizar las mediciones en tiempo real. El diagrama de conexión se muestra en la **Figura 2**.

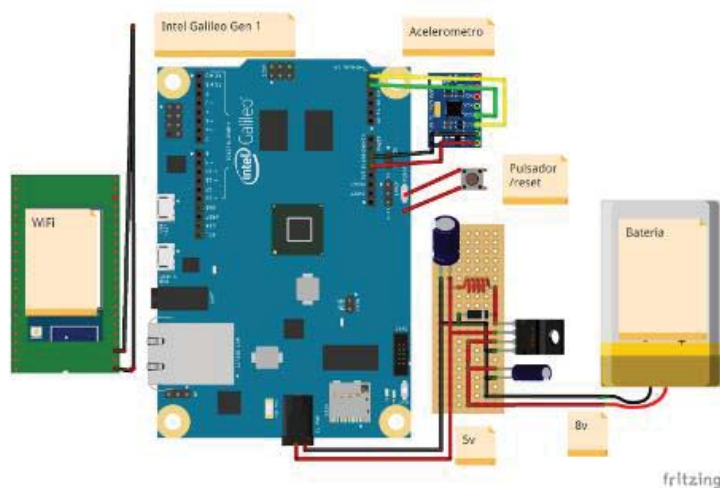


Fig. 2 Diagrama de conexión del prototipo

Uno de los problemas que surgió al utilizar el prototipo, se debió a la ubicación del acelerómetro dentro del gabinete del prototipo, provocó errores por descalibración del sensor MPU6050. Debido a que es sensible a los cambios bruscos de temperatura (rango de trabajo, -40°C a 85°C) [5]. Su hoja de datos especifica que el desvío de las mediciones del acelerómetro es de $\pm 0.02\%$ por grado centígrado y el giroscopio de $\pm 20\%$ por grados/segundos. El que afecta indirectamente a la calibración del sensor. Por ese motivo, para que el MPU6050 no se vea afectado a la temperatura que genera el microprocesador, se aisló el sensor y se lo colocó en la parte exterior del prototipo. Debido a que el microprocesador en un uso intensivo de la CPU llega a alcanzar una temperatura de 70°C . De esta forma gracias a la aislación, se logró que se afecten mínimamente las mediciones, realizadas en distintas temperaturas ambiente una vez calibrado el sensor. Además como el rango de sensibilidad de los sensores fue aumentado al máximo soportado por estos, 16G y $2000^{\circ}/\text{seg.}$, se concluyó que las mediciones no serían afectadas por la temperatura.

2.3 Ubicación del prototipo.

En la primera etapa las mediciones fueron realizadas [6] sin tener en cuenta un eje de referencia, con lo cual resultó dificultoso poder determinar correctamente el ángulo inicial y final en que se encontraba la persona en determinado momento. Esto se debió a que el prototipo no se encontraba en una posición fija durante las mediciones. Luego de diferentes pruebas se resolvió ubicarlo sujeto en la cintura del individuo **Figura 3**, resolviendo este inconveniente ya que de esta manera el eje Y sería siempre el afectado por la gravedad en lugar del eje Z. Luego de esto se realizó el muestreo de mediciones de las caídas de una persona, para poder obtener así los valores correctos de aceleración y orientación.



Fig. 3 Ubicación del prototipo en una persona

2.4 Mediciones realizadas

Una persona puede realizar distintas actividades diarias, cada una de ellas involucra distintos movimientos. A lo largo de dicha acción la persona va adquiriendo diferentes valores de aceleración y de orientación, que deben ser tenidos en cuenta al momento de analizar si se produjo o no una caída. Permitiendo de esta manera evitar falsos positivos durante la ejecución del algoritmo de detección. Por eso dos individuos diferentes contextura física han realizado distintas actividades, entre ellas caídas, y se han capturado los valores devueltos por el sensor MPU6050, con la finalidad de poder determinar correctamente el funcionamiento del algoritmo de detección. Los sujetos de prueba son de sexo masculino, de aproximadamente 35 años y de diferentes alturas. El primero mide 1.80 metros de altura y pesa 80 kg, mientras que el segundo mide 1.70 metros y pesa 72 kg. Las actividades realizadas por los sujetos de pruebas fueron las siguientes:

- **Actividades de la vida diaria:** Caminar, saltar, sentarse y levantarse de una silla, subir y bajar por una escalera
- **Actividades de caídas:** Caerse estando de pie hacia adelante, atrás o de costado. También se realizaron mediciones cayéndose hacia adelante o de costado estando sentado .

A través de las mediciones de las actividades realizadas, se pueden determinar los patrones que siguen las personas al realizar dichas tareas y observar cómo se comporta el cuerpo humano cuando sufre una caída. Por ese motivo a continuación se muestran los tres gráficos más significativos, obtenidos con los valores de las mediciones realizadas. Por un lado se muestran gráficamente los valores resultantes del cálculo de la norma de la aceleración [6]. Mientras que por otro, se muestra gráficamente el ángulo que tiene el cuerpo del sujeto de prueba en determinado momento.

a) Al caminar.

En la **figura 4** se muestra la norma de la aceleración que obtiene una persona al caminar. Como se puede observar la aceleración nunca supera los 2.5G y no llega a disminuir a los 0G. Este detalle es importante ya que nos sirve para poder determinar el alcance del algoritmo de detección.



Fig. 4 Norma de la aceleración al caminar.

b) Una persona de pie , que cae hacia adelante.

En la **figura 5** se muestran dos gráficos, el primero expone los valores de la norma de la aceleración cuando una persona se encuentra de pie y se cae hacia adelante. Mientras que el segundo gráfico muestra las variaciones de ángulos que va adquiriendo el cuerpo de la persona a medida que transcurre la caída.

Cómo se puede observar en **A** la persona se encuentra en el estado de Reposo Inicial. En esta fase, se encuentra quieta, sin embargo su aceleración es de 1G debido a que se ve afectado su eje Y por la fuerza de gravedad. Además se puede visualizar en **A**, que el ángulo inicial es de 0° . Luego, una vez que la persona pierde el equilibrio, empieza en **B** la fase de Caída Libre. En la cual se observa una disminución brusca de la aceleración, desde 1G hasta casi llegar a 0G. A continuación, apenas la persona toca el suelo se produce la transición al estado de impacto **C**, anteriormente mencionada. En esta fase se puede observar un pico abrupto en la aceleración de casi 7G, superando de esta forma el límite de los 2.5G que se puede llegar a alcanzar cuando se camina. Finalmente la aceleración decrece,

normalizándose hasta 1G. Produciéndose así el cambio de estado a la etapa Reposo Final **D**. Esta última etapa se diferencia de la fase **A**, debido a la existencia de un cambio de orientación en el cuerpo de la persona, de 0° a 90° . Indicando así que efectivamente la persona se ha caído al suelo.

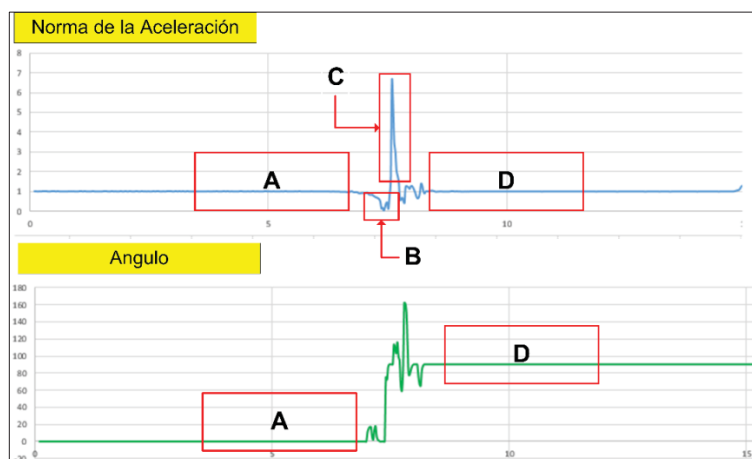


Fig. 5 Norma de la aceleración y ángulo de la persona al caerse hacia adelante.

2.5 Solución de software

Al comienzo se desarrolló el software embebido en el prototipo utilizando el lenguaje de programación Wiring de Arduino. Sin embargo por cuestiones de desempeño al utilizar el prototipo con la placa Intel Galileo Gen I, se decidió cambiar este lenguaje por Node.js, ya que esta plataforma emula determinados componentes de Arduino. No obstante en determinadas situaciones se utilizan sketches⁴ de Arduino para realizar tareas muy específicas. Desde Node.js se está utilizando la biblioteca [7] de UPM para leer el sensor MPU6050 en la placa Intel Galileo.

Para el desarrollo del detector de caídas utilizando Node.js, se está utilizando programación basada en eventos, debido a que este lenguaje se apoya en este paradigma. En consecuencia, el algoritmo de detección de caídas se basa en los eventos mencionado anteriormente. Por ese motivo se muestra en la **figura 6** la transición de estados y eventos que deben cumplirse en nuestro algoritmo para que sea detectada la caída de una persona. Se puede observar además, que si se da una secuencia completa, desde el estado reposo inicial hasta el estado reposo final, significa que se ha identificado una caída de la persona observada.

En las mediciones obtenidas primeramente, se había configurado al acelerómetro del MPU6050 con un grado de alcance de 2G y al giroscopio con $250^\circ/\text{seg}$, con el objeto de poder tener mayor sensibilidad al tomar las mediciones. Sin embargo,

⁴ **Sketch** - Parte de código fuente listo para abrir y ser cargado sobre el dispositivo, con el entorno de desarrollo integrado de Arduino.

durante el desarrollo del prototipo se observó que dicho rango no era suficiente, debido a que a una caída normal no superaba los 2G de aceleración y una persona caminando normalmente alcanzaba fácilmente ese valor. Por dicho motivo fue necesario cambiar el rango soportado por el acelerómetro a 16G y a 2000°/seg el giroscopio. De esta forma se intenta mejorar el rango de detección del MPU6050, descartando a su vez aceleraciones mínimas que se obtienen al mover levemente el prototipo.

El MPU6050, como la mayoría de los sensores, presenta desviaciones en los valores que vaya sensando a lo largo del tiempo. Pueden ser diversas las causas de estas desviaciones, también conocidas como offsets, ya que su origen puede deberse a las soldaduras que presenta el integrado desde su momento de fabricación, o a cambios de temperatura y humedad, entre otras. Por lo tanto resulta indispensable su calibración antes de la comenzar a utilizarlo, siendo necesario realizar este paso, para encontrar los offsets correctos para calibrar el sensor.

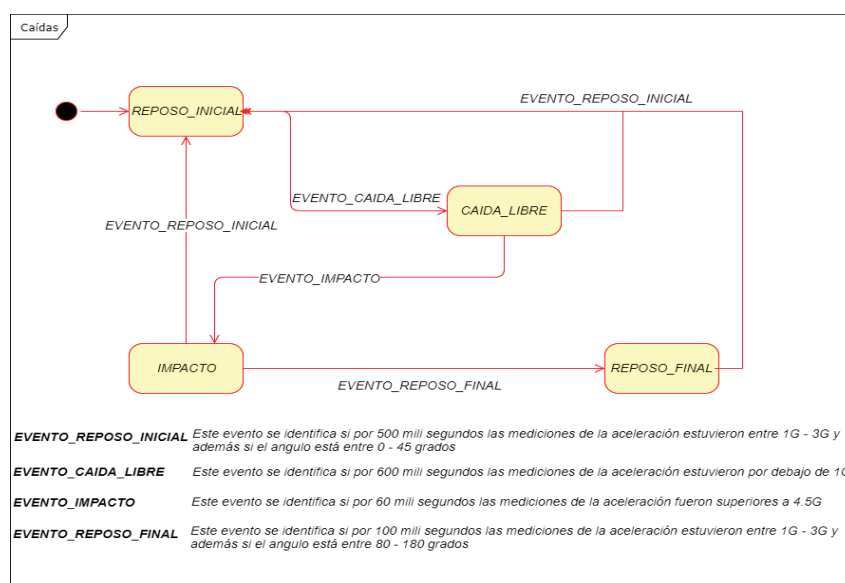


Fig. 6 Secuencia de transición de estados para detectar una caída

En primera instancia de la calibración del MPU6050 se había conseguido realizar utilizando el sketch calibrador desarrollado por Naylampmechatronics [8]. Sin embargo, durante el desarrollo del módulo detector de caídas, se descubrió que este programa era poco eficiente, debido a la demora excesiva en su ejecución. Esta situación es poco performante y de poca usabilidad si el prototipo es utilizado por un usuario adulto. Por ese motivo actualmente se reemplazó este mecanismo de calibración utilizando el sketch desarrollado por Luis Ródenas [9]. El cual permite obtener los offsets correctos para cada eje y almacenarlos en los registros offsets del procesador DMP del MPU6050 [5]. Para poder ejecutar este código, al igual que cualquier otro calibrador de acelerómetro y giroscopio, es necesario que la primera

vez que se utiliza el prototipo se lo coloque sobre un plano horizontal durante su ejecución, apuntando el eje Z positivo hacia el techo. Después de que sean almacenados los offsets en el DMP, cada vez que se ejecute el programa de Node.js, se utilizará indirectamente los offsets que se encuentran en el DMP restándolos automáticamente a los valores leídos del acelerómetro y del giroscopio. De esta manera se están obteniendo las mediciones sin desviación. En el programa de Node.js se debió emplear la biblioteca [7] para poder utilizar el MPU6050 desde ese lenguaje de programación.

Es importante mencionar, que el sketch de Luis Ródenas está desarrollado para funcionar con una escala de 2G. Por lo tanto al modificar la sensibilidad del acelerómetro a 16G, fue necesario modificar este sketch para que el valor de los offsets sea calculado de acuerdo a la escala con la que se está trabajando.

2.6 Desempeño del funcionamiento de Node.js en un sistema embebido.

Uno de los problemas encontrados durante el análisis de las primeras mediciones realizadas [6], fue el intervalo de tiempo en el que se habían capturado los valores sensados por acelerómetro y giroscopio, debido a que se realizaba repetidamente una lectura del sensor cada 50 milisegundos. Por ese motivo la cantidad de datos capturados, en el intervalo de tiempo, no eran suficientes para poder determinar si se había efectuado una caída. Con lo cual fue necesario bajar el tiempo de muestreo a 20 milisegundos. Siendo necesario realizar adaptaciones en el código fuente para subsanar las limitaciones que presenta Node.js. Una de las modificaciones necesarias en el programa fue paralelizar la lectura del sensor MPU6050 que se realizan desde Node.js. Dado que este lenguaje se ejecuta en un único hilo de ejecución, se debió utilizar la biblioteca *child_process* para poder paralelizar su ejecución utilizando procesos [9]. No obstante, se pudo observar que cualquier programa realizado en Node.js que lee valores de sensores y se ejecute en la Placa Intel Galileo Gen I, es poco performante en cuanto a rendimiento. Debido a que consume considerables recursos de hardware, como CPU y memoria. Pero se consigue disminuir mínimamente su consumo, si se lo paraleliza con procesos hijos. Esto es debido a que Node.js no fue desarrollado para ser utilizado en sistemas embebidos, sino principalmente para el desarrollo principal de servidores web. Sin embargo debido a que el prototipo del detector de caídas se está desarrollando en una placa Intel Galileo, que posee suficiente recursos de hardware para poder ejecutar el programa realizado en Node.js sin percance, se decidió que era conveniente seguir desarrollando el módulo detector de caídas utilizando este lenguaje de programación.

3 Conclusiones

De acuerdo a lo expuesto se ha podido determinar los patrones de movimientos que tiene una persona al realizar sus actividades diarias y los diferentes tipos de caídas que puede sufrir ante determinados sucesos. A partir del análisis realizado se pudieron detectar los eventos que permite diferenciar las caídas de una actividad rutinaria,

diseñando un algoritmo de detección de caídas con las restricciones antes mencionadas, que puede descartar falsos positivos teniendo en cuenta las fases de la caída.

Se continuará en la investigación con la implementación del algoritmo de detección, realizando pruebas unitarias con el prototipo construido, teniendo en cuenta las distintas acciones que puede realizar el adulto en su vida diaria, y ante la detección de una caída notificarle automáticamente al familiar dicha ocurrencia a través de dispositivos móviles. Además se pretende agregar mayor precisión incorporando otros tipos de sensores al prototipo en el futuro, como pulso cardíaco que se incorporará a nuestro algoritmo de caída. En fases siguientes de la investigación, a partir del prototipo, se proyecta elaborar un producto comercial útil para la sociedad, utilizando componentes de hardware más específicos, en vez de placas de desarrollo que son empleadas para prototipado. Se espera que éste proyecto realice un aporte considerable a una necesidad concreta de nuestra sociedad, la supervisión de personas ancianas, utilizando metodologías IoT y poniéndolas a disposición de los individuos que no dispongan de recursos o tiempo para ésta actividad, teniendo la premisa de que la solución utilice herramientas y/o dispositivos masivos para reducir costos en la solución final, permitiendo a personas de cualquier nivel socioeconómico la posibilidad de acceder a estos dispositivos.

4 Referencias

- [1] M.L.Roqué, Guía para el autocuidado de la salud de las personas mayores, Buenos Aires: Secretaría Nacion de la Niñez, 2010.
- [2] Casiliari y Oviedo-Jiménez, Automatic Fall Detection System Based on the Combined Use of Smartphone and a Smartwatch, Malaga, España: Universidad de Málaga, 2015.
- [3] E. Oporto Díaz, «Diseño de un sistema inalámbrico de detección de caídas aplicado a personas de la tercera edad basado en acelerómetro y teléfono móvil,» Lima-Peru, 214.
- [4] R. Blanco, «Sistema de detección de caída en personas de la tercera edad para uso en centro geriaticos,» Bogota, 2010.
- [5] I. InvenSense, *MPU-6000 and MPU 6050 Product Specification Revision 3.4.*, Sunnyvale: Inc.InvenSense, 2013.
- [6] S. Barillaro, W. Valiente, G. De Luca, E. Carnuccio, G. García, M. Volker, D. Guilianelli y N. Casas, «Diseño de sistema IoT de monitoreo para personas mayores,» de *XIX Workshop de Investigadores en Ciencias de la Computación(WICC 2017)*, Buenos Aires, 2017.
- [7] Intel, «http://iotdk.intel.com/docs/master/upm/group__mpu9150.html,» [En línea].
- [8] S. Barillaro, W. Valiente, G. De Luca, E. Carnuccio, G. Gracia, M. Volker y D. Guilianelli, «IoT fall detection system for the elderly using,» de *IEEE CACIDI 2016 - IEEE Conference on Computer Sciences*, IEEE, 2016.
- [9] L. Ródenas, «<https://www.i2cdevlib.com/forums/topic/96-arduino-sketch-to-auto>

Ejecución de comandos de voz mediante Web Speech API

Javier Pérez¹, Javier Diaz², Ivana Harari²

¹ Facultad de Informática – Universidad Nacional de La Plata (UNLP)
javi.pzv@gmail.com

² LINTI – Facultad de Informática – Universidad Nacional de La Plata (UNLP)
jdiaz@unlp.edu.ar, iharari@info.unlp.edu.ar

Abstract. The predominance of exclusively manual interaction required when interacting with websites is a restrictive condition. This limitation is taken as a starting point and shapes the main motivation for the development of web applications driven by voice commands. It is described and provided a process useful to allow the user to interact with web applications invoking voice commands.

Keywords: Speech recognition, voice commands, voice interfaces, web applications.

1 Introducción

Una interfaz de comando de voz propone un medio de interacción oral que permite a las personas comunicarse con los dispositivos de forma natural utilizando exclusivamente la voz.

Al contrario de los otros mecanismos de interacción, donde las personas deben adaptarse al uso de dispositivos artificiales, en la interacción por medio de la voz se invierten los roles; es la computadora quien se ocupa de entender al ser humano.

Al hacer la analogía entre una interfaz con reconocimiento de voz y el teclado convencional de una computadora, se observa que en el primero la persona accede a un teclado virtual compuesto de miles de teclas de acceso directo, donde cada una de ellas representa una acción a ejecutar a diferencia del teclado que requiere presionar una secuencia de teclas establecidas.

Las ventajas de interactuar a través del uso de la voz producen un incremento significativo en la calidad de comunicación que se da entre el dispositivo y la persona, al mismo tiempo que reduce la cantidad de componentes visuales que deben ser exhibidos para lograr una buena experiencia de usuario.

En la actualidad los sistemas web pueden integrar servicios de reconocimiento de voz externos a la aplicación y de esta forma brindar un medio alternativo de interacción.

A continuación, se describe el sistema de reconocimiento utilizando en el desarrollo de la aplicación "Handsfree for Web" correspondiente al trabajo final de grado "Navegación web dirigida por comandos de voz" [2].

2 Servicio de reconocimiento de voz

Actualmente navegadores de uso masivo como Google Chrome, Firefox y Opera [1] implementan la denominada Web Speech API.

La API de reconocimiento de voz tiene como objetivo dar servicios de análisis y síntesis del habla [2]. Esto permite a los usuarios integrar un sistema de reconocimiento de voz en aplicaciones web. Regularmente estas aplicaciones utilizan el procesamiento de voz para transformar un discurso oral en texto o viceversa y de esta forma proveerle al usuario nuevos mecanismos de interacción.

La API de reconocimiento de voz provista por los navegadores que la soportan cuentan con las siguientes ventajas que facilitan la integración con sistemas web [3]:

- Servicio de procesamiento de voz externo a la aplicación
- Procesamiento remoto
- Procesamiento en tiempo real
- Obtención de resultados parciales durante el proceso de reconocimiento
- Uso gratuito y sin límites
- Sistema de reconocimiento independiente del usuario, de propósito general y capaz de procesar frases con palabras conectadas
- Soporte de múltiples idiomas y dialectos
- Mejora continua del servicio
- Posibilidad de definir un servicio de reconocimiento alternativo al que viene asignado por defecto



Fig. 1. Soporte de API de Reconocimiento de Voz por parte de los navegadores más populares [1].

3 Proceso de ejecución de un comando de voz

Se define un comando de voz como un par (*nombre, acción*) donde *nombre* es una frase que puede ser pronunciada por el usuario y *acción* es una tarea que será ejecutada por el sistema una vez que el nombre asociado sea invocado.

En esta sección se describen los componentes y las etapas que conforman el proceso de ejecución de comandos de voz que desempeña la herramienta Handsfree for Web, mientras los usuarios interactúan oralmente [2].

3.1 Captura y envío de señal sonora

Cuando una aplicación web solicita el inicio de los servicios de reconocimiento de voz por primera vez, el navegador solicita al usuario permisos de captura de sonidos a través del micrófono.

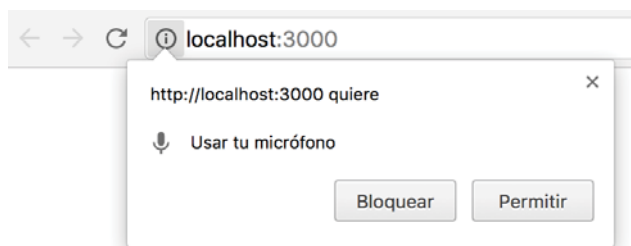


Fig. 2. Permisos solicitados por el navegador Google Chrome al iniciar el servicio de reconocimiento de voz.

La API de Reconocimiento de Voz propone dos formas distintas de captura de sonido.

Reconocimiento de voz continuo

Cuando este modo es seleccionado, el cliente de reconocimiento de voz, captura y procesa el sonido de forma continua hasta que el servicio de reconocimiento es detenido de forma explícita.

Reconocimiento de voz interrumpido

A diferencia del modo continuo, cuando se opta por la modalidad de reconocimiento de voz interrumpido, el servicio finaliza automáticamente una vez que el sistema detecta una pausa en el sonido capturado.

3.2 Recepción de transcripción

Mientras el flujo de datos correspondientes al sonido capturado es enviado activamente, se reciben transcripciones parciales y no finales correspondientes al sonido procesado.

La herramienta puede recibir, como resultado del proceso de reconocimiento del sonido, hasta diez transcripciones posibles, cada una de ellas con un grado de precisión y con una marca que indica si son resultados finales o parciales.

3.3 Inferencia y resolución de un comando de voz

Las transcripciones parciales recibidas pueden ser utilizadas para dar feedback al usuario acerca del estado del proceso de reconocimiento de voz. A la hora de determinar un comando de voz, solo debe tenerse en cuenta las transcripciones de tipo final que superen cierto nivel de precisión.

A pesar de que la Especificación de Web Speech API describe la posibilidad de definir reglas gramáticas [2], actualmente las implementaciones del servicio no soportan la incorporación de las mismas, cuando son definidas por el cliente. Esto provoca en que las transcripciones recibidas resultantes del proceso de reconocimiento corresponden a texto que no obedece ninguna regla gramatical asignada por la aplicación web.

Regularmente se obtienen múltiples transcripciones que pueden ser comandos de voz y no coinciden con exactitud con las acciones disponibles para ser ejecutadas. Por ejemplo, al indicar verbalmente el comando de voz “abrir” se puede obtener la transcripción “habría”. Al buscar por el comando “habría” se encuentra que el mismo no existe. Situaciones de este tipo dan origen en la necesidad de realizar aproximaciones con el fin de determinar si existe un comando similar al mencionado que satisfaga la intención del usuario.

Con la finalidad de realizar aproximaciones precisas, se propone en principio transformar todos los comandos disponibles y las transcripciones recibidas en fonemas utilizando el algoritmo Metaphone [4]. Posteriormente, utilizando el Coeficiente de Sorensen-Dice [5], se comparan los fonemas correspondientes a las transcripciones y los comandos disponibles con el propósito de obtener el par (transcripción, nombre de comando) que presenten mayor grado de similitud.

La obtención del par (transcripción, nombre de comando) con mayor similitud no implica que el mismo sea un buen resultado. Puede ser que el grado de similitud obtenido sea bajo y por consecuencia lejos de representar la intención del usuario.

Luego de sucesivas pruebas realizadas a través de ensayo y error utilizando la herramienta Handsfree for Web [6], se concluyó que el par (transcripción, nombre de comando) debe poseer al menos un 70% de nivel de similitud para que el comando sea considerado como representativo de las intenciones del usuario.

La etapa de resolución de comando de voz puede finalizar con dos resultados; la existencia de un comando a ejecutar o la ausencia del mismo. En caso de no existir un comando a ejecutar se puede informar por pantalla que no existe el comando de voz indicado y/o solicitar que el mismo sea pronunciado nuevamente. De lo contrario se procede a la etapa siguiente.

3.4 Ejecución de un comando de voz

Una vez elegido el comando de voz, la reproducción del mismo consiste simplemente en ejecutar el código definido para el mismo.

4 Resultados obtenidos

Durante la fase de evaluación de la aplicación Handsfree for Web se realizaron cinco entrevistas a personas pertenecientes a distintos rangos etarios, idioma y género. A las mismas se les solicitó que naveguen la web invocando comandos de voz [2].

	Maitena	Alicia	Carlos	Belén	Ivan
Idioma	Español	Español	Español	Español	Inglés
Comandos Invocados	56	92	30	38	94
Comandos interpretados correctamente	51	83	27	33	80
Comandos interpretados incorrectamente	5	9	3	5	15
Tasa de acierto	91%	90%	90%	86%	85%

Fig. 3. Resultados obtenidos durante las pruebas de usabilidad correspondientes a la aplicación Handsfree for Web

Los resultados obtenidos durante las evaluaciones realizadas indican una tasa de acierto promedio cercana al 90%. Al analizar los comandos de voz fallidos, se observó que la mayoría de ellos fueron producidos por la incapacidad de reconocer la palabra "click" cuando la aplicación estaba configurada en modo español. Una situación similar sucedió con los comandos de voz correspondientes a números. Al momento de realizar la evaluación, el software de reconocimiento brindaba la representación literal de los números, en vez de la numérica. Por ejemplo, resolvía la palabra "quince" en vez de el número "15". Luego de ajustar la aplicación para que sea capaz de procesar representaciones literales de números y agregar un alias al comando "click" llamado "presionar", se obtuvieron tasas de acierto cercanas al 100% para los idiomas español e inglés.

A continuación se brindan mecanismos adicionales mediante los cuales se puede incrementar la tasa de acierto correspondiente al reconocimiento de comandos de voz.

5 Mejoras al reconocimiento de comandos de voz

En la sección anterior se describió una forma simplificada en la cual se puede inferir un comando de voz a partir del texto resultante del análisis de la expresión oral del usuario. Se considera que este proceso puede ser mejorado desde los siguientes aspectos:

5.1 Mejorar la captura del sonido

Desde el punto de vista del hardware el medio convencional por el cual se captura la voz del usuario es el micrófono, el cual puede ser interno o externo. Estos micrófonos, al no ser de uso profesional, poseen deficiencias a la hora de capturar el sonido.

Atendiendo al software, el controlador instalado en el sistema operativo suele aplicar filtros con la intención de disminuir el ruido ambiental. Si bien el contexto en el cual el usuario utiliza la herramienta no es un ambiente controlado y estaría fuera del alcance de este trabajo, se podría proveer al usuario un dispositivo especializado de captura de audio con filtros de sonidos apropiados tales como; balance de graves y agudos, reducción de ruido ambiental, y otras normalizaciones. De esta forma la captura específica de la voz de la persona sería más eficiente y como consecuencia de ello se obtendrían mejores resultados a la hora de realizar el reconocimiento de voz.

5.2 Mejorar el reconocimiento de voz

La API de reconocimiento de voz descrita permite la definición de un servicio de reconocimiento diferente al provisto por defecto. Con el fin de mejorar los resultados del proceso de reconocimiento, se podrían utilizar otros servicios alternativos que existen en el mercado.

En la actualidad existen diversos sistemas de reconocimiento de voz, cada uno de ellos presentan diversas características, las cuales son relevantes a la hora de obtener mejores tiempos de respuesta y tasas de reconocimiento de voz [6].

5.3 Mejorar la inferencia del comando de voz

Una vez obtenida la transcripción resultante del proceso de reconocimiento de voz que dio lugar a la expresión verbal del comando por parte del usuario, es necesario determinar en concreto cual es el comando de voz a ejecutar. Por lo general, la identificación del comando no concuerda en un 100% con la transcripción recibida por el servicio de reconocimiento de voz. Por ejemplo, cuando el usuario dice "click", se suele obtener la transcripción "clip". Al intentar determinar el comando correspondiente a "clip" se encuentra que el mismo no existe.

Al no contar con sistemas de reconocimiento de voz con tasas de acierto del 100%, es importante analizar los resultados obtenidos, con el fin de inferir cuál fue la intención del usuario, y a partir de ello ejecutar el comando de voz que tenga una mayor probabilidad de acierto.

Como se vio en la sección anterior, es posible realizar comparaciones fonéticas entre los posibles comandos que pueden ser ejecutados y las transcripciones resultantes de la expresión del usuario, con el objetivo de determinar los posibles comandos de voz que él mismo quiso ejecutar. El comando de voz de mayor probabilidad de acierto es el elegido.

Si bien se pueden obtener buenos resultados a la hora de inferir los comandos, el proceso descrito puede ser mejorado. A continuación, se describen algunas posibles mejoras.

Luego de la ejecución de un comando, el usuario podría indicar si el comando ejecutado fue el deseado. De esta forma, se podrían almacenar los casos fallidos de resolución de comandos con el fin de no volver a reproducir el fallo.

Cada vez que una persona menciona un comando de voz desconocido para la aplicación, se le podría mostrar una lista de comandos de voz disponibles y consultarle qué comando de voz deseó invocar. De esta forma, la próxima vez que el usuario invoque las mismas palabras, se contará con la información necesaria para inferir el comando apropiadamente.

La correcta selección de los comandos de voz es un aspecto importante a tener en cuenta. A la hora de inferir los comandos, se obtienen mejores resultados si la pronunciación de los mismos es fonéticamente diferente. Por ejemplo, si analizamos el comando “dos”, correspondiente al número "2", es en cierta forma es similar al comando “favoritos”. En caso de visualizar errores en la resolución de comandos, se podrían renombrar alguno de ellos a “número dos” o “lista de favoritos” respectivamente.

6 Conclusión

El servicio de reconocimiento de voz provisto por los navegadores web posibilita a las aplicaciones web brindar nuevos mecanismos de interacción.

La incapacidad de especificar gramáticas a la hora de inicializar los servicios de la Web Speech API motiva la necesidad de inferir comandos de voz a partir de lo mencionado por el usuario.

Se propuso un método de reconocimiento e inferencia de comandos, el cual realiza comparaciones fonéticas entre las transcripciones textuales resultantes de lo expresado verbalmente por el usuario y las acciones soportadas por el sistema.

El proceso de ejecución de comandos de voz, permite capturar la intención del usuario de una forma efectiva y confiable. Estas características hacen posible su utilización como punto de entrada de una interfaz de voz.

Referencias

1. Speech Recognition API, <http://caniuse.com/#feat=speech-recognition>
2. Web Speech API Specification, <https://dvcs.w3.org/hg/speech-api/raw-file/tip/speechapi.html>
3. Adorf, J.: Web Speech API. KTH Royal Institute of Technology, Stockholm (2013)
4. Lawrence Philips' Metaphone Algorithm, <http://aspell.net/metaphone>
5. Rodriguez-Salazar, M. E., Álvarez-Hernández, S., Bravo-Núñez, E.: Coeficientes de Asociación. Plaza y Valdés (2001)
6. Perez, J.: Navegación web dirigida por comandos de voz. UNLP, La Plata (2017)

Estación de monitoreo en tiempo-real de parámetros agrometeorológicos para determinar la necesidad de riego en plantaciones agrícolas

Alberto Eduardo Riba¹, Emmanuel Alejandro Portugal Murcia¹, Fernanda Beatriz Carmona¹, Jorge Damián Tejada¹, Matías Agustín Pérez¹

¹Departamento de Básicas y Aplicadas, Universidad Nacional de Chilecito
9 de Julio 22, Chilecito, La Rioja, Argentina
{ariba, jtejada, fbcarmona}@undec.edu.ar
{emmanuel.portugal.91, mataguper20}@gmail.com

Abstract. La automatización y control de los sistemas de riego han evolucionado notablemente existiendo diferentes tecnologías que permiten calcular en tiempo real el contenido de agua en el suelo, parámetros climáticos y el estatus hídrico de la planta. Este trabajo implementa un prototipo que integra tecnologías Open Hardware para la captura y transmisión de los datos y un software que permite integrar los índices sensados por varias estaciones para visualizar, generar alertas y estimaciones relacionadas con el riego utilizando interfaces gráficas que contribuyan al proceso de toma de decisiones.

Keywords: Open Hardware, IoT, Data Driven Agriculture, sensores, monitoreo.

1 Introducción

La tecnología es la impulsora de lo que hoy se denomina agricultura de precisión [1]. La generalización del uso de los sistemas de posicionamiento globales y los avances en las herramientas Open Hardware para el procesamiento y sensado de datos, permiten mejorar la productividad agrícola [2]. Data-driven agriculture no solo implica la digitalización y almacenamiento de información sino también desplegar políticas de gestión de los datos enfocadas a su análisis, explotación y protección [3].

En las regiones áridas y semiáridas, como el noroeste argentino, la escasez y competencia del recurso hídrico causa problemas de abastecimiento e incrementos en los costos de extracción, especialmente en el sector agrícola, por lo que es de vital importancia la racionalización y uso eficiente del mismo para incrementar la productividad [4].

Actualmente los sectores de la región capturan la información relacionadas con las variables necesarias para el pronóstico y planificación del riego en el cultivo de olivo, de manera aislada e independiente por cada tipo de sensor, utilizando para el análisis e interpretación de datos la interface de software suministrada por el fabricante del dispositivo. No existe una solución que sea escalable e integre toda esta información. De esta manera, los encargados de tomar la decisión de cuánto y cuándo regar

necesitan utilizar herramientas adicionales para poder procesar y relacionar esta información.

En este artículo se presenta un prototipo de estación de monitoreo en tiempo-real que integra los índices de varios sensores (temperatura, humedad relativa, presión atmosférica, luminosidad, humedad de suelo a tres niveles, velocidad y dirección del viento, precipitaciones) montados sobre un microcontrolador Arduino Mega. El prototipo incluye un módulo de comunicaciones que se implementa con un shield Ethernet conectado a una antena Wifi pero también se podría implementar sobre un shield 3G – GPRS dependiendo de la disponibilidad de señal; la transmisión de datos se basa en una arquitectura cliente servidor entre una o varias estaciones y un servidor de aplicaciones, que inserta e integra los registros en una base de datos para la generación de informes estadísticos, alertas y estimaciones relacionadas con el riego, utilizando interfaces gráficas que contribuyan al proceso de toma de decisiones. En caso de falla en la conectividad la estación posee una memoria micro SD, que almacena los datos localmente permitiendo su posterior sincronización.

2 Metodología

Entre las metodologías empleadas en el presente trabajo podemos mencionar:

- Metodología de la investigación: se empleó el proceso de la investigación científica de Tamayo y Tamayo [5] como marco metodológico de referencia.
- Metodología del área aplicada: para el desarrollo del sistema se empleó la Metodología de la Red Nacional de Integración y Desarrollo De Software Libre [6]. Desarrollada en el Centro Nacional de Tecnologías de Información en Venezuela, en el año 2007.
- Metodología agronómica: Para el cálculo del riego y las necesidades hídricas se emplearán diferentes fórmulas establecidas por la FAO [7]. Estas fórmulas son aceptadas en cualquier territorio del planeta tras el resultado de muchos años de investigación en el cálculo de la evapotranspiración, valor esencial para poder determinar correctamente la necesidad de agua de un cultivo.

3 Arquitectura Hardware

El prototipo está compuesto por los siguientes dispositivos Open Hardware, un microcontrolador Arduino Mega programable [8]; un shield ethernet [9]; shield SparkFun weather [10]; un reloj de tiempo real; sensores de humedad de suelo DFRobot a tres niveles; sensores meteorológicos; un punto de acceso inalámbrico outdoor; una batería de 33 amper; un panel solar de 20 watts; un regulador de voltaje de 5 amperes.

Para capturar los parámetros meteorológicos se utiliza el shield SparkFun weather que tiene integrado los sensores de presión atmosférica (MPL3115A2), luminosidad (ALS-PT19), temperatura y humedad relativa (Si7021). Además se conectan a este shield una veleta, que nos permite calcular la dirección del viento en 36 puntos

cardinales, un anemómetro, para la velocidad del viento hasta una velocidad máxima de 200 km/h, un pluviómetro encargado de medir las precipitaciones (weather meters [11]) y un puerto para conectar una antena GPS

Se utiliza un reloj de tiempo real DS3231 RTC [12], que permite llevar un registro detallado del tiempo en las capturas y envío de los datos. La transmisión se realiza a través de los pines SCL y SDA utilizando el protocolo I2C de comunicación.

Para el sensado de la humedad de suelo, se utilizan los sensores DFRobot [13], que miden los niveles de humedad mediante detección capacitiva y es construido con material resistente a la corrosión que prolonga su vida útil.

Los sensores detallados anteriormente son conectados a un dispositivo Arduino mega 2560, que contiene un microcontrolador ATmega2560, y todo lo necesario para realizar la programación e integración de los componentes.

El microcontrolador envía los datos a un servidor a través de un shield ethernet arduino que cuenta con un conector ethernet estándar RJ45 y trae integrado lector de tarjeta Micro SD . El prototipo se abastece de energía eléctrica a través de una batería 12V de 33 amper gel, que es cargada por medio de un panel solar de 20 Watts. Para evitar la sobrecarga que dañen la batería, es necesario utilizar un regulador de 5 amper. En la Figura 1, se puede observar la estructura física de la estación de monitoreo.

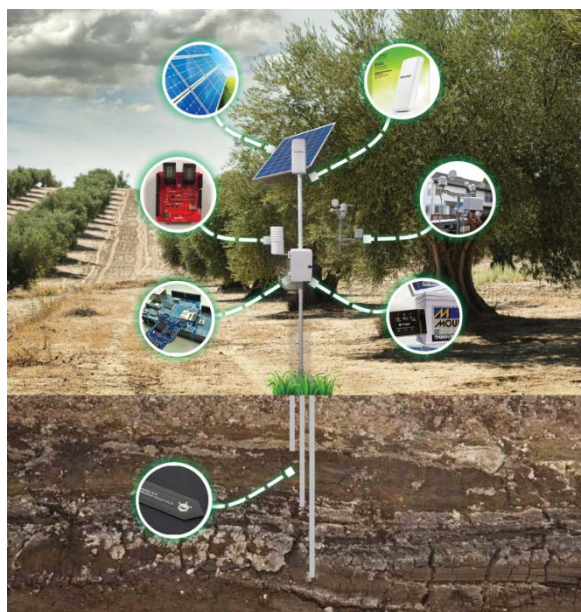


Figura 1. Estructura Física de la estación de monitoreo

En la figura 2 se muestra el diagrama de conexión de los componentes utilizados en el diseño de la estación.

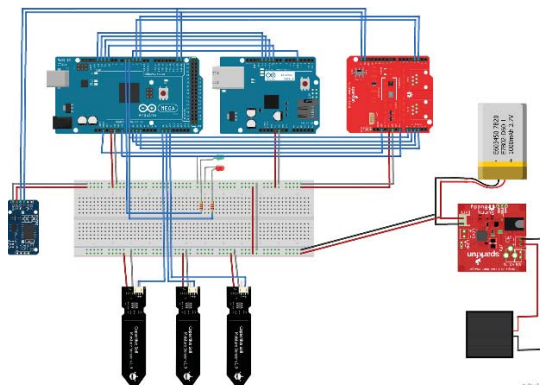


Figura 2. Esquema de conexión de la estación.

4 Arquitectura Software

El software está compuesto por cuatro módulos: de captura; de comunicación; de recepción; de aplicación. Las estaciones crean un flujo de índices agro meteorológicos constantes, que se procesan en cada uno de estos módulos de software.

4.1 Módulo de Captura

La captura de los índices sensados es el primer paso que se realiza para la obtención de datos agrometeorológicos reales, que luego se procesan para la generación de informes estadísticos, alertas y estimaciones que contribuyen al proceso de toma de decisiones para la gestión del sistema de riego.

Los sensores conectados al prototipo, permiten detectar las variables de instrumentación (temperatura, humedad relativa, humedad del suelo, luminosidad, presión atmosférica, velocidad y dirección del viento, precipitaciones, nivel de batería, etc.) y transformarlas en variables eléctricas medibles en una longitud. Estos dispositivos siempre están en contacto con las respectivas variables de instrumentación a través de comportamientos disímiles a nivel software y hardware. Esta captura se realiza en distintos medios físicos o químicos, como puede ser: el aire, el agua, el suelo, la luz, entre otros.

Los sensores reciben instrucciones desde el microcontrolador para capturar señales discretas en un momento determinado (N mediciones por día de cada variable, en base al período seleccionado) y transformarlas a unidades medibles, que se almacenan solo por un intervalo de tiempo en una variable global ocupando memoria SRAM del microcontrolador.

La estación solo realiza la captura de señales cuando está en funcionamiento y se encuentra activa. El período es manipulable y puede variar su rango de acuerdo a las necesidades agrometeorológicas.

4.2 Módulo de Comunicación

Para la comunicación, la estación cuenta con un shield Ethernet con chip W5100 Wiznet, que utiliza el protocolo de comunicación TCP-IP. La comunicación se establece utilizando las diferentes librerías de software provistas por el fabricante permitiendo configurar cada estación en modo ethernetClient. Esta placa tiene integrado un conector RJ45 al que se conecta a una antena Wifi outdoor.

El microcontrolador Arduino Mega establece una nueva conexión cliente por medio del Shield Ethernet, en caso de ser ésta exitosa, se invoca una solicitud GET HTTP al servidor. Esta solicitud está compuesta por un string de datos a enviar que especifica la dirección del servidor de aplicaciones, el directorio de ubicación, el código ejecutable, el archivo y todos los datos relevantes a transmitir [14].

En el servidor se ejecuta un bloque de código de recepción de datos que se encarga de crear un archivo (o situarse al final de este en caso que ya esté creado), formatear e insertar los datos enviados desde la estación.

Si la conexión Ethernet Cliente no fue exitosa, los datos son almacenados en una memoria micro SD ubicada en el Shield Ethernet. Cuando se restablece la conexión, se procede a enviar las cadenas string almacenadas y no enviadas, liberando el archivo en la memoria micro SD ante un envío exitoso.

4.3 Módulo de Recepción

Una vez que los índices de las diferentes estaciones han sido capturados por los sensores y comunicados a través del canal de comunicaciones pueden ser afectados por ruidos en la transmisión. Para evitar que datos erróneos sean cargados son analizados, utilizando técnicas de detección y corrección de errores, de análisis de patrones y mecanismos de razonamiento, antes de ser almacenados en la base de datos central.

Este módulo ejecuta un software en segundo plano, que se encarga de tomar los archivos provenientes de las estaciones, para luego acceder, verificar y almacenar los índices capturados en una base de datos central implementada en el gestor de base de datos relacional MySQL

4.4 Módulo de Aplicación

El módulo de aplicación está compuesto por tres submódulos: seguridad, gestión y procesamiento y generación de informes.

4.4.1 Submódulo de seguridad

Este es el encargado de determinar las opciones de acceso al módulo de aplicación sobre la base de los diferentes roles y perfiles de los usuarios. La validación de usuario garantiza el acceso a las diferentes opciones del sistema de acuerdo a cada perfil de usuario.

4.4.2 Submódulo de Gestión y Procesamiento

Este contiene un tablero de comandos, que incluye información de control y de tiempo real promedio de las estaciones con la representación gráfica de los principales indicadores que intervienen en la toma de decisiones.

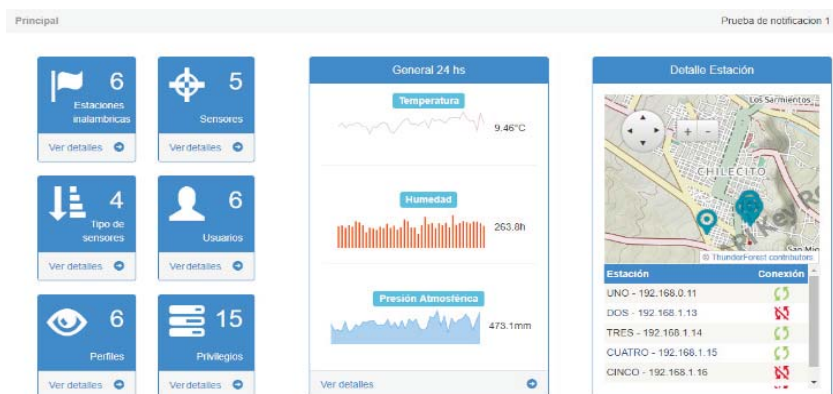


Figura 3. Tablero de comandos con información en tiempo real.

Contiene una interfaz que permite visualizar datos históricos y en tiempo real de las distintas variables sensadas, generando gráficos que facilitan la lectura de la información y la toma de decisiones. La aplicación admite la generación e impresión de informes mostrando históricos de lecturas con gráficos estadísticos, personalizando el periodo de tiempo, como así también permite incluir parámetros de configuración para personalizar las características disponibles, dotando de versatilidad a la aplicación, extendiendo su uso a más casos y escenarios (cantidad de índices sensados, sectores, estaciones, etc.).

El sistema procesa los datos disponibles combinando y aplicando funciones matemáticas cuyos resultados se traducen en la generación de alertas y estimaciones relacionadas con el riego utilizando interfaces gráficas que contribuyen al proceso de toma de decisiones.



Figura 4. Lamina inicial, Riego y Evapotranspiración.

La interfaz diseñada también permite verificar diferentes aspectos de la estación, como: su estado (en funcionamiento o no), niveles de la batería, conectividad, ubicación geográfica y toda la información pertinente.

4.4.3 Submódulo Generación de Informes

Este submódulo permite generar informes estadísticos, integrando la información registrada y procesada por el sistema de gestión de riego [15], ya implementado, con los índices sensados y procesados por este prototipo. Los índices sensados en tiempo real por la estación son combinados y concatenados con los datos registrados por sistema de gestión de riego.

La aplicación admite la generación e impresión de informes mostrando históricos de lecturas con gráficos estadísticos, personalizando el periodo de tiempo.

5 Escalabilidad y Soporte a la toma de Decisiones

Actualmente en la región, el proceso de toma de decisiones relacionados con la gestión de riego y la fertilización de los cultivos se lleva a cabo a través del procesamiento de los datos proporcionados por algunas estaciones meteorológicas y/o dataloggers, ubicadas en lugares distantes representativos de una amplia zona de cobertura con características disímiles. El costo elevado de adquisición de estos equipos, su mantenimiento y la necesidad de software adicional para el procesamiento de los índices capturados, son factores que limitan a los productores en la adquisición de la cantidad de equipamiento necesario para una adecuada planificación y gestión del riego.

La estación agrometeorológica y el software asociado presentado en este trabajo, integrado al software de gestión de riego y fertilización, suministran a un bajo costo, los índices agrometeorológicos e indicadores necesarios para la gestión de un sector de la plantación en particular.

Los dispositivos utilizados incorporan una tecnología que permite un alto grado de escalabilidad, tanto en los índices que se pueden medir y monitorizar como en el tamaño de la red. El prototipo desarrollado puede ampliarse fácilmente a más sectores de la finca, además de extender el ámbito de índices agrometeorológicos a ser monitorizados.

La implementación de varias estaciones agrometeorológicas permitirá medir la necesidad real de agua del cultivo y las características del clima en aquellos sectores de la finca en donde los índices agrometeorológicos tienen un comportamiento disímil; contribuyendo a una adecuada y certera distribución del riego y la fertilización en el cultivo de la región.

6 Trabajos relacionados

El presente trabajo incluye tanto desarrollo de hardware (estación inalámbrica y sensores) como de software (recolección, análisis y procesamiento de índices

agrometeorológicos) y se desarrolló como Trabajo Final de la carrera de Ingeniería en Sistemas de la Universidad Nacional de Chilecito, como parte del proyecto “Red de Sensores Inalámbricos basado en microcontroladores para la monitorización del riego presurizado en plantaciones de olivo”, el cual se presenta como continuidad de la línea de investigación iniciada con del proyecto “Programación y generación de pronósticos de riego presurizado para cultivo de olivo en regiones áridas, utilizando TIC”, y como complemento del proyecto “Integración y administración de índices de sensado en la programación y pronósticos de riego presurizado para cultivo de olivo en regiones áridas”. Aprobados por la Secretaría de Ciencia y Tecnología, convocatoria de financiamiento para estímulo y desarrollo de la investigación científica y tecnológica (Ficyt–UNDEC) .[16]

Como parte de estos últimos proyectos se desarrolla el Sistema de Gestión de Riego y Fertilización implementado en una empresa de agroinsumos de la región, a partir del cual y como complemento se diseña y construye el trabajo presentado en este artículo.

7 Conclusiones y Trabajos a Futuro

El objetivo fundamental es determinar el momento más adecuado para regar y la cantidad de agua a aplicar en función de, entre otros factores, el estado de humedad del suelo, de la planta y la uniformidad en el reparto de agua del sistema. La elección del nivel de automatización del riego debe realizarse siguiendo criterios técnico-económicos, según las características de la finca y las preferencias de los agricultores. El máximo nivel a alcanzar sería el control total del sistema, la humedad del suelo, el estado hídrico de la planta, del clima, y otros, incluyendo los sistemas de adquisición de datos para una adecuada gestión del riego y fertilización.

El seguimiento y control de estas variables ofrece la posibilidad de modificar las pautas prediseñadas de turnos de riego, volúmenes de agua y fertilizantes aplicados de acuerdo a la evolución meteorológica, a la disponibilidad de agua y al estado fenológico del cultivo, desde un punto centralizado que comande lo que sucede en distintos sectores de la finca.

El desarrollo de una estación con fines agrometeorológicos construida íntegramente con materiales que se consiguen fácilmente en el mercado local y de bajo costo, permite medir la necesidad real de agua del cultivo en aquellos sectores de la finca en donde las variables agrometeorológicas tienen un comportamiento disímil; contribuyendo a una adecuada y certera planificación y distribución del riego y la fertilización en los cultivos de la región.

El cálculo de la evapotranspiración de referencia (ET₀) podrá ajustarse a sectores más acotados y por consiguiente reflejar valores adecuados a las características de cada zona, permitiendo mejorar la gestión del recurso energético, factor crítico en el costo de producción en estas regiones de climas áridos.

Como trabajos futuros se proyecta complementar el software desarrollado con otra aplicación que permita la normalización de las series agrometeorológicas y posibilite métodos avanzados de detección y corrección de errores.

8 Referencias

1. M. L. Marote, "Agricultura de Precisión", Ciencia y Tecnología 10, ISEU, (2010).
2. Alberto E. Riba, J. Tejada, F. Carmona, F. Frati, M. Pérez, E. Portugal Murcia, N. Acosta, J. M. Toloza, "Agricultura de precisión y Data-driven agriculture en regiones de climas áridos", XIX Workshop de Investigadores en Ciencias de la Computación, (2017).
3. Schuster, E. W., Kumar, S., Sarma, S. E., Willers, J. L., & Milliken, G. A. Infrastructure for data-driven agriculture: identifying management zones for cotton using statistical modeling and machine learning techniques. In Emerging Technologies for a Smarter World (CEWIT), 8th International Conference & Expo on (pp. 1-6). IEEE, (2011).
4. Alberto E. Riba, F. Carmona, F. Frati, J. Tejada, N. Acosta, J. M. Toloza, "Integración y administración de índices de sensado en la programación y pronósticos de riego presurizado para cultivo de olivo en regiones áridas," XIV Workshop de Investigadores en Ciencias de la Computación, eje: Innovación en sistemas de software, (2012).
5. Tamayo, M. "El proceso de la investigación científica", Editorial Limusa, (2004).
6. Carmen R. Salazar G, Francisca Losavio, Alfredo Matteo, "MeRinde process model adaptation with Requirements Engineering techniques supported by Free Software tools", Computing Conference (CLEI), XXXIX Latin American, (2013).
7. Allen, R., L. Pereira, D. Raes y M. Smith. "Evapotranspiración del cultivo. Guías para la determinación de los requerimientos de agua de los cultivos. Estudio FAO Riego y Drenaje" 56. Roma, Italia. 298 p, (2006).
8. Arduino, Arduino Mega, <https://www.arduino.cc/en/Main/arduinoBoardMega>
9. Arduino, Arduino Ethernet Shield, <https://store.arduino.cc/arduino-ethernet-shield-2>
10. Sparkfun Start something, <https://www.sparkfun.com/products/13956>
11. Sparkfun Start something, <https://www.sparkfun.com/products/8942>
12. Maxim Integrated, <https://www.maximintegrated.com/en/products/digital/real-time-clocks/DS3231.html>
13. Dfrobot, Capacitive Soil Moisture Sensor, https://www.dfrobot.com/wiki/index.php/Capacitive_Soil_Moisture_Sensor_SKU:SEN0193
14. Viñas Gutiérrez, M. Control de acceso mediante NFC con Arduino. (2017)
15. Jorge Tejada, Sistema de Gestión de riego y fertilización. Tesis de grado, Universidad Nacional de Chilecito (2014).
16. Carmona, F. B., Riba, A. E., Frati, F. E., Tejada, J. D., Martínez del Pezzo, A., Robador, E. R. D. & Toloza, J. M. Tecnologías de la información y de las comunicaciones aplicadas a problemas de agricultura regional. En XVI Workshop de Investigadores en Ciencias de la Computación. (2014)

Generation and Use of a Digest System by Integrating OCR and Smart Searches

Germán Cáseres¹, Lisandro Delía¹, Pablo Thomas¹, Verónica Aguirre¹,

¹ Instituto de Investigación en Informática III-LIDI, Universidad Nacional de La Plata,
La Plata, Buenos Aires, Argentina
{gcaseres, ldelia, pthomas, vaguirre}@lidi.info.unlp.edu.ar

Abstract. A digest can be defined as a regulations repository which is manipulated by organizations for extended time periods. The search for information in this repositories can be tedious without assistance from an ad-hoc software application. This work presents the development of a Digest Software System with its architecture and integration with other base tools. Finally, two study cases are presented where the developed product is used.

Keywords: digest, full text search, ocr, solr, tika

1 Introduction

Organizations usually have a vast set of pre-established regulations that have to be consulted regularly. Oftentimes, applicable laws defined in those standards are not known and searches must be done that result in tedious document reviews.

Clearly, it would be convenient to have some sort of repository where the entire legislation is stored. However, searching the regulatory framework in relation to any given issue may involve some time to collect and analyze current regulations.

In this context, there is a clear need to have a digest, i.e., a collection of all regulations applicable to a specific field. This digest should also be digital in nature, since it would not have a significant physical fingerprint, but it should also be integrated to a system that provides a certain level of intelligence to search for ordinances or resolutions on a specific topic.

Thus, there are two essential processes in a software system that acts as an institutional digest. First, regulations have to be digitized, considering that many of those might be available in printed format only, and which will therefore have to be converted to a digital format. Naturally, any dispositions or regulations issued today by institutions are already produced in digital format. All this constitutes a digital repository or database of regulations.

Secondly, such repository should include a feature to perform “smart” searches that consider the entire text of the regulations, showing results sorted by relevance (word similarity, proximity, and so forth). These functionalities are available in current popular web search engines.

In this article, we introduce the development and use of a software system that is built as an institutional digest. The rest of this paper is organized as follows: in

Section 2, the problem at hand is introduced. Then, in Section 3, the tools selected to solve the problem are described. Sections 4 and 5 deal with the digest system Digesto and its use. After that, study cases and the results obtained are detailed. Finally, conclusions and future lines of work are discussed.

2 Description of the problem

The digitization process of a standard that is on paper support does not involve just obtaining a digital representation of its contents, but it also requires a mechanism that can interpret the textual information present in the digitized document. In general, this mechanism is not simple and requires applying image recognition algorithms capable of identifying language symbols.

Once this information is obtained, a mechanism that not only allows storing the text content of the regulations, but also ensures a fast response time for arbitrary content searches, has to be selected.

To achieve this, information search indexes must be generated based on the syntactic and semantic characteristics of the language used in the standards.

These mechanisms should be also integrated in a single system that can perform the necessary processes transparently to the user in charge of loading up the regulations or performing searches on them.

2.1 Digitization of Regulations

The digitization of a regulation is done using a device that is capable of capturing the content printed on the paper and saving it as a digital image. To do this task, a scanner or a digital photographic camera can be used.

Once a digital format image of the regulation is obtained, a recognition process has to be run to extract the text information that will be used as foundation for the searches. This process, called OCR (Optical Character Recognition) [1], emulates the ability of the human eye to recognize objects, identifying the symbols in an image that correspond to those of a specific alphabet.

Images enter the OCR process in a digital format (pixel matrix) and go through a number of stages that apply transformations on them to simplify the task of identifying the symbols in different contexts of colors, locations, designs, etc.

Process output information consists of the symbols extracted from the image in some character coding format (ASCII, UTF8, etc.). This output format is the most suitable to facilitate text processing and storage in any computer system.

The quality of the result depends on the quality of the digitization process and the characteristics of the algorithms used to carry out the process.

2.2 Storing Information for Content Searches

Storing the textual content of a regulation is a task that could be solved by simply storing the text as an attribute of a relational database table. However, solving this following such a simple method would involve leaving out search “smart” features, such as the ones that can be found in the most popular search engines.

To store text that is ready to allow quick searches and yields relevant results, different indexing techniques are required, which are usually not developed in the relational database engines [2] that are currently used (MySQL, PostgreSQL, SQL Server, and so forth).

3 Selecting the Right Tools

To solve the problems described above, specific knowledge is required in the field of image recognition as well as in relation to storing and indexing large volumes of information. For this reason, we decided to select tools that can solve these issues and can also be configured and integrated into a system at a specific domain, such as the Digesto system.

To perform the tasks related to text recognition on images, the TesseractOCR library [3] was used, which is an open source optical character recognition engine whose development began in 1995 and which has been optimized and improved by the Google organization since 2006 and up to the present time.

In the case of the search engine, an existing technology called Apache Solr [4] was selected; this technology makes the development, configuration and utilization of the various information indexing techniques easier, while providing a dynamic querying mechanism and ease of integration with other systems.

Solr is an open source search platform built on the Apache Lucene [5] project.

Its main features are:

- Advanced text search capabilities (through Lucene)
- Optimization for large traffic volumes
- Based on standards (XML, JSON and HTTP) [6]
- Scalability and fault-tolerance
- Flexibilization and adaptability
- Extensible architecture through plug-ins

Solr uses the Lucene library to carry out indexation and text querying tasks. At the same time, it also presents an adaptable and extensible platform that can manage the different indexing or querying requirements, without affecting system stability.

Indexed information storage is also managed by Solr through its own implementation of a non-SQL database [7], which allows easily and quickly adapting to any data structure.

The use of standards such as XML, JSON and HTTP allows simple integration with other systems, since indexing the contents of a regulation or obtaining a set of results based on a search criterion simply involves running an HTTP requirement against the Solr server.

The plug-in extensible architecture allows integrating other components to extend tool capabilities. In this sense, one of the plug-ins available in the platform belongs to the same organization that develops Solr, and is called Apache Tika [8].

Tika is a plug-in that is independent from Solr and whose purpose is obtaining information from different popular file formats:

- Images (gif, png, bmp, etc.)
- Office packages (xls, doc, ppt, etc.)
- Compression and packing (zip, tar, gz, etc.)
- Structured text (xml, html, xaml, etc.)
- Portable document format (.pdf)
- Etc.

With Tika, metadata and text can be extracted from a wide variety of different file formats. Tika runs a parsing process on the different file formats to obtain all possible information from them.

In the case of image processing, Tika is integrated with TesseractOCR and it gives Solr all the information obtained from the optical character recognition process.

4 Developing the Digesto System

The Digesto system, the same as other information systems, has a relational database whose structure is related to basic system features:

- Administration of users with different access levels
- Regulation administration:
 - Metadata
 - ◆ Different visibility levels
 - ◆ Characterization by type or issuer
 - ◆ Identification number
 - ◆ Relation to other regulations and type of relation
 - Digitized contents of the paper print-out

For the Digesto system, paper-based regulations are digitized using a scanner and stored in PDF format (Portable Document Format). Initially, the process consists in just obtaining a digital representation of the paper-based regulation, and does not involve any type of text recognition.

In particular, the digitized content is not stored directly in the relational database, but in a folder within the file system, and a unique file name is assigned to each file. This helps avoid too large databases that could hinder backup tasks.

As regards the Digesto system, the storage scheme consists of a relational database and a file system, as shown in Figure 1.

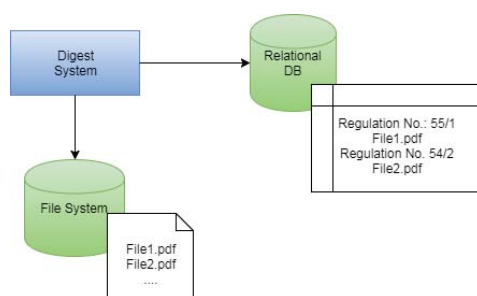


Figure 1. Initial storage scheme.

4.1 Integration of the Digesto System with Solr

The storage scheme used in the Digesto system does not consider the textual contents of the regulations, but rather stores the information in PDF format together with the attributes defined for each of them as a set of metadata.

The content extraction work is managed by Solr and handled by Tika. Using Tika allows recognizing text information within each PDF file attached to each regulation, as well as any images present in the documents. These images can also be analyzed using Tika, so that they are also processed using optical character recognition (OCR) techniques to extract their textual content if possible. OCR tasks are processed by the TesseractOCR library.

The process of generating search indexes and returning search results is handled and run by Solr.

The communication between the Digesto system and Solr is done through HTTP requirements, with information being exchanged in XML or JSON format. Figure 2 depicts the storage scheme integrated with Solr.

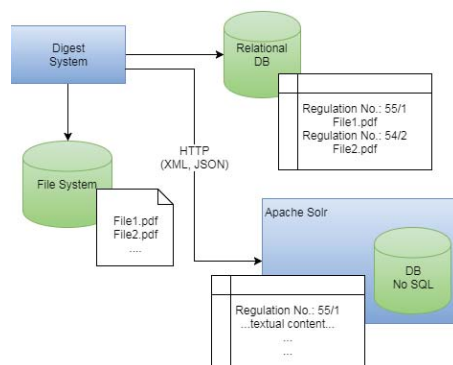


Figure 2. Storage scheme integrated with Solr.

4.2 Synchronization Between the Digesto System and Solr

Synchronizing the information included in a regulation between the Digesto system and Solr is important to guarantee that the search results returned by the system are up to date.

When a new regulation is loaded to the Digesto system, it is stored in the relational database and sent to Solr to be indexed and stored in its own, search-oriented format.

When a regulation is modified, its information has to be totally or partially re-indexed.

Finally, if a regulation is removed, the corresponding information has to be removed from the index as well to prevent the outdated information to be included in future search results.

Even though both databases must be synchronized for the proper operation of the system, the indexed information can be automatically re-built from the relational database of the Digesto system.

4.3 Adding a Regulation to the Digesto System

With the storage infrastructure mounted and configured as shown in Figure 3, the process for adding a regulation to Digesto involves the following steps:

1. Digitization of the paper-based regulation to PDF format using a scanner. (Manual process, external to the system).
2. Loading regulation metadata: date, number, issuer, attached digitized file, etc. (Manual process, in the system).
3. Storing regulation metadata in the relational database. (Automated process).
4. Storing the digitized file in server's file system. (Automated process).
5. Optical Character Recognition of the digitized file to obtain the textual content of the regulation. (Automated process in Solr through Tika and TesseractOCR).
6. Indexing recognized content to allow text searches. (Automated process in Solr).

Figure 4 shows a simplified diagram of the process for adding regulations.

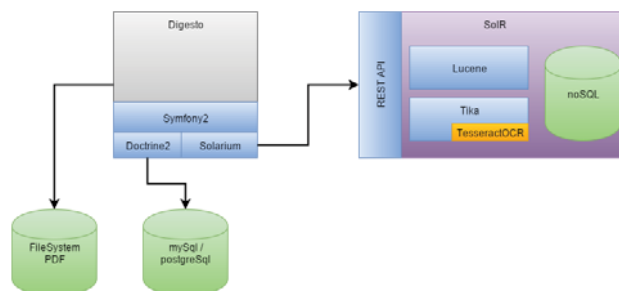


Figure 3. Detailed structure of components in the integrated infrastructure.

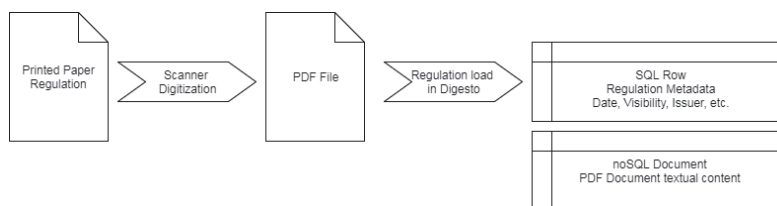


Figure 4. Process for adding regulations

5 Using the Digesto System

The integration of an ad hoc system to perform searches over a set of regulations opens up a number of possibilities that would be too expensive to develop.

The relevant aspects offered by Solr's search engine in the Digesto system are as follows:

Non-SQL storage - Flexibility. The storage method used by Solr is implemented through a non relational database (non-SQL). This type of databases allows indexing entirely heterogeneous documents as regards their structure.

In the case of the Digesto system, each regulation has a number of attributes that are specific to the domain and have to be indexed in the search engine together with the textual information of the regulation.

Some of these attributes should not be treated as text, since they represent dates or numeric values and must allow carrying out comparative searches using mathematical operators such as greater than, less than, not equal to, and so forth. Solr can handle

this situation through the specification of a scheme that defines the type of data that each attribute will contain.

The search engine easily adapts to the specific data structure of the regulations and is tolerant to changes that occur with time due to its non-SQL nature.

Non-SQL storage - Scalability. The non-relational database model presents advantages when it comes to escalating in terms of performance through the use of computer clusters for distributed query processing [9]. Solr can work with clusters, so it can provide mechanisms to improve response speed with large volumes of information.

Query Language. Through Lucene, Solr provides a powerful query language that is easy to understand for the end user. It has many similarities with the searches that can be done with the most popular engines nowadays.

Through Solr, Digesto allows searching for regulations that meet complex conditions, while for the person using Digesto, finding regulations is as easy as finding a page on the Internet.

Plug-In Expandible Architecture. Solr provides an expandible architecture that allows modifying or adding new information indexing methods.

The Digesto system uses this advantage to improve national identification number indexation so that searches can be done either using thousand separators or omitting them.

This is a specific requirement from one of the institutions that is currently using the system, and it was solved through the development of a specific plug-in.

Deferred Indexing. Since the Digesto system and the search engine are separate, when users add regulations to the system, they do not have to wait until the OCR+indexation process ends; they can keep working while the necessary tasks to ready the regulation for inclusion on search results are run in the background.

6 Study cases

The development of the Digesto system was originally aimed at addressing the requirements of the School of Computer Science of the National University of La Plata (UNLP). It went into production during the second half of 2015.

The second application case for the system was at the School of Economic Sciences of the UNLP, where specific changes were implemented as required by the institution and an automated import and indexation feature to add digitized regulations was implemented. The system went live in the first half of 2016.

Both instances of the system, Digesto Info (School of Computer Science) and Digesto Econo (School of Economic Sciences), are currently running on stable versions and available for everyday use at both institutions. Figure 5 shows a

screenshot of the main screen for searching regulations, which is common to all instances.

Búsqueda

Tipos de norma: Disposición Ordenanza Resolución

Estado de norma:

Fecha Desde:

Emisores:

Ejemplos de búsqueda libre:

- Juan Perez - Busca las normas que contengan la palabra "Juan", "Perez" o ambas.
- "Universidad de La Plata" - Busca las normas que contengan la frase completa.
- "Juan Perez"-Ricardo - Busca las normas que contengan "Juan Perez" pero que no contengan "Ricardo".

Búsqueda libre:

Normas

Número	Extracto	Fecha de norma	Tipo	Emisor	Estado	Visibilidad	Acciones
216-2011	Norma importada automáticamente.	01/01/2011	Resolución	Decano	Aprobado	Pública	
217-2011	Norma importada automáticamente.	01/01/2011	Resolución	Decano	Aprobado	Pública	
201-2013	Norma importada automáticamente.	01/01/2013	Resolución	Decano	Aprobado	Pública	

Figure 5. Screenshot of the regulations search screen, Digesto Econo.

7 Results obtained

Both the School of Computer Science and the School of Computer Sciences are currently using the system. The Digesto Econo instance currently holds more than 7200 loaded regulations that can be queried.

Response times for searches on regulation contents are comparable to those of any search process done at popular Internet portals such as Google, Yahoo! or Bing, and they meet the expectations of the users working with the system.

Finding a specific regulation or a set of regulations in relation to specific content is as simple and familiar as searching for a website in a web browser. More advanced queries respond to the same syntax used by Google, and results are presented sorted by relevance.

8 Conclusions

A study was carried out aimed at finding a feasible alternative for the development of a system with requirements that are strongly related to image recognition and agile information search.

The tools selected during this study were analyzed and the Digesto system was developed, which streamlines tasks related to loading and parsing historic and current regulations.

The use of a specific tool such as Solr to carry out the searches provides an added value to the Digesto system that would have been very difficult to achieve if a new tool had been developed from scratch.

Even though the configuration, synchronization and adaptation of Solr is not a trivial process, and specific knowledge of the subject area on which it is going to be used is required, the time invested doing this is negligible compared to the benefits obtained.

The Digesto system that was developed, as well as each of its instances (Info and Econo) meet the requirements and expectations of the end users, since they have a familiar and user-friendly tool that not only makes regulation querying easy, it also simplifies regulation digitization.

9 Future Lines of Work

Since the generic nature of Solr can be applied to different subject areas, in the future we are considering analyzing, configuring and optimizing the tool to improve both metadata extraction and the relevance of the search results returned by Digesto.

Additionally, the regulation digitization process will be improved through the development of a mechanism that will make digital information to remain available after the digitization of paper-based regulations to allow the automated extraction and load of metadata with no user interaction. This would allow users to only carry out the digitization process during the initial load stage; indexation in the Digesto system would be automated.

References

1. Shalin A. Chopra, Amit A. Ghadge: Optical Character Recognition. IJARCCCE, Vol. 3, January 2014.
2. Pablo Thomas, Rodolfo Bertone: Introducción a las bases de datos. Prentice Hall - Pearson Education, 2011, ISBN: 9789876153515
3. Tesseract OCR, <https://github.com/tesseract-ocr/tesseract> [Accessed 21/07]
4. Apache Solr, <http://lucene.apache.org/solr/> [Accessed 21/07]
5. Apache Lucene, <http://lucene.apache.org/> [Accessed 21/07]
6. Leonard Richardson, Sam Ruby. RESTful Web Services. O'Reilly Media; 1 edition (May 18, 2007), ISBN: 9780596529260
7. Pramod J. Sadalage, Martin Fowler: NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence. Addison-Wesley Professional; 1 edition (August 18, 2012). ISBN: 9780321826626
8. Apache Tika, <https://tika.apache.org/> [Accessed 21/07]
9. Rajkumar Buyya: High Performance Cluster Computing: Architectures and Systems. Prentice Hall; 1 edition (May 31, 1999). ISBN: 9780130137845

Implementación con Raspberry PI de un Servidor Portátil de Contenidos

Rocío Rodríguez, Pablo Vera, Daniel Giulianelli, Pablo Cammarano

Universidad Nacional de La Matanza, Grupo de Investigación, Desarrollo y Formación en Innovación de Software (GIDFIS), San Justo, Buenos Aires, Argentina
{rocio.rodriguez, pvera, dgiulian, pcmamarano}@unlam.edu.ar

Resumen. En diversos ámbitos la conexión a internet no está disponible o la misma no es lo suficientemente estable como para poder ser utilizada en forma constante por un conjunto de personas. Para estos casos donde el uso de servicios web puede ser beneficioso, es necesario contar con un servidor que no dependa de la conexión a internet, es decir un servidor de contenidos portátil. La Raspberry Pi es una minicomputadora portátil de bajo costo y con el poder de cómputo suficiente como para ejercer este rol. Este trabajo se enfoca en cómo utilizar la Raspberry Pi no sólo como servidor de contenidos portátil sino también como punto de acceso. Haciendo que este servidor portátil no requiera de infraestructura de red adicional, permitiendo a los usuarios finales acceder desde diversos dispositivos. Los diversos usos de este enfoque serán analizados en el presente trabajo.

Palabras Clave: Raspberry Pi, Servidor, Dispositivos Móviles

1 Introducción

La Raspberry Pi (RPi) es un SBC (Single Board Computer) que tiene el tamaño de una tarjeta de crédito, desarrollada en Reino Unido por la fundación Raspberry Pi.

La tabla 1, muestra las prestaciones en cuanto a hardware de los distintos modelos. El modelo más reciente RPi 3 [1] incorpora wifi. A los modelos previos era necesario agregarle por USB una pequeña placa externa.

Tabla 1. Características de los Modelos de Raspberry Pi

	1A	1B	1B+	2B	3B
RAM	256 MB		512 MB		1GB
Almacenamiento	SD				Micro SD
Procesador	ARM 11		ARM Cortex – A7		ARM v8
Velocidad	700 MHz		900 MHz		1,2 GHz
Conectividad de Red	no	Ethernet			Ethernet,WiFi,Bluetooth
Nº USB	1	2	4		
Nº GPIO	26		40		
Tamaño	85 x 56 x 17 mm				

La RPi se ha hecho masiva por su bajo costo, su pequeño tamaño que permite hacer portable la solución y la posibilidad de incorporar sensores si se requieren, lo cual permite una gran variedad de aplicaciones. Su popularidad puede analizarse a partir del interés por ella, medido por medio de Google Trends, en donde puede observarse en la figura 1, un pico de interés en el 2015, que produce el valor máximo de búsquedas efectuadas, considerado como 100%. Puede observarse también que la mayor parte del tiempo la cantidad de búsquedas se encuentran entre el 50% de dicho valor o mayor al 50%.



Fig. 1. Interés por el término “Raspberry Pi” medido por Google Trends

En cuanto a las regiones desde las cuales se han realizado la mayor cantidad de búsquedas, el máximo de búsquedas realizadas lo tiene Alemania (con ese valor se fija el 100%), en la tabla 2 se presentan los rangos de porcentajes (resumiendo la información obtenida en Google Trends). Los países marcados en negrita corresponden al porcentaje marcado en negrita en la misma fila, es decir que tienen por porcentaje uno de los valores de los límites del rango (por ejemplo: Alemania 100%, Chequia 50%, España 40%). En el intervalo de 80-70 no hay países por ello queda el renglón vacío en la tabla. Cabe destacar que el rango más bajo que iría de 10-0 se ha consignado de 10-4 dado que el país con menor tasa de búsqueda sobre la temática es Vietnam con el 4%.

Tabla 2. Interés por País

Rangos	Países
100-90	Alemania
90-80	Austria
80-70	----
70-60	Países Bajos, Suecia
60-50	Suiza, Chequia
50-40	Eslovenia, Finlandia, Estonia, Bélgica, Reino Unido, Portugal, Dinamarca, Noruega, España
40-30	Francia, Nueva Zelanda, Eslovaquia, Irlanda, Hungría, Singapur, Canadá, Australia
30-20	Sudáfrica, Estados Unidos, Italia, Hong Kong, Lituania, Croacia, Polonia, Corea del Sur, Grecia, Romania
20-10	Israel, Tunes, Bulgaria, Taiwan, China, India, Japón, Malasia, Chile, Serbia, Filipinas , Rusia
10-4	Emiratos Arabes Unidos, Ucrania, Argentina, Mexico, Brasil, Tailandia, Irán, Palistán, Colombia, Indonesia, Turquía, Perú, Vietnam

En base a lo indicado previamente la RPi es una solución de bajo costo, de interés para muchas regiones, con una gran diversidad de posibilidad para generar soluciones con el hardware que dispone. En cuanto al software, por medio de una tarjeta de memoria se puede instalar un sistema operativo. Existe una distribución de Linux particular para la RPi denominada Raspbian, pero también se pueden instalar otros sistemas operativos existiendo por ejemplo una versión específica de Windows 10 para este dispositivo. En cuanto al uso que se le puede dar a la RPi, existen diversos trabajos realizados en los últimos años, entre ellos se pueden mencionar aplicaciones dedicadas a:

- Domótica [2], [3], [4], [5];
- Monitoreo para Seguridad [6], [7], [8], [9];
- Streaming de Video/Audio [10], [11];
- Estación meteorológica [12], [13], [14];
- Agro [15], [16], [17], [18];
- Salud [19], [20], [21], [22];

Estos trabajos mencionados previamente muestran el gran abanico de acción que es posible alcanzar utilizando la Raspberry Pi. Lo cual la hace más que interesante para planificar diversas aplicaciones innovadoras. En el presente artículo se plantea el uso de la Raspberry Pi como servidor de contenidos, siendo este un punto de acceso para usuarios móviles. Sin requerir conexión a internet la Raspberry tendrá instalados los contenidos a ser consumidos. El presente artículo está organizado de la siguiente manera: En las secciones 2, 3 y 4 se explica cómo implementar dicho servidor de contenido, tomando en cuenta consideraciones de configuración y de software necesario, lo que incluye el desarrollo de un programa de monitoreo de performance. Luego en la sección 5 se analizan los múltiples usos que podría tener el contar con un servidor de contenidos, finalmente en la sección 6 se presentan las conclusiones y trabajos futuros.

2 Servidor de Contenidos

En este artículo se propone generar una red en la cual la Raspberry Pi (RPi) sea un servidor de contenidos y desde diversos dispositivos móviles puedan accederse a dichos contenidos. Para lo cual se realiza una red local en donde el punto de acceso a la misma estará provisto por la RPi.

Esta solución tiene por ventajas el acceso a contenidos:

- (1) Desde diversos dispositivos, entre ellos teléfonos celulares;
- (2) Sin necesitar tener una infraestructura de red existente.;
- (3) Sin contar con conectividad a internet.

Por otra parte, la portabilidad de la misma es importante, como se mencionó en la introducción la RPi tiene un tamaño pequeño, puede conectarse a la misma un teclado en miniatura (comercializables para Smart TV) que incorpora el mouse a través de un pad y una pantalla empotrada si se desea. Puede como pantalla por HDMI enchufarse a un monitor, televisor o proyector. También es posible que para determinadas soluciones no se requiera tener constantemente una pantalla. El resultado final será una solución de bajo costo.

En cuanto a los requerimientos se pueden mencionar principalmente:

1. Que diversos tipos de clientes (notebooks, tablets, smartphones) puedan conectarse a una red inalámbrica creada y administrada por la Raspberry Pi. Dicha red no tiene que depender de una conexión a Internet.
2. Que un servidor web se pueda ejecutar en la Raspberry Pi, y que permita correr aplicaciones web que interaccionen con una base de datos.
3. Que exista un sistema de resolución de nombres (DNS) en la Raspberry Pi, permitiendo que los clientes accedan a las herramientas ofrecidas por el servidor mediante nombres fáciles de recordar. Por ejemplo: www.ejercicios.com.ar

3 Implementación

En cuanto a sistema operativo, para crear una solución de bajo costo, se selecciona Linux en su distribución particular para la RPi denominada Raspbian. “Linux es una gran opción para la raspberry pi ya que es gratuito y de código abierto. Por un lado mantiene el precio de la plataforma bajo y por el otro lo hace más configurable” [23]. “Raspbian es un sistema operativo gratuito basado en Debian (distribución de linux) y optimizado para el Hardware de la Raspberry Pi” [24]. Se instala además: Apache (Servidor Web), MySQL (Base de Datos). Sin costos asociados con el software, se desarrolló una solución que permita monitorear la performance ante múltiples accesos (lo cual es detallado en la sección 4). A continuación, se presentan en forma resumida, los pasos necesarios para la configuración del entorno:

- a) Configuración y acceso a la red: Comparando diferentes arquitecturas de red: redes ad-hoc, redes mesh y redes centralizadas. Se decide implementar la opción de red centralizada, siendo la Raspberry el único elemento que forma parte de la infraestructura de la red (junto con su adaptador wifi USB), dado que la raspberry disponible es Modelo 2 y no cuenta con placa integrada (tal como se detalló en la tabla 1). Para ello, es necesario configurar a la Raspberry como un punto de acceso (Access Point – AP), también llamado “hotspot”. Es importante destacar que no todos los adaptadores wifi pueden funcionar como un punto de acceso. El que se utilizó, “Ralink Technology Corp RT 5370 Wireless Adapter”, es capaz de hacerlo. Del lado del software, hay que instalar dos componentes: hostapd y udhcpd, que brindan diferentes funcionalidades. Mientras que hostapd se encarga de administrar el punto de acceso y ofrecer servicios de autenticación, udhcpd actúa como un servidor DHCP, para asignarles direcciones IP a los clientes de la red dinámicamente. Cabe destacar que existen otros componentes para estas mismas funcionalidades. La instalación de los mismos se realiza mediante la utilidad de gestión de paquetes de la Raspberry Pi (apt-get) y la configuración se hace a través de archivos, ubicados en la carpeta “/etc”. Existe la posibilidad de hacer un forwarding del tráfico de la red wifi hacia el puerto Ethernet de la Raspberry Pi que, si se encuentra conectado a Internet, puede ofrecer conectividad a los clientes de la red.

- b) Servidor Web: Se instaló el tradicional stack LAMP (Linux-Apache-MySQL-PHP). El proceso de instalación se hace utilizando la herramienta apt-get. También se instala por separado el servidor web, Apache 2, al igual que php5 y MySQL como motor de base de datos. También se instaló phpMyAdmin, que ofrece una interfaz gráfica para interactuar con bases de datos propias. Luego de la instalación, se agregan algunas páginas en “/var/www/html”, para verificar el correcto funcionamiento del servidor web.
- c) Servidor DNS: Para acceder a las páginas web del servidor, es necesario introducir la dirección IP de la Raspberry Pi en el navegador, algo que es incómodo y poco usable para los clientes de la red. Por lo tanto, se instaló un servidor DNS en la Raspberry Pi. El paquete software se llama “dnsmasq” y se instala mediante apt-get. La instalación requirió cambiar una configuración en el servidor DHCP, cambiando el servidor DNS por defecto a la propia Raspberry, de forma tal que las peticiones DNS de los clientes sean resueltas por dnsmasq. Esta utilidad cuenta con bastantes configuraciones, por el momento sólo se limitó al mapeo de las URLs a la dirección IP de la Raspberry, donde se está ejecutando el servidor web. Por ejemplo, la siguiente línea de configuración:
`address=/www.ejercicios.com.ar/192.168.0.1`
 permite que los clientes accedan a esa URL en lugar de conocer la dirección IP de la Raspberry. Cabe destacar, que no hay conflictos con dominios ya existentes en la web, debido a la naturaleza offline de este proyecto.

En la tabla 3, se presenta a modo de resumen los paquetes que fueron necesario instalar.

Tabla 3. Paquetes utilizados

Paquete	Descripción	Utilidad
Hostapd	Demonio (servicio) que se encarga de administrar el punto de acceso y además ofrece servicios de autenticación.	Para transformar la Raspberry Pi en un punto de acceso, con capacidades de servidor DHCP y DNS.
Udhcpd	Servidor DHCP. Está incluido en la utilidad “busybox”.	
Dnsmasq	Servidor DNS.	
apache2	Servidor Web.	Hosting de páginas dinámicas que interactúen con una base de datos.
php5	Lenguaje server-side.	
mysql-server	Motor de base de datos.	
phpmyadmin	Interfaz web para administrar bases de datos MySQL.	
Npm	Gestor de paquetes de JavaScript.	Monitoreo de performance de la red que corre en segundo plano. También podrían utilizarse para hostear servicios web que sean consumidos por aplicaciones web.
Nodejs	Node.js, servidor JavaScript	
pm2	Gestor de procesos para Node.js	
Iw	Utilidad de Linux para gestionar dispositivos y conexiones inalámbricas.	

4 Monitoreo de performance

Teniendo en cuenta que el ámbito de aplicación podría requerir de la conexión de múltiples usuarios, es importante saber cuántos podrán conectarse simultáneamente a la red creada por la Raspberry Pi. Para ello, se desarrolló una aplicación Node.js, que corre en segundo plano mientras la Raspberry está en modo punto de acceso. Dicha aplicación monitorea la cantidad de usuarios conectados a la red, almacenando la cantidad máxima de usuarios y el tiempo que dichos usuarios permanecieron conectados a la red en un archivo de log en la Raspberry. En la figura 2, se muestra un log generado.

Monitoreo N°	Fecha y hora de inicio	Máxima cantidad de usuarios	Duración
21	11-07-2017--19:46:48	5	45 minutos
32	12-07-2017--19:37:07	10	30 minutos

Fig. 2. Log con los datos de los usuarios conectados

Además, la aplicación ofrece una interfaz web donde se puede ver la cantidad actual de usuarios conectados a la red, con información específica de cada uno, por ejemplo: cantidad de tráfico enviado y recibido (medido en bytes y en paquetes), intensidad de la señal (medido en dBm) y la tasa de bits (medido en MBit/s). Se utilizó el paquete de Linux “iw” para obtener dicha información. Para ejecutar la aplicación Node.js en segundo plano se usó el gestor de procesos “PM2”. Además, fue necesario configurar un reverse proxy en el servidor web Apache, para acceder a la aplicación Node.js sin necesidad de incluir en la URL el puerto en el cual corre la aplicación. Se usó “socket.io” para transferir información en tiempo real sobre la red desde la Raspberry hacia los clientes. En la figura 3 se muestra la captura de pantalla tanto del lado de la RPi como desde uno de los dispositivos móviles donde puede observarse los datos de monitoreo de performance en un caso reducido de 2 usuarios conectados.

Actualmente las pruebas se realizaron dentro del Laboratorio del GIDFIS (Grupo de Investigación, Desarrollo y Formación en Innovación de Software), con un total de 15 usuarios accediendo a la RPi en simultáneo, esto evidenció el acceso sin inconvenientes. Los usuarios en el día de prueba accedieron desde diversos equipos tal como se muestra en la tabla 4.

Tabla 4. Prueba con Usuarios conectados en simultáneo

Cantidad de Usuarios	Equipo
5	Computadoras de Escritorio
2	Notebooks
6	Celulares
1	Tablet
1	Lector de Libros

Si bien 15 usuarios ya es una cantidad interesante para poder pensar aplicaciones de diversos tipos, se ha planificado más adelante incrementar la cantidad de usuarios para las pruebas solicitándoles a docentes y alumnos que ingresen a la RPi para visualizar un contenido. Previéndose pruebas con distinta cantidad de usuarios, por ejemplo: 50, 60, 100 conectados simultáneamente. Esto permitirá poder evidenciar que sucede cuando se aumenta la cantidad de usuarios.

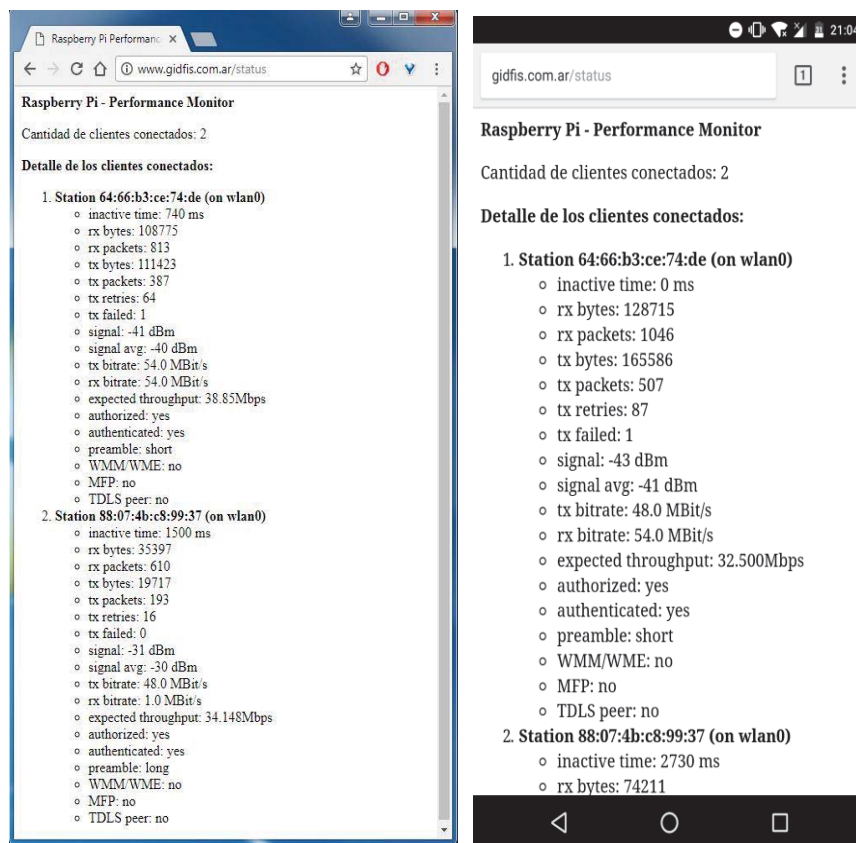


Fig. 3. Capturas de Pantallas del Software de Monitoreo Desarrollado

5 Usos Posibles

La Raspberry Pi como punto de acceso y servidor de contenidos sin necesidad de conexión a internet puede aprovecharse en diversos escenarios:

- Educación: en aulas donde no hay conexión a internet o la misma no está disponible permanente, puede utilizarse para:
 - Poner a disposición contenidos web pre-diseñados (texto, imágenes, audio, video).

- Generar preguntas o pequeñas trivias que puedan ser respondidas por los alumnos desde sus dispositivos durante la clase, a fin de sondear la opinión, conocimiento o entendimiento de determinado tema.
 - En aulas rurales donde no hay ninguna posibilidad de acceso a internet los docentes podrían incluso clonar algunos sitios útiles para trabajar en clase.
- Vehículos: en movimiento la conexión móvil a internet no siempre está disponible y además se pueden atravesar zonas sin cobertura, por lo tanto, una solución de este tipo podría ser de utilidad en distintos contextos:
 - Turismo: las agencias de turismo, en sus micros o mini-van al realizar excursiones pueden ofrecer contenido para que los turistas puedan ir consultando desde sus propios dispositivos para complementar las explicaciones del guía. Incluso pueden brindarse audio guías en distintos idiomas, videos, etc.
 - Autos particulares: para viajes largos la Raspberry Pi puede configurarse como un media center móvil pre-cargando, algunas películas, series o música que mediante streaming se pueda consumir desde un dispositivo móvil.
 - Taxis, Remises y Colectivos: si bien podría ofrecerse el mismo servicio que para autos particulares, en general, los viajes realizados en estos medios de transporte son más cortos, por lo que podrían clonarse por ejemplo algunos sitios web, de actualidad o de noticias para que el pasajero pueda ir leyendo durante el viaje sin utilizar su conexión a internet. También podría incluirse un módulo GPS para que el pasajero pueda conocer la velocidad del vehículo en el que se desplaza.
 - Como nodos de una red VANET (Vehicular Ad-Hoc Network) permitiendo la comunicación entre diferentes vehículos en forma directa. Estas redes pueden intercambiar en tiempo real información con otros vehículos como posición, velocidad y ser utilizadas para la prevención de accidentes alertando al usuario sobre otros vehículos en la zona.
- Supermercados: para brindar un servicio de consulta de precios sin necesidad de acceder físicamente a un lector. Incluso podría ofrecerse un servicio que permita ir calculando el gasto total realizado por el cliente.
- Salas de espera: además de brindar algunos de los servicios de entretenimiento antes mencionados, también podría generarse un sistema para reservar turnos para distintos trámites, sin necesidad de que todas las personas tengan que acceder a un único punto para sacar un número. Otra aplicación es consultar en que número está cada tipo de trámite ya que muchas veces no es visible por falta de espacio en las pantallas de los establecimientos (por ejemplo: en bancos donde se ven los últimos números llamados, pero como hay tanta variedad no se pueden mostrar los últimos números de todos los trámites disponibles).

6 Conclusiones

La implementación de la RPi como servidor de contenidos, permite tener un servidor de bajo costo que despliegue una gran cantidad de contenidos y servicios. La implementación requiere la selección de software adecuado y la configuración de la Raspberry para que pueda funcionar como nodo central de la red.

Las pruebas realizadas muestran que tiene una buena performance incluso ante diversos accesos simultáneos. La posibilidad de acceder a una red incluso sin la necesidad de contar con internet y con usuarios que accedan desde dispositivos móviles, le da una gran flexibilidad. Con lo realizado es posible consumir contenidos alojados en la Raspberry Pi. Como trabajo futuro se ha planificado generar mayor cantidad de accesos de usuarios móviles para poder realizar pruebas de performance. Pero no sólo performance en cuanto a accesos simultáneos sino también medir el rendimiento si se utiliza para streaming de video. Se considera además el desarrollo futuro algunos de los múltiples tipos de aplicaciones posibles.

Referencias

1. Raspberry Pi 3 Model B, <https://www.raspberrypi.org/products/raspberry-pi-3-model-b/>
2. Suryawanshi, S. G., & Annadate, S. A. (2016). Implementation of Smart Home Automation System through E-mail using Raspberry Pi and Sensors. <http://ijireeice.com/upload/2016/march-16/IJIREEICE%2047.pdf>
3. Pampattiwar, K., Lakhani, M., Marar, R., & Menon, R. (2017). Home Automation using Raspberry Pi controlled via an Android Application. <http://inpressco.com/wp-content/uploads/2017/05/Paper33962-967.pdf>
4. Kulkarni, B. P., Joshi, A. V., Jadhav, V. V., & Dhamange, A. T. (2017). IoT Based Home Automation Using Raspberry PI.
5. Segrera, C. C., & Calderón, M. (2017). Hogar Inteligente por Control de Voz Usando Redes Neuronales. *Revista Colombiana de Tecnologías de Avanzada (RCTA)*, 1(25).
6. Ibrahim, M., Elgamri, A., Babiker, S. and Mohamed, A., 2015, October. Internet of things based smart environmental monitoring using the Raspberry-Pi computer. In *Digital Information Processing and Communications (ICDIPC), 2015 Fifth International Conference on* (pp. 159-164). IEEE.
7. Ahmad, T., Studiawan, H. and Ramadhan, T.T., 2014. Developing a Raspberry Pi-based Monitoring System for Detecting and Securing an Object. <http://kbj.if.its.ac.id/wp-content/uploads/2015/03/2014-IES-PENS.pdf>
8. Snyder, R.M., 2014. Power monitoring using the Raspberry Pi. *Association Supporting Computer Users in Education "Our Second Quarter Century of Resource Sharing"*, p.82.
9. Yaldaie, A., 2016. Home automation and security system with the Raspberry Pi. <http://www.theseus.fi/handle/10024/106677>
10. Giulianelli, D. A., Rodríguez, R. A., Vera, P. M., Conca, A., Valles, G. Y., & Cammarano, P. (2016, May). Implementación de un sistema de video conferencia HD autónomo utilizando Raspberry Pi. In *XVIII Workshop de Investigadores en Ciencias de la Computación (WICC 2016, Entre Ríos, Argentina)*.
11. Khamankar, A., Phirke, A., Shah, K., Rangare, D., & Shinde, P. A. (2017). Portable voice communication system on raspberry pi. <https://www.irjet.net/archives/V4/i2/IRJET-V4I2295.pdf>

12. Salcedo Tovar, M.L. and Cendrós, J., 2016. Uso del minicomputador de bajo costo "Raspberry Pi" en estaciones meteorológicas. *Télématique*, 15(1), pp.62-84.
13. Anchundia, R. and Michael, J., 2015. Telemetría de Estación Meteorológica. <http://www.dspace.espol.edu.ec/xmlui/handle/123456789/31545>
14. Malfatti, G.M., Pavan, W. and Cunha, J.M., 2014. Rede de estações meteorológicas automáticas sem fio. <http://ainfo.cnptia.embrapa.br/digital/bitstream/item/133214/1/ID-43372-2014-TCC-MiniColetorDados.pdf>
15. Escalas Rodríguez, G., 2015. Diseño y desarrollo de un prototipo de riego automático controlado con Raspberry Pi y Arduino. <http://upcommons.upc.edu/handle/2099.1/25074>
16. Parrales, L., Enrique, L., & Safadi Figueroa, H. E. (2017). Diseño y control de ambiente interno de un vivero modular para cultivo en interiores de plantas comestibles (Bachelor's thesis, Espol). <http://www.dspace.espol.edu.ec/xmlui/handle/123456789/37278>
17. James, J. (2016, December). Plant growth monitoring system, with dynamic user-interface. In Humanitarian Technology Conference (R10-HTC), 2016 IEEE Region 10 (pp. 1-5). IEEE. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7906781&isnumber=7906774>
18. Vaccaro Acosta, F.E., Soriano, A. and Alberto, J., 2015. Telecontrol de secadora de granos de cacao con energía solar y biogas. <http://www.dspace.espol.edu.ec/xmlui/handle/123456789/30149>
19. Trivi, M. R., Rabal, H., Grumel, E. E., Cap, N. L., & Buffarini, L. (2017, May). Técnicas de speckle dinámico: desarrollos y perspectivas. In IV Jornadas de Investigación, Transferencia y Extensión de la Facultad de Ingeniería (La Plata, 2017) http://sedici.unlp.edu.ar/bitstream/handle/10915/60107/Documento_completo.pdf-PDFA.pdf?sequence=1
20. Thota, J., Vangali, P., & Yang, X. (2017). Prototyping An Autonomous Eye-Controlled System (AECS) using Raspberry-Pi on Wheelchairs. *International Journal of Computer Applications*, 158(8).
21. Haney, J., Lim, S., & Chong, R. (2017, March). Automated sleep data extraction and streaming using Zeo sleep personal manager and Raspberry Pi. In SoutheastCon, 2017 (pp. 1-2). IEEE. <http://ieeexplore.ieee.org/abstract/document/792527/>
22. Sethia, D., Singh, S., & Singhal, V. (2017). ABE Based Raspberry Pi Secure Health Sensor (SHS). In *Advances in Ubiquitous Networking 2* (pp. 599-610). Springer Singapore. https://link.springer.com/chapter/10.1007/978-981-10-1627-1_47
23. Richardson, M. and Wallace, S., 2012. Getting started with raspberry PI. " O'Reilly Media, Inc."
24. Raspbian, <https://www.raspbian.org/>

Integración de Sistemas ERP para el Monitoreo de Planes y Detección de Eventos Disruptivos en Cadenas de Suministros

Erica S. Fernández^{1,2}, Pedro L. Querini¹, Carlos M. Toledo³

¹ Depto. Ing. Industrial UTN, FRRA.

M. Acuña 49, Rafaela, Argentina

² GEMPRO – UTN, FRSF

Lavaisse 610, Santa Fe, Argentina

³ Instituto de Desarrollo y Diseño INGAR (CONICET-UTN)

Avellaneda 3657, Santa Fe, Argentina

erica.fernandez@frra.utn.edu.ar

pedroquerini@gmail.com

cmtolledo@santafe-conicet.gov.ar

Resumen. Los sistemas de Planificación de Recursos Empresariales (ERP Enterprise Resource Planning) se definen como un conjunto de aplicaciones que permiten gestionar de manera integrada los procesos de negocio de las empresas. En la actualidad, los sistemas ERP carecen de flexibilidad suficiente para generar Cadenas de Suministros (CS) flexibles capaces de reaccionar rápidamente a eventos disruptivos. Es por ello, en este trabajo se propone un servicio MASM-ERP que integra los planes de abastecimiento, producción y distribución definidos en los ERP de las empresas involucradas en una CS. También permite monitorear la ejecución de estos planes y notificar la ocurrencia de eventos disruptivos a los actores pertinentes. El monitoreo durante la ejecución de planes permite anticiparse a los cambios que podrían tener lugar en el horizonte de tiempo considerado y mejorar los procesos de toma de decisión.

Palabras claves: ERP, Cadena de Suministro, Eventos Disruptivos.

1 Introducción

Una Cadena de Suministro (CS) es una red de empresas autónomas, colectivamente responsables por la obtención, producción y distribución de productos [1]. Los ambientes altamente competitivos y dinámicos en los que actualmente opera la CS generan fuertes obligaciones entre los participantes con el fin de lograr ventajas competitivas. Tal condición hace necesario establecer vínculos más estrechos entre las empresas que la integran. Los proveedores, mayoristas y minoristas se ven como “socios”, comparten información, delinean planes de negocios, ventas y promociones, analizan en forma conjunta la demanda, planifican y generan los planes de

abastecimiento y distribución. Esta nueva forma de relacionarse se denomina *colaboración*.

Las CS deben ser gestionadas de tal forma que los productos sean producidos y distribuidos de modo correcto (cantidad, lugar y fecha) para minimizar los costos totales cumpliendo con los requerimientos de los clientes. En los últimos años, las empresas que pertenecen a una CS han incorporado a sus prácticas habituales la utilización de los sistemas de Planificación de Recursos Empresariales (ERP, Enterprise Resource Planning) [2,3,4]. Los ERP, compuestos por módulos completamente integrados, posibilitan la disminución de los costos operativos, mejoran el servicio al cliente y la productividad, entre otros.

Tradicionalmente, el desarrollo de sistemas ERP se focalizaba en proveer funcionalidades para dar respuestas a requerimientos propios de una empresa. Sin embargo, en un entorno competitivo, el éxito de una empresa depende de su habilidad para establecer relaciones de colaboración con otras empresas, razón por la cual los sistemas ERP evolucionaron incorporando módulos relacionados a la gestión de la CS (SCM, Supply Chain Management), sistemas avanzados de planificación (APS, Advances Planning and Scheduling), gestión de relaciones con el cliente (CRM, Customer Relationship Management) y negocios inteligentes (BI, Business Intelligent).

A pesar de estos nuevos módulos, en la actualidad, los ERP carecen de flexibilidad suficiente para generar CS flexibles capaces de detectar/anticipar eventos disruptivos, tales como: la disminución de la cantidad de una orden o el retraso de la fecha de entrega de un producto. A partir de la detección/anticipación de estos eventos, las empresas pueden reaccionar rápidamente y ajustar sus planes de ejecución [5,6].

Los Sistemas de Información para la Gestión de Eventos de la CS, conocidos por su sigla en inglés SCEM (Supply Chain Event Management) [7-11] son herramientas tecnológicas que dan respuesta a los requerimientos funcionales de monitoreo de los planes en ejecución para detectar y/o anticipar la ocurrencia de un evento disruptivo [1,12-14]. Además de estas funcionalidades estos sistemas son capaces de evaluar la factibilidad de un plan afectado por un evento disruptivo y definir una estrategia de solución para reparar el plan afectado utilizando holguras previamente consideradas [15-16].

En la literatura se han propuesto enfoques de monitoreo reactivo [17,18], predictivos [19, 20] y reactivo/predictivo [21,22]. Cada uno de estos trabajos recolecta datos de ejecución de los planes para el monitoreo de órdenes [23, 21], recursos [17, 19, 20] u órdenes/recursos [18, 23]. Entre estos enfoques, Fernández et al. desarrollaron un sistema multi-agente de monitoreo (MASM, Multi-Agent Monitoring System) que realiza la función de monitoreo de planes de un sistema SCEM [13]. Durante la ejecución de un plan integrado, los agentes recolectan datos de las órdenes y recursos de dicho plan y; del entorno a efectos de evaluarlos para detectar/ anticipar eventos disruptivos.

A pesar de las funcionalidades, MAMS permite únicamente el monitoreo de planes previamente sincronizados. Además, MAMS no se encuentra integrado a los sistemas ERPs de las empresas que integran la CS y la recopilación de los datos debe realizarse en diferentes lugares de las empresas afectando los tiempos de búsqueda y procesamiento y, aumentando la complejidad en la implementación de la recolección y duplicación de datos.

Ante esta falta de integración, este trabajo propone una extensión (MASM-ERP, Multi-Agent Monitoring System for Enterprise Resource Planning) que permite integrar los ERP mediante la sincronización de los planes de abastecimiento, producción y distribución en una CS, como así también monitorear la ejecución de estos planes y notificar la ocurrencia de eventos disruptivos a los actores pertinentes a fin de realizar las correcciones necesarias.

El resto del trabajo se estructura de la siguiente manera: la Sección 2 presenta la arquitectura MASM-ERP, la integración y monitoreo de planes para detectar eventos disruptivos, la Sección 3 presenta un caso de estudio para mostrar el funcionamiento de MASM-ERP y por último la Sección 4 presenta las conclusiones y trabajos futuros.

2 Integración de sistemas ERPs para sincronizar y monitorear planes en una CS

2.1 Arquitectura de integración

Para la sincronización y monitoreo de planes y la detección de eventos disruptivos en una CS, la extensión MASM-ERP implementa un Servicio de Datos ERP que permite a los Agentes ERP ubicado en las empresas enviar datos extraídos de la base de datos de los ERP (Fig. 1). Además, cada Agente ERP implementa un Servicio de Notificación de Eventos que le permite a MASM-ERP comunicarse para solicitar datos adicionales ante la posible presencia de un evento disruptivo como así también informar a las empresas sobre su ocurrencia.

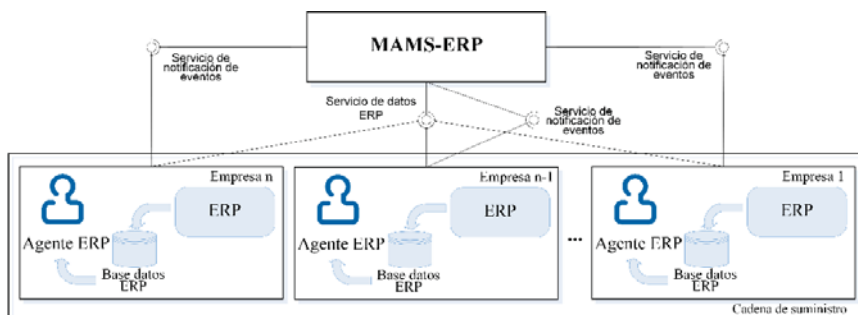


Fig. 1. Arquitectura del servicio MASM-ERP

Durante la ejecución de un plan integrado, cada Agente ERP extrae datos relacionados a los módulos de compras, manufactura, depósitos y ventas que se encuentran en las bases de datos del ERP. Los principales datos de compras incluyen aquellos relacionados al proceso de aprovisionamiento de productos/subproductos/materia prima incorporados en las órdenes de compra. La Fig. 2 presenta un ejemplo de los datos de un plan de compras de una empresa que

tiene planificado recibir de un proveedor X, 100 unidades del producto A en la fecha 21 de octubre de 2017.

Número de Orden	Producto/Subproducto	Proveedor	Fecha de entrega	Cantidad
O1	Producto A	Empresa X	21/10/2017	100

Fig. 2. Datos extraídos de un plan de abastecimiento

Los principales datos de manufactura están relacionados a las órdenes de fabricación, incluyendo la cantidad a producir de un producto/subproducto, fecha de producción. Una vez que las empresas definen una orden de producción, se consumen los materiales necesarios del depósito. Para ello, las empresas deben definir la BOM (Bill of Materials, Lista de Materiales) y los datos de los artículos que ingresan a los almacenes. La BOM contiene una descripción completa del producto enumerando los materiales, las partes, los componentes y la secuencia en que el producto es fabricado.

Los principales datos de ventas están relacionados con los planes de ventas de los productos/subproductos/materia prima. Si los productos incluidos en un plan no se encuentran en el inventario, el módulo de compra del ERP genera automáticamente los pedidos para que estén a disposición los materiales necesarios y comunica esta información. En caso de ser una fábrica, se emite además una orden de producción. Seguidamente, se contabiliza los recursos materiales disponibles en el almacén y en caso que estos no sean suficientes, y se emite un presupuesto hacia el proveedor del producto, el cual se transformará en orden de compra una vez confirmado manualmente el mismo.

Con estos datos incorporados a las bases de datos de los ERP de cada empresa de la CS, el Agente ERP accede a los mismos y los notifica al MASM-ERP. A partir de estos datos, el MASM-ERP sincroniza los planes de suministro utilizando un algoritmo específico que se detalla a continuación.

2.2 Sincronización de planes

Dado un conjunto de empresas $E = \{e_1, e_2, \dots, e_n\}$ que pertenecen a una CS que se integran sus sistemas ERP para la detección de eventos disruptivos y un plan P compuesto por un conjunto de órdenes $O_j \in P$ que representan la compra, venta o producción del producto j . A cada empresa e_i se les asigna un rol: Proveedor-Fabricante-Cliente. Una empresa proveedora e_i^p tendrá un conjunto de órdenes de venta (O^v) que pertenecen a su programa de distribución. Una empresa cliente e_i^c tendrá un conjunto de órdenes de compra (O^c) que pertenecen a su programa de abastecimiento. Una empresa fabricante e_i^f tendrá un conjunto de órdenes de compra (O^c), de producción (O^p) y de venta (O^v).

Definida las empresas interesadas en participar de la colaboración y el horizonte de planificación, se selecciona la empresa e_n^c que pertenece al último eslabón de la CS con el rol cliente y busca las órdenes de compra (O_j^c) de un producto j de dicha empresa y que es objeto de la colaboración. A continuación, identifica las órdenes de

venta O_j^v de la empresa e_{n-1}^v y producto j que van a cumplimentar con dichas O_j^c , resultando una relación $O_j^c - O_j^v$.

Luego se pueden generar dos alternativas. En la primera alternativa, la empresa e_{n-1}^f a la que le pertenece la O_j^v tiene un rol Fabricante. En esta alternativa, se establece la asociación entre la orden de venta O_j^v y la orden de producción O_j^p del producto j . Posteriormente, se asocia la orden de producción O_j^p con las órdenes de compra O_k^c de cada subproducto o materia prima k necesaria para la elaboración de j . Finalmente, se asocia cada orden de compra O_k^c de la empresa e_{n-1}^f con las órdenes de venta O_k^v de la empresa con el rol proveedor e_{n-2}^p .

En la segunda alternativa, la empresa e_{n-1}^p a la que pertenece la O_j^v tiene un rol de proveedora, entonces se debe establecer la relación entre las órdenes O_j^c de compra de la empresa e_n^c y las órdenes de venta O_j^v de la empresa e_{n-2}^p .

Este proceso se repite hasta obtener la trazabilidad de cada producto/subproducto/materia prima (estableciendo asociaciones entre órdenes) resultando el plan sincronizado P de la CS. Dicho plan estará compuesto de un conjunto de órdenes con sus atributos cantidad de producto j o subproducto/materia prima k y la fecha de envío, arribo o producción según el rol de la empresa. A partir de este plan se monitorea y detectan eventos disruptivos.

2.3 Monitoreo y detección de eventos disruptivos

Obtenido el plan sincronizado P para una CS, P debe ser monitoreado a efectos de detectar dichos eventos disruptivos que pueden afectar a O^v, O^c u O^p del conjunto $O_j \in P$.

Cuando comienza la ejecución del plan P , cada orden $O_j \in P$ es controlada. La empresa a la que pertenece la orden mantiene informado al MASM-ERP sobre el estado de la orden (recibida, en producción, enviada) junto con la cantidad especificada y en la fecha planificada. El MASM-ERP analiza la existencia de desvíos. Si no hay desvíos, la orden ha sido cumplimentada satisfactoriamente. Caso contrario se debe analizar si el desvío representa una interrupción. Para ello, la empresa informa al MASM-ERP la holgura permitida que puede ser de cantidad de producto/subproducto/materia prima (recibir menor cantidad) o tiempo (una orden se retrasa o adelanta) [15-16].

La empresa que posee el inconveniente (que puede dar lugar a una interrupción) debe especificar también cuánto material del planificado va a poder satisfacer y/o con cuánto tiempo de atraso se entregará. Esta información es necesaria para que las empresas de la CS puedan tomar mejores decisiones. Por ejemplo, una empresa es notificada que va a recibir una menor cantidad de material de una orden de abastecimiento, dicha empresa podría verificar su nivel de inventario y decidir que recibir esa menor cantidad no afecta la ejecución de su plan, entonces la interrupción no ocurrirá.

Para la anticipación de eventos disruptivos, el sistema MASM-ERP utiliza funciones de evaluación predictivas implementadas en MASM propuestas por Fernández et al. [1,12-14]. Estas funciones utilizan una Red Bayesiana que permite

representar conocimiento incierto y razonar con base en teoría de probabilidades. Esta red incluye una serie de nodos que permite recolectar datos de variables del proceso de suministro y/o producción (proceso para ejecutar una orden), y están conectados entre sí con cierto grado de influencia expresados a través de un grado de probabilidad de ocurrencia de un evento disruptivo. A partir de estos valores de probabilidades y de un razonamiento probabilístico, se puede anticipar con suficiente grado de certeza que, dada las condiciones actuales del proceso, en un futuro puede ocurrir un evento disruptivo. Esto les permite a las empresas anticiparse y generar planes de contingencia para poder corregir y/o disminuir el impacto de los eventos que se produzcan.

Cuando hay una disrupción, el MASM-ERP notifica mediante el envío de un mensaje a las empresas que tienen órdenes relacionadas al evento. Esta notificación se propaga a todas las empresas afectadas en la CS.

2.4 Implementación

La extensión MASM-ERP como el sistema MASM fueron implementados con la plataforma de programación de agentes JaCaMo [13,24]. JaCaMo abarca los tres principales niveles de abstracción de un sistema multi-agente: agente, organización y entorno. JaCaMo permite la programación de agentes BDI (Belief, Desire and Intention – Creencias, Deseos e Intenciones) gobernados por modelo organizacional Moise, y el soporte del metamodelo A&A [25]. Moise describe las colaboraciones organizacionales y aborda el comportamiento colectivo por las restricciones de comportamiento individuales [26]. El metamodelo A&A permite la implementación de mecanismos de coordinación basada en el medio ambiente, y servicios no autónomos y herramientas [27].

JaCaMo fue seleccionada ya que es una plataforma de programación operacional que incluye soporte integrado para la programación de artefactos, lo que facilitó la integración de funcionalidades no orientadas a objetivos tales como los modelos de detección y anticipación de eventos disruptivos del sistema MASM. Además de JaCaMo permite la implementación de servicios web mediante el lenguaje WSDL (Web Services Description Language) [28].

3 Caso de Estudio

3.1 Presentación

El caso de estudio consiste en una CS integrada por tres empresas. La empresa A es un tambo el cual posee 192 vacas y una capacidad de ordeño de 144.000 litros mensuales. Se encuentra ubicado en la ruta provincial N°13 en la localidad el Trebol – Santa Fe y posee como cliente exclusivo la empresa B.

La empresa B produce quesos duros, semiduros y blandos, manteca y crema. Debido a su posicionamiento como empresa líder del sector, la elaboración de quesos representa el 40% de su producción. Actualmente, procesa 360.000 litros de leche

mensuales y recibe 144.000 litros mensuales de la empresa A. También se abastece de leche de otros tambos no integrados a la CS. Su planta industrial está emplazada en la localidad de Rafaela y la comercialización de sus productos la hace a través de ventas directas, distribuidores del rubro y supermercados hiper-mayoristas.

La empresa C es un distribuidor hiper-mayorista y uno de los productos que comercializa son hormas de queso duro. Se encuentra ubicada en la ciudad de Buenos Aires. Posee varios proveedores, pero recibe la mayor cantidad de hormas de queso duro de la empresa B debido a la calidad del producto.

La empresa A y B se encuentran a una distancia aproximada de 130 km por lo que se envían todos los días de la semana 4.800 lts/día. La B y C se encuentran a mayor distancia (550 km aproximadamente) por lo que el tiempo de entrega de las hormas de queso duro a la empresa C son cada día por medio y se envían 200 hormas de queso en cada viaje. La empresa B aprovecha el camión para el envío de otros productos a la provincia de Buenos Aires.

Para usar el servicio MASM-ERP, las empresas que integran la CS establecen un contrato el cual define que se va a monitorear la materia prima leche (Le) y el producto queso duro (Qd). La codificación Le y Qd se asigna a efectos de evitar problemas de interoperabilidad semántica. Además, se define el horizonte de planificación en 14 días (periodo de tiempo durante el cual se llevará a cabo el monitoreo del plan sincronizado). Luego, cada empresa define su plan de abastecimiento, producción y/o distribución en su sistema ERP (Sección 3). La empresa A (rol Proveedor) tendrá un conjunto de órdenes de venta O^v que pertenecen a su plan de distribución. La empresa B (rol Fabricante) tendrá un conjunto de órdenes de compra O^c , producción O^p y venta O^v que pertenecen a 3 planes (abastecimiento-producción-distribución). La empresa C (rol cliente) tendrá un conjunto de órdenes de compra O^c que pertenecen a su plan de abastecimiento. A continuación, se definen un conjunto de posibles escenarios que pueden presentarse.

3.2 Escenario inicial

Definidos los planes, el agente ERP extrae los datos relacionados a los módulos de compras, manufactura y ventas que se encuentran en la base de datos del ERP de cada empresa y son enviados al MASM-ERP, el cual genera un plan sincronizado para un horizonte de planificación de 14 días. La Fig.3 resume y sintetiza los datos de un plan sincronizado extraído de las bases de datos del sistema MASM-ERP. Esta sincronización implica que la $O^c = \{Qd: 200, fecha: 06/06/2017\}$ de la empresa C, se encuentra relacionada a las $O^v = \{Qd = 200, fecha: 05/06/2017\}$ y a dos $O^p = \{Qd: 100, fecha: 03/06/2017, Qd: 100, fecha: 05/06/2017\}$ de la empresa B. A su vez, las 2 O^p se encuentran relacionadas a las cuatro O^c de la misma empresa y a las cuatro O^v de la empresa A.

Empresa A: Ov			Empresa B: Oc			Empresa B: Op			Empresa B: Ov			Empresa C: Oc		
Fecha	Materia Prima	Cantidad	Fecha	Materia Prima	Cantidad	Fecha	Producto	Cantidad (hormas)	Fecha	Producto	Cantidad (hormas)	Fecha	Producto	Cantidad (hormas)
01/06/2017	Le	4800 litros	01/06/2017	Le	4800 litros	01/06/2017	Qd	100	01/06/2017	Qd	200	01/06/2017	Qd	
02/06/2017	Le	4800 litros	02/06/2017	Le	4800 litros	02/06/2017	Qd		02/06/2017	Qd		02/06/2017	Qd	200
03/06/2017	Le	4800 litros	03/06/2017	Le	4800 litros	03/06/2017	Qd	100	03/06/2017	Qd	200	03/06/2017	Qd	
04/06/2017	Le	4800 litros	04/06/2017	Le	4800 litros	04/06/2017	Qd		04/06/2017	Qd		04/06/2017	Qd	200
05/06/2017	Le	4800 litros	05/06/2017	Le	4800 litros	05/06/2017	Qd	100	05/06/2017	Qd	200	05/06/2017	Qd	
06/06/2017	Le	4800 litros	06/06/2017	Le	4800 litros	06/06/2017	Qd		06/06/2017	Qd		06/06/2017	Qd	200
07/06/2017	Le	4800 litros	07/06/2017	Le	4800 litros	07/06/2017	Qd	100	07/06/2017	Qd	200	07/06/2017	Qd	
08/06/2017	Le	4800 litros	08/06/2017	Le	4800 litros	08/06/2017	Qd		08/06/2017	Qd		08/06/2017	Qd	200
09/06/2017	Le	4800 litros	09/06/2017	Le	4800 litros	09/06/2017	Qd	100	09/06/2017	Qd	200	09/06/2017	Qd	
10/06/2017	Le	4800 litros	10/06/2017	Le	4800 litros	10/06/2017	Qd		10/06/2017	Qd		10/06/2017	Qd	200
11/06/2017	Le	4800 litros	11/06/2017	Le	4800 litros	11/06/2017	Qd	100	11/06/2017	Qd	200	11/06/2017	Qd	
12/06/2017	Le	4800 litros	12/06/2017	Le	4800 litros	12/06/2017	Qd		12/06/2017	Qd		12/06/2017	Qd	200
13/06/2017	Le	4800 litros	13/06/2017	Le	4800 litros	13/06/2017	Qd	100	13/06/2017	Qd	200	13/06/2017	Qd	
14/06/2017	Le	4800 litros	14/06/2017	Le	4800 litros	14/06/2017	Qd		14/06/2017	Qd		14/06/2017	Qd	200

Fig. 3. Plan sincronizado para una CS realizado por MASM-ERP

3.3 Escenario de interrupción

Definido el plan sincronizado, comienza su ejecución. El día 02/06/2017 el MASM-ERP recibe información de la empresa A sobre el estado de la $O^v = \{Le: 4800 \text{ litros, fecha: } 02/06/2017\}$. Si la O^v fue cumplimentada, el MASM-ERP continúa el monitoreo de las órdenes de las restantes empresas para el horizonte de planificación.

Dado que algunas vacas sufrieron una enfermedad que ocasionó que la producción de leche del tambo disminuyera de 4800 litros a 4500 litros. Entonces, el día 02/06/2017 el MASM-ERP evalúa si la $O^v = \{Le: 4800 \text{ litros, fecha: } 02/06/2017\}$ de la empresa A fue cumplimentada a partir de los datos recibidos de la empresa A. Dado que la empresa A no podrá cumplir con dicha orden notifica la cantidad efectiva de leche que podrá enviar a la empresa B, $O^v = \{Le: 4500 \text{ litros, fecha: } 02/06/2017\}$. En base a estos datos, MASM-ERP detecta un posible evento y envía un mensaje a la empresa B consultando si esta situación de escasez de envío de leche podría afectar a su producción de quesos duros. Para ello, la empresa B posee definida una holgura, que dependerá si puede cubrir el faltante con leche enviada por otro proveedor que no se encuentra integrado a la CS. Suponiendo que la holgura de la empresa B es de 200 litros de leche, se analiza: $4800 - 4500 = 300 > 200$ (holgura). Como la diferencia entre ambas cantidades es mayor que la holgura, se produce un evento disruptivo y MASM-ERP notifica a la empresa C la ocurrencia del mismo. La Fig.4 sintetiza los datos resultantes en la base de datos del MASM-ERP del plan sincronizado afectado por la interrupción.

Empresa A: Ov			Empresa B: Oc			Empresa B: Op			Empresa B: Ov			Empresa C: Oc		
Fecha	Materia Prima	Cantidad	Fecha	Materia Prima	Cantidad	Fecha	Producto	Cantidad (hormas)	Fecha	Producto	Cantidad (hormas)	Fecha	Producto	Cantidad (hormas)
01/06/2017	Le	4800 litros	01/06/2017	Le	4800 litros	01/06/2017	Qd	100	01/06/2017	Qd	200	01/06/2017	Qd	
02/06/2017	Le	4500 litros	02/06/2017	Le	4500 litros	02/06/2017	Qd		02/06/2017	Qd		02/06/2017	Qd	200
03/06/2017	Le	4800 litros	03/06/2017	Le	4800 litros	03/06/2017	Qd	100	03/06/2017	Qd	200	03/06/2017	Qd	
04/06/2017	Le	4800 litros	04/06/2017	Le	4800 litros	04/06/2017	Qd		04/06/2017	Qd		04/06/2017	Qd	200
05/06/2017	Le	4800 litros	05/06/2017	Le	4800 litros	05/06/2017	Qd	100	05/06/2017	Qd	200	05/06/2017	Qd	
06/06/2017	Le	4800 litros	06/06/2017	Le	4800 litros	06/06/2017	Qd		06/06/2017	Qd		06/06/2017	Qd	200

Fig. 4. Plan sincronizado afectado por una interrupción

4 Conclusiones

Este trabajo presentó una propuesta que permite sincronizar planes de abastecimiento, producción y distribución y anticipar/detectar eventos disruptivos de

las empresas involucradas en una CS utilizando sistemas ERP. Esta propuesta permite que las empresas que apoyan sus actividades diarias con estos sistemas no requieran su reemplazo o convivencia con nuevos sistemas de planificación para gestionar de manera integrada su CS.

Dada la independencia del algoritmo que realiza la sincronización de los planes del sistema ERP, las empresas solamente debieron respetar la carga de datos específicos (cantidad de producto, fecha de entrega, etc.) en los módulos de compra, manufactura, depósito y venta a efectos de poder realizar la extracción de datos necesarios. Durante la ejecución de los planes integrados, MASM-ERP fue capaz de acceder a la base de datos de los ERP para obtener datos de ejecución. Esto evitó recopilar datos en sistemas externos u otros utilizados por las empresas a efectos de analizar la posible ocurrencia de un evento disruptivo. MASM-ERP, como extensión del sistema MASM, fue capaz de gestionar eventos disruptivos independientemente de las características de los procesos involucrados en la CS y notificar a las empresas involucradas sobre su ocurrencia.

Como trabajo futuro, se pretende la integración completa con MASM y otros sistemas que permiten corregir los eventos disruptivos mediante métodos de optimización y replanificación, como así también agregar herramientas semánticas que permitan independizar la nomenclatura de productos, subproductos y materia primas utilizada en empresas que forman parte de CS internacionales.

Referencias

- [1] Fernández, E. Proceso de Monitoreo Colaborativo de Planes de Abastecimiento para anticipar Eventos Disruptivos en la Cadena de Suministro. Tesis Doctoral, ISBN 978-987-33-2613-4 (2012).
- [2] Davenport T. H. Misión crítica: Promesas y riesgos de los sistemas empresariales. Primera edición, Editorial Oxford University Press México (2002).
- [3] Gattikera T. F., Goodhueb D. L. Understanding the local level costs and benefits of ERP through organizational information processing theory, *Information & Management*, 41, pp. 431–443 (2004).
- [4] Davenport T. H. Misión crítica: Promesas y riesgos de los sistemas empresariales. Primera edición, Editorial Oxford University Press México (2002).
- [5] Ben Clegg Yi Wan. Managing enterprises and ERP systems: a contingency model for the enterprization of operations. *International Journal of Operations & Production Management*, 33, 1458 – 1489, DOI: <http://dx.doi.org/10.1108/IJOPM-07-2010-0201> (2010).
- [6] Yi-fen Su, Chyan Yang. Why are enterprise resource planning systems indispensable to supply chain management? *European Journal of Operational Research*, 203, 81–94 (2010).
- [7] Masing N. SC Event Management as Strategic Perspectiva – Market Study: CMDESC Software Performance in the European Market. Master Thesis. Université du Québec en Outaouasis (2003).
- [8] Software Agent Technologies, Eds: Marius Walliser, Stefan Brantschen, Monique Calisti and Thomas Hempfling (2006).
- [9] Guarnaschelli A., Fernández E., Chiotti O., Salomone E. A Service-Oriented Approach to Collaborative Management of Disruptive Events in Supply Chains. *International Journal of Innovative Computing, Information and Control (IJICIC)*, 8, 5341-5368, ISSN 1349-4198, (2012).

- [10] Liu R., Kumar A., Van Der Aalst W. A formal modeling approach for supply chain event management. *Decision Support Systems*, 43, 3, 761-778 (2007).
- [11] Montgomery N., Waheed R. Event Management Enables Companies to Take Control of Extended Supply Chains. *AMR Research* (2001).
- [12] Fernández E., Salomone E., Chiotti O, A model driven development approach based on a reference model for predicting disruptive events in a supply process, *Computers in Industry* 63, 482–499 (2012).
- [13] Fernández E., Toledo C. M, Galli M. R., Salomone E., Chiotti O. Agent-based Monitoring Service for Management of Disruptive Events in Supply Chains. *Computers in Industry*. 70, 89-101. DOI 10.1016/j.compind.2015.01.009 (2015)
- [14] Fernández E., Bogado V., Salomone E., Chiotti O. Framework from Modelling and Simulating the Supply Process Monitoring to Detect and Predict Disruptive Events. *Computers in Industry*, 80, 30-42 (2016)
- [15] Rittiner J., Gonzalez M., Fernandez E., Chiotti O. Modelo de negociación cooperativa del nivel de servicio de almacenes. Argentina. Sarmiento 440 - (C1041AAJ) - Ciudad Autónoma de Buenos Aires - Argentina. 2015. Libro. Artículo Completo. Congreso. VIIIº Congreso Argentino de Ingeniería Industrial. UTN-Facultad Regional Chubut (2015).
- [16] Fagiolani B., Fernandez E., Chiotti O. Modelo de negociación de holguras para reparar programas de abastecimiento afectados por eventos disruptivos. Argentina. Cesar Bustelo, et.al. VII Congreso Argentino de Ingeniería Industrial. UTN FRCH (2014).
- [17] Winkelmann A., Fleischer S., Herwig S., Becker J., A conceptual modeling approach for supply chain event management, in: *Proceedings of the 17th European Conference on Information Systems* (2009).
- [18] Bansal M., Adhitya A., Srinivasan R., Karimi I., An online decision support framework for managing abnormal supply chain events, in: L. Puigjaner, A. Espuña (Eds.), *European Symposium on Computer-Aided Process Engineering*, volume 20 of *Computer Aided Chemical Engineering*, Elsevier, 985–990 (2005).
- [19] Kurbel K., Schreiber D., Agent-based diagnostics in supply networks, *Issues in Information Systems* 8 (2007).
- [20] K. Kwangmyeong, C. Injun, Active data acquisition for proactive exception handling in manufacturing, *International Journal of Advanced Manufacturing Technology* 43 (2009) 365–378.
- [21] Zimmermann R., S. Käs, Butscher R., Bodendorf F., An ontology for agent-based monitoring of fulfillment processes, in: *Whitestein Series in Software Agent Technologies and Autonomic Computing*, 2005, 323–345.
- [22] Xu H.H., A web-based system for proactive management of supply exceptions, *Journal of Manufacturing Systems* 29 (2010) 91–101.
- [23] Ozteme E, Tekez E.K., A general framework of a reference model for intelligent integrated manufacturing systems (remims), *Engineering Applications of Artificial Intelligence* 22 (2009) 855–864, *Artificial Intelligence Techniques for Supply Chain Management*.
- [24] Boissier O., Bordini R.H., Hübner J.F., Ricci A., Santi A., Multi-agent oriented programming with JaCaMo, *Science of Computer Programming*, 78, 747–761 (2013).
- [25] Omicini A., Ricci A., Viroli M., Artifacts in the A&A meta-model for multi- agent systems, *Autonomous Agents and Multi-Agent Systems* 17, 432–456 (2008).
- [26] Hübner J., Bordini R., G Picard., Jason Using, MOISE+, and CArtAgO to develop a team of cowboys, in: K. Hindriks, A. Pokahr, S. Sardina (Eds.), *Programming Multi- Agent Systems*, volume 5442 of *LNCIS*, Springer, Berlin/Heidelberg, 238–242 (2009).
- [27] Weyns D., Omicini A., Odell J., Environment as a first-class abstraction in multiagent systems, *Autonomous Agents and Multi-Agent Systems* 14, 5–30 (2007).
- [28] Erl T., A Karmarkar., Walmsley P., Haas H., Yalcinalp U., Liu K., Orchard D., Tost A., Pasley J., *Web Service Contract Design and Versioning for SOA*, Prentice Hall, New York, (2009)

Modelo de Análisis para el Dominio de Tránsito Ciclista

Leo Ordinez¹, Damián Barry¹, Alex Torrico¹, Emanuel Mallon¹, José Luis Devia¹, Lucas Abella¹, Nahuel Alamo¹

¹LINVI, Departamento de Informática, Facultad de Ingeniería, UNPSJB, 9120
Puerto Madryn, Argentina,
leo.ordinez@gmail.com, damian_barry@unpata.edu.ar,
<http://madryn.unp.edu.ar>

Resumen En este trabajo se presenta un modelo del dominio de aplicación enmarcado en el análisis del tránsito ciclista en una ciudad mediana, como la de Puerto Madryn. El objetivo del modelo es construir conocimiento que permita determinar los trayectos más adecuados para emplazar en esa ciudad ciclovías. El núcleo de la propuesta es el análisis de información de *recorridos*, que permita sintetizar *trayectos* con sus respectivas longitudes, frecuencias, densidades y ponderaciones. Además del modelo teórico, se presenta una herramienta de análisis de datos capturados (recorridos realizados por los ciclistas) y un generador de recorridos sintéticos.

Keywords: big data analytics, datos espacio-temporales, ciclovías, urbanismo, sistemas colaborativos

1. Introducción

Actualmente, muchos gobiernos municipales no cuentan con información de calidad, ni mecanismos de adquisición continua de datos, que permitan un monitoreo permanente del estado de los sistemas que hacen al funcionamiento global de la ciudad. La propuesta aquí presentada apunta a estudiar, experimentar y optimizar una infraestructura de Big Data enfocada en información heterogénea fuertemente ligada a datos espaciales y temporales, de un dominio particular como es el tránsito ciclista.

En línea con muchas de las principales ciudades del país (y del mundo), la localidad de Puerto Madryn tiene en carpeta la instalación de ciclovías que garanticen tanto el orden del tráfico en la ciudad, como la seguridad de los ciclistas. Este proyecto también persigue otro objetivo implícito que es el promover el uso de la bicicleta como medio de transporte, mejorando la calidad de vida de las personas y aportando a la ciudad en aspectos tales como: el tráfico, la seguridad, el medio ambiente, el turismo, etc.

El propósito del presente trabajo es la determinación y visualización de las alternativas más convenientes para la instalación de ciclovías en la ciudad de

Puerto Madryn, considerando los lugares con mayor afluencia de personas (centralidades), las características de las calles y los recorridos actuales con mayor tránsito ciclista.

La propuesta incluye la generación artificial de datos. En particular, aquellos referidos a recorridos realizados por ciclistas en la ciudad, serán generados en forma sintética. Asimismo, se incluye el diseño de las estrategias de elección y ponderación de los trayectos más convenientes para la instalación de ciclovías. Por otro lado, se contempla la visualización de los trayectos más convenientes en un mapa y con una escala de ponderación clara.

Aunque la problemática no es nueva y sobre todo ha cobrado gran relevancia en los últimos años, en general los trabajos científicos se enfocan en la aplicación de técnicas de optimización matemática y se concentran en grandes ciudades con sistemas de transporte multimodal. En [1] se presenta una metodología para modelar simultáneamente los modos de transporte de automóviles, autobuses y bicicletas, considerando las interacciones entre los tres modos a través del modelado de la división modal y la asignación de red de los diferentes recorridos de cada modo. Posteriormente, este modelo se utiliza para optimizar el diseño de la red de caminos ciclistas que obtengan un sistema de transporte eficiente y sostenible. En [2], aunque el objetivo es el análisis de estaciones para un sistema de bicicletas compartidas, se expone un modelo detallado útil para el modelado de dominio de tránsito ciclista. Por otro lado, en [3] se presenta un mapa que ilustra, por primera vez, el flujo real de ciclistas casuales y mensajeros de bicicletas en la ciudad de Madrid. Este mapa fue obtenido a partir del desarrollo y los resultados de la iniciativa Madrid Cycle Track, una plataforma en línea lanzada con el objetivo de recoger rutas ciclistas y otra información de voluntarios. Esta iniciativa es complementaria al trabajo aquí propuesto y se está en fase de desarrollo para lanzarla en la ciudad de Puerto Madryn. En relación a esto, en [4] se realiza un análisis para determinar si los datos de crowdsourcing de aplicaciones deportivas pueden usarse para cuantificar y mapear la variación espacial y temporal del número de ciclistas. Por otro lado, diversos autores han desarrollado modelos para describir el comportamiento de los ciclistas, ante la escasez de datos de los mismos [5], [6], [7], [8].

En particular, el presente trabajo propone un modelo de dominio específico y una aplicación prototipo que implementa dicho modelo. El modelo considera diferentes fuentes de información (recorridos, centralidades y calles) de una ciudad, para realizar un análisis integrado de las mismas. El objetivo principal del análisis es la determinación de los segmentos de calles (trayectos) más convenientes para instalar ciclovías, lo cual implica responder a otras preguntas previamente como son los trayectos con mayor tránsito ciclista, los que tienen menor, entre otras.

El resto del trabajo se organiza de la siguiente manera: en la sección 2 se presenta el modelo de dominio para el tránsito ciclista; en la sección 3 se expone el núcleo de análisis, el cual es la ponderación de trayectos; en la sección 4 muestra cómo la abstracción planteada en las secciones anteriores es utilizada

en la construcción de conocimiento útil para la toma de decisiones; finalmente, la sección 5 expone las conclusiones y trabajos futuros.

2. Modelado del Dominio

Se parte de que la ciudad se encuentra representada por una malla de puntos p , donde cada punto es una intersección de dos calles (esquina). Aquellos cruces de más de dos calles, se representan considerando las combinaciones en forma de pares. La unión entre dos puntos consecutivos se denomina *cuadra* $b_{ab} = p_a \rightarrow p_b$. Nótese que de lo anterior se desprende que $b_{ab} \neq b_{ba}$, lo cual implica indicar que las *cuadras* tienen sentido de circulación.

Se supone que un *ciclista* c_i realiza n recorridos r_{ij} , con $j = 1 \dots n$. A la vez, cada recorrido r_{ij} se *normaliza*, para adecuarlo a la representación de puntos, y cuadras en la malla de la ciudad (*i.e.*, el recorrido más corto que podría realizar un ciclista es de una cuadra, el cual se representa por dos puntos consecutivos). El recorrido r_{ij} normalizado se denota R_{ij} . En esta misma línea, el conjunto¹ de recorridos \mathcal{R}_i que realiza un ciclista a lo largo de su vida, se expresa como

$$\mathcal{R}_i = \sum_{j=1}^n R_{ij}$$

De igual modo, se construye el conjunto universal de recorridos \mathbf{R} de todos los ciclistas

$$\mathbf{R} = \prod_{i=1}^n \mathcal{R}_i$$

Por otro lado, se supone que cada recorrido R_{ij} se compone de *tramos*, que son conjuntos ordenados de cuadras, siendo t_{ij}^k el k -ésimo *tramo* del j -ésimo *recorrido* del i -ésimo *ciclista*. De este modo, cada recorrido se puede expresar como

$$R_{ij} = \sum_{k=1}^n t_{ij}^k$$

Así, cada tramo es un conjunto ordenado de cuadras consecutivas

$$t_{ij}^k = \sum_{b=1}^m b_{ab}, \quad m \leq n$$

De lo anterior, surge que cada tramo tiene un sentido de circulación.

A partir de los tramos que componen los recorridos normalizados de los ciclistas, se construye la entidad *trayecto*, que se representa con τ , donde

$$\tau_{uv} = t_u \cap t_v, \quad \forall t \in \mathbf{R}$$

¹ En adelante se asume la siguiente convención para conjuntos \sum representa un conjunto ordenado temporalmente (secuencia); y \prod representa un conjunto que no necesariamente respeta un orden.

De la definición de trayecto surge que éstos son las intersecciones de dos tramos (*i.e.*, las cuadras compartidas por ambos). Los trayectos podrían tener un sentido de circulación, sin embargo para los fines de este trabajo no se considera esa característica.

2.1. Medidas de Análisis de Trayectos

A continuación se definen una serie de características de los trayectos, las cuales se utilizarán como base para la elaboración de los algoritmos de búsqueda y análisis de los recorridos, que permitan determinar los lugares más adecuados para instalar ciclovías.

Longitud: Se define la *longitud* de un trayecto τ_{uv} como $long(\tau_{uv}) = |\tau_{uv}| = |t_u \cap t_v|$.

Frecuencia: Se define la *frecuencia* de un trayecto τ_{uv} , simbolizada como $frec(\tau_{uv})$, como la cantidad veces que se ha realizado ese trayecto, independientemente de si lo ha realizado uno o varios ciclistas. Esto es, $frec(\tau_{uv}) = \omega$

Densidad: Se define la *densidad* de un trayecto τ_{uv} , simbolizada como $dens(\tau_{uv})$, como la cantidad de ciclistas diferentes que realizaron ese trayecto. Esto es, $dens(\tau_{uv}) = \gamma$

Similitud: Se dice que dos recorridos R_i y R_j son *similares* si existe un trayecto construido a partir de ambos, cuya longitud es mayor a un cierto porcentaje del menor. Esto es, $\exists \tau_{uv} : \tau_{uv} = R_i \cap R_j \wedge long(\tau_{uv}) \geq \mu \cdot \min\{|R_i|, |R_j|\} \wedge \mu \in (0, 1)$

3. Análisis de Trayectos

Con el objetivo de determinar los lugares más aptos para la instalación de ciclovías, se busca obtener el conjunto de trayectos más transitados. A continuación se describe el algoritmo desarrollado. El resultado de los trayectos más transitados es fundamental para cualquier análisis que se haga. No obstante, éste no es único ni determinante, ya que se deben considerar otros factores, como por ejemplo las características de las calles (pendiente, ancho, material, etc.).

Trayectos más frecuentes

Dado un conjunto de recorridos normalizados, ya sea obtenido desde la base de datos o de un conjunto preseleccionado, este algoritmo se encarga de realizar el análisis correspondiente para determinar los trayectos similares entre todo el conjunto y ponderar cada uno de los mismos, asignándole un peso mediante una fórmula determinada.

Factores de peso Para determinar el peso de cada trayecto, se tiene en cuenta la *frecuencia* y la *densidad*.

A partir de estos datos se determinan otros dos factores que influyen en el peso del trayecto (P_t), como lo son:

1 - *Factor de frecuencia del trayecto* ($F_{frec()}$): se calcula teniendo en cuenta:

- La frecuencia actual (F_{act}): la cantidad de veces que se realizó el trayecto
- La frecuencia mínima ($F_{mín}$): la frecuencia mínima del conjunto de recorridos analizado
- La frecuencia máxima ($F_{máx}$): la frecuencia máxima del conjunto de recorridos analizado
- El rango de frecuencia ($\Delta(F)$): la diferencia entre la frecuencia máxima y mínima ($F_{máx} - F_{mín}$)

$$F_{frec()} = \frac{(F_{act} - F_{mín})}{\Delta(F)}$$

con $F_{act} - F_{mín} > 0$, dado que los trayectos analizados son realizados por lo menos una vez.

2 - *Factor de densidad del trayecto* ($F_{dens()}$): se obtiene a partir de:

- La densidad del trayecto ($dens()$): el número de usuarios distintos que realizan el trayecto.
- La densidad total ($dens_t()$): la cantidad de usuarios distintos del conjunto de recorridos analizado.

$$F_{dens()} = \frac{dens()}{dens_t()}$$

con $dens_t() > 0$, dado que un trayecto pertenece como mínimo a un usuario.

3 - *Factores de Importancia*: A fin de ponderar el trayecto, se establecen dos factores que pesan las medidas anteriores, ya que éstos no tienen la misma importancia. Se denominan I_{frec} e I_{dens} , haciendo referencia al Factor de frecuencia $F_{frec()}$ y al Factor de densidad $F_{dens()}$, respectivamente. En general,

$$I_{frec} + I_{dens} = 1$$

Fórmula de ponderación De acuerdo a los factores descritos anteriormente, a continuación se presenta la descripción de la fórmula de ponderación utilizada:

$$P_t = (I_{frec} \cdot F_{frec()}) + (I_{dens} \cdot F_{dens()})$$

con $0,1 \geq P_t \geq 1$, donde $I_{frec} = 0,7$ e $I_{dens} = 0,3$, puesto que se considera que $F_{frec()}$ es más determinante que $F_{dens()}$.

4. Construcción de Conocimiento

La construcción de conocimiento está ligada a la interpretación de hechos de un dominio específico. Así, el análisis de los modelos presentados tiene un correlato en las metas de conocimiento esperadas. En este sentido, el dominio específico se orienta a facilitar la construcción de conocimiento buscando alcanzar las siguientes métras de negocio:

- Ordenamiento urbano, especialmente en lo que se refiere a la movilidad de los habitantes.
- Mejoramiento de la calidad de vida de la población, especialmente en la promoción de mecanismos de vida saludable, fomentando actividades que permitan interactuar con los planes de medicina preventiva.
- Mejorar mecanismos que permitan optimizar el presupuesto de infraestructura urbana relacionada con la movilidad.

A fin de alcanzar estas metas se plantean una serie no exhaustiva de interrogantes, los cuales se pretende responder a través del modelo propuesto. A continuación se enuncian algunos de ellos:

- ¿Cuáles son los trayectos mínimos que tienen mayor cantidad de tráfico ciclista?
- ¿Cuáles son los trayectos máximos que tienen mayor cantidad de tráfico ciclista?
- ¿Cuáles son los trayectos mínimos que tienen menor cantidad de tráfico ciclista?
- ¿Cuáles son los trayectos mínimos que tienen menor cantidad de tráfico ciclista?
- ¿En qué mes del año se producen los recorridos de los trayectos mínimos y máximos?
- ¿En qué días de la semana se realizan los mayores desplazamientos?
- ¿En qué franjas horarias se realizan los mayores desplazamientos?
- ¿Qué zonas de la ciudad (barrios) son los menos transitados por los ciclistas?
- ¿Qué centralidades se encuentran cercanas a los trayectos con mayor cantidad de tráfico?
- ¿Cuáles son los radios mínimos y máximos entre las centralidades halladas y los trayectos de mayor tráfico?

En base al modelo propuesto, se desarrolló un prototipo de aplicación de simulación y análisis de recorridos ciclistas. El simulador genera conjuntos sintéticos de recorridos y el analizador aplica el modelo planteado anteriormente, a fin de que los tomadores de decisión puedan contar con la información suficiente para determinar los lugares por los cuales instalar las ciclovías.

4.1. Arquitectura

En la Figura 1 se muestra la arquitectura completa de la aplicación desarrollada. El diseño se basó en capas. En particular, se dividió la arquitectura

en cuatro capas. Esto se debió a cuestiones de diseño y a imposiciones de la tecnología utilizada en algunas capas, como Laravel.

1. Capa de Presentación
2. Capa de Servicio
3. Capa de Datos
4. Capa de Base de Datos.

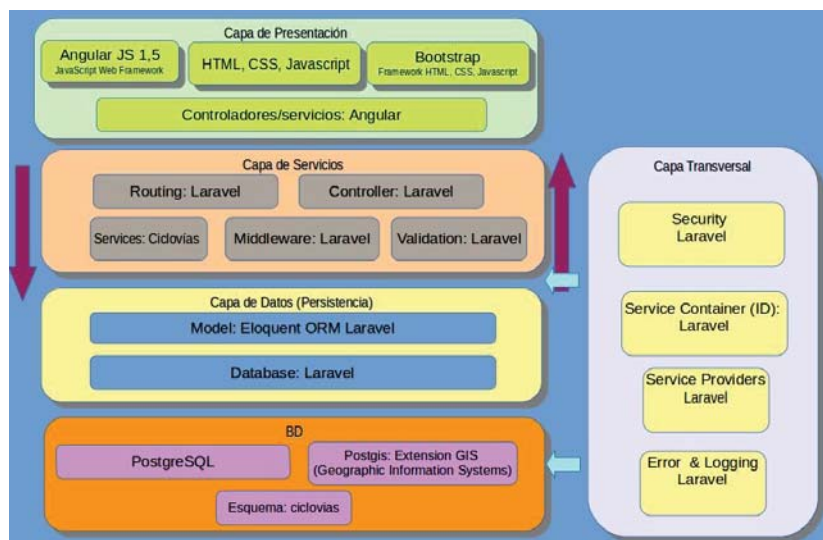


Figura 1. Arquitectura completa de la aplicación.

En la Figura 2 se muestra el modelo de datos de la aplicación, con las principales entidades involucradas.

- **Centralities** (Centralidades): muestra los datos referidos a Centralidades, que son puntos de interés para los ciclistas. En ellos se destacan establecimientos educativos, farmacias, etc. Los datos que se representan, son el nombre, dirección, ubicación geográfica.
- **Zones** (Zonas): muestra los datos referidos a las Zonas en las que está delimitada la ciudad. Los datos que se representan, son el nombre, dirección, puntos que delimitan la extensión de una zona.
- **Roads** (Calles): muestra los datos referidos a las calles de la ciudad. Los datos que se representan, son el nombre y los puntos de todas las esquinas referidas a esa calle.
- **Trips** (Recorridos): muestra los datos referidos a los Recorridos realizados por los ciclistas en la ciudad. Los datos que se representan, son el nombre, descripción, usuario, tiempo del recorrido, distancia, el recorrido que realiza (Linestring), y una referencia al datalog.

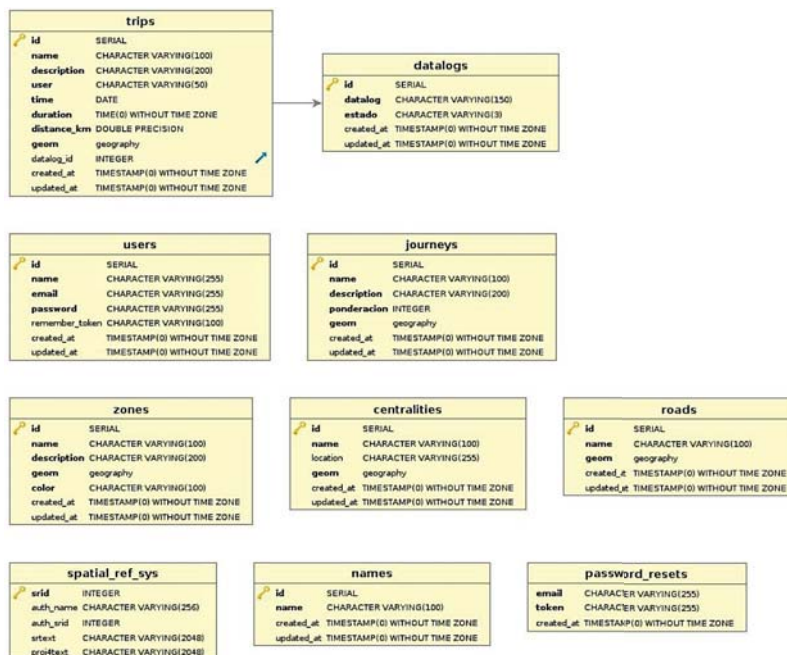


Figura 2. Modelo de Datos.

- **Journeys** (Trayectos): representación de los trayectos en línea recta que unidos forman un Recorrido. Los datos que se representan son el nombre, descripción, usuario, ponderación, la línea del trayecto realizado (Linestring).
- **Names** (Nombres): representación de ciclistas ficticios que realizan Recorrido generados en la ciudad. Representa el nombre.

4.2. Funcionalidades

La aplicación cuenta con características que permiten visualizar los comportamientos presentados en el modelo de análisis, presentado en las secciones 2 y 3. El mismo cuenta con segmentos para:

- Visualizar y analizar recorridos.
- Visualizar y analizar trayectos.
- Visualizar y analizar densidades y ponderaciones:
 - Trayectos similares.
 - Recorridos por Zona.
 - Recorridos cercanos a un punto o centralidad (Figura 3).
 - Recorridos por rangos de distancia (Figura 4).
- Visualizar y analizar recorridos más frecuentes.
- Visualizar y analizar comportamiento de movilidad respecto de centralidades.

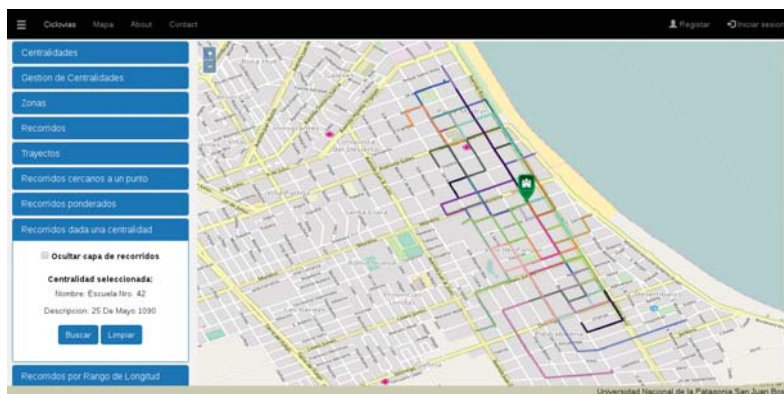


Figura 3. Recorridos dada una centralidad

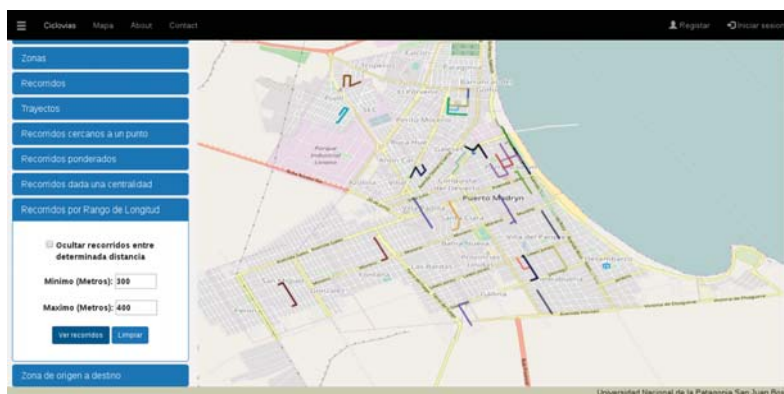


Figura 4. Recorridos por rango de longitud

- Visualizar y analizar las densidades poblacionales y los compromientos según la zona de residencia (Figura 5).

5. Conclusiones y Trabajos Futuros

En este trabajo se propuso un modelo para el análisis de un dominio específico, como es el tránsito ciclista en una ciudad pequeña como Puerto Madryn. A la vez, se expuso una aplicación prototipo que implementa dicho modelo. El modelo considera diferentes fuentes de información (recorridos, centralidades y calles) de una ciudad. Al mismo tiempo, en base a los datos obtenidos de esas fuentes construye conceptos nuevos, a fin de determinar los segmentos de calles (trayectos) más convenientes para instalar ciclovías. El principal concepto introducido es el de trayecto, el cual posee características como longitud, frecuencia, densidad y peso.



Figura 5. Zona de origen-destino

En cuanto a trabajos futuros, se continuará profundizando el modelo descrito, a partir de la utilización de grafos en la descripción de algunos aspectos y se avanzará en la construcción de una aplicación para la recolección de recorridos reales. Estos aspectos se utilizarán a la vez, para continuar mejorando la aplicación desarrollada.

Referencias

1. Gaspar, I., Benavente, J., Bordagaray, M., Alonso, B., Moura, J.L., Ibeas, A.: A bilevel mathematical programming model to optimize the design of cycle paths. In: *Transportation Research Procedia*. (2015)
2. Kim, E.S.: Mathematical model for cost-efficient installation of public transportation system. PhD thesis, University of California at Berkeley (2014)
3. Romanillos, G., Zaltz Austwick, M.: Madrid cycle track: visualizing the cyclable city. *Journal of Maps* (10 2015) 1–9
4. Jestico, B., Nelson, T., Winters, M.: Mapping ridership using crowdsourced cycling data. *Journal of Transport Geography* **52** (4 2016) 90–97
5. Gosse, C., Clarens, A.: Estimating Spatially and Temporally Continuous Bicycle Volumes by Using Sparse Data. *Transportation Research Record: Journal of the Transportation Research Board* **2443** (12 2014) 115–122
6. Broach, J., Dill, J., Gliebe, J.: Where do cyclists ride? A route choice model developed with revealed preference GPS data. *Transportation Research Part A: Policy and Practice* **46**(10) (12 2012) 1730–1740
7. Casello, J., Usyukov, V.: Modeling Cyclists' Route Choice Based on GPS Data. *Transportation Research Record: Journal of the Transportation Research Board* **2430** (12 2014) 155–161
8. El Esawey, M.: Estimation of Annual Average Daily Bicycle Traffic with Adjustment Factors. *Transportation Research Record: Journal of the Transportation Research Board* **2443** (12 2014) 106–114

Monitoring System for Elderly Care with Smartwatch and Smartphone

Pablo Rosales^{1,2}, Alejandro Vega¹, Carlos De Marziani^{1,2}, José Gallardo¹, Jorge Pires^{1,2}, Rómulo Alcoleas¹

¹ Universidad Nacional de la Patagonia San Juan Bosco (UNPSJB).
Comodoro Rivadavia, Argentina

² Centro de Investigaciones y Transferencia Golfo San Jorge
Consejo Nacional de Investigaciones Científicas y Técnicas (CIT-GSJ CONICET)
{prosales, alejandrovega}@unpata.edu.ar

Abstract. In recent years, mobile devices, such as Smartphones and Smartwatches, have increasingly been used as personal healthcare gadgets. This work presents the development of a monitoring system based on a software application over Android OS for mobile devices. The main goal is to use it as a tool for monitoring and tracking elderly people. The proposed application allows to: detect falls to the ground of the users; monitor the heart rate; count the number of steps that the users perform during the day and also track their position on a map. Despite the presence of several cases of possible false positives (e.g. answer a call or sending a message could trigger a wrong detection of fall or increase the steps counters.), the results obtained show that it is possible, by using current mobile devices in the market, to supervise and track the user's activities.

Keywords: elderly care, mobile applications; wearable; Android.

1 Introduction

Currently, there are more than two billion Smartphones (SmP) around the world and it is expected that, by 2019, 36% of the population will have one [1]. This exponential growth, added to the higher performance computing, sensing and communication capabilities that current devices present [2], have aroused the interest of different research groups and companies. Current research efforts are focused on the development of functionalities that could assist in the daily life of persons in order to monitor or provide data about: 1) User falls [3]; 2) Heart rate [4]; 3) Number of steps [5] and 4) Location of users [6] among other parameter useful in healthcare.

Cardiovascular diseases (CVD) are the leading cause of death in the worldwide, in 2012 it is estimated that 17.5 million people have suffered some type of CVD. Real time supervision of users' heart rate could detect irregularities [7] in the normal cardiovascular system functioning, that may help to avoid these types of accidents more frequently.

Falls are the second world's leading cause of death from accidental or unintentional injuries, it has been estimated that 424,000 people in the world die every year from this

cause. People over sixty-five years-old suffer most fatalities. There are 37.3 million non-fatal falls per year that need medical attention [8]. For these reasons, in this age-group, the supervision and reporting of falls is an important improvement of the quality of life.

It has been estimated that 24 million people in the world are suffering from Alzheimer's disease and that by the year 2040, this number will reach 81 million. The most affected regions are those that are more densely populated [9]. People who suffer from this disease are prone to disorientation, so it may be desirable to track the movements made by this type of users through their geolocation.

Currently six out of ten people do not perform adequate physical activity according to the World Health Organization (WHO). This can lead to diseases such as obesity and diabetes. According to WHO, a sedentary lifestyle causes 2 million deaths per year. Children, adults and the elderly do not perform enough exercise and instead occupy their time with sedentary activities related to video games, cell phones, television and computer use [10]. For these reasons, providing information about the number of steps and the distance traveled would help users to know how many calories they burn per day, in this way users would be encouraged to overcome daily goals imposed by themselves in order to maintain a healthy state.

This work presents a monitoring system in order to provide information about the described variables of interest, based on the use of Smartphone and Smartwatch mobile devices, in order to be used, for example, in residential elderly. The paper is organized as follows: Section 2 describes the system architecture and the functionalities to be implemented. Section 3 presents a description of the algorithms used. Then, in section 4, the obtained experimental results are shown, and finally, in section 5 the conclusions of the work are given.

2 Monitoring and tracking system

The global architecture of the proposed system is shown in Fig. 1. The SmP device reports the data collected to a central server. Smartwatches (SmW) are used to provide data that is transmitted to the SmP as shown in Fig. 2 such that it is processed and sent to the server. The SmP not only sends information by this mean, in special cases, like in the fall detection, it sends a report by a text message, with the location and time at which the event occurred. The connection between the SmP and SmW is established through Bluetooth. Depending on the functionality, the SmW sends the data on demand to the SmP, as it is done in the Heart Rate report, but for example, the Step Counter functionality reports whenever the SmW needs it and not when the phone requires it.

2.1 Selection of mobile devices

According to the specifications of parameters to be supervised, mobile devices must have the following sensors: accelerometer and pedometer on both devices and at least GPS on the SmP. Also, a heart rate sensor must be available in the SmW. In order to develop the software application, Android Operating System has been selected, since it is the most used in the world with 82.8% for the second quarter of 2015 [11]. This OS

offers many resources to develop software, and most of the applications available in the literature for the detection of falls were developed on this platform [12]. In the current market, different models of SmP are available. In this case, due to their availability, the following devices have been used:

1. Samsung Galaxy S4
 - a. OS: Android Lollipop
 - b. Processor Exynos 5 Octa-core 1.6Ghz
2. Samsung Galaxy S6
 - a. OS: Android Marshmallow
 - b. Processor: Exynos 7420 CPU Cortex A53 cores 4 1.50 GHz GPU Mali T760 CPU Cortex A57 4 cores 2.10 Ghz
3. Samsung Galaxy S7
 - a. OS: Android Nougat
 - b. Processor: Exynos 8 Octa 8890 8 cores 2,3 GHz

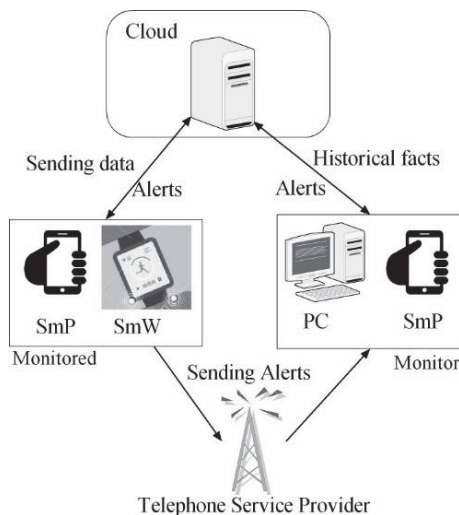


Fig. 1 Architecture of monitoring system for supervising elderly people.

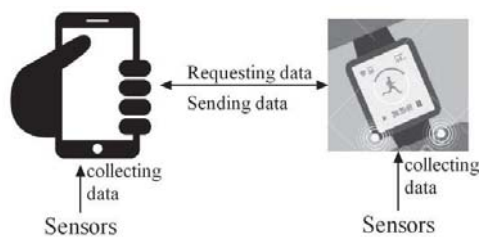


Fig. 2. Communication and gathering of information among mobile devices.

3 Algorithms for measuring parameters of interest

3.1 Pedometer

The pedometer functionality has been designed to detect the steps that the user performs in their daily life. Thus, taking into account the weight and the distance traveled, an approximation of the burned calories is calculated. To perform this function both devices, SmW and SmP, report their data. These devices have the sensor called "Step counter". This sensor provides step information and the period of time in which those occurred (time intervals). According to this, the highest step value computed between both devices is selected and reported to the server. In addition to the steps count, the distance traveled and the burned calories are computed.

3.2 Heart beat

The heart rate measurements are automatically performed every hour or when the user sends a request using the graphical interface. It also displays the historical data of the day and the evolution of the current measurement. The heart rate sensor is located in the SmW, but the control and storage of the data occur on the Smartphone, each device has a service that performs the actions outside of the main thread, keeping constant communication, ensuring the sending of requests and reception of data.

3.3 Georeferencing

This function has been designed to detect the user's location every 100 meters and/or at intervals of twenty minutes. In this way, with the data provided by the GPS, markers are deployed on a map, allowing drawing a path between them and leaving a record of the displacement activity. In this case, the mobile network can also be used to detect the location, but since it was decided to use a fine coarse detection, only the GPS sensor is used. Once the data is obtained, it is saved in the SmP, and the algorithm establishes a connection with the central server to send the information. At that moment, all previously unsent data is transferred, and already sent data is deleted from the local storage, preventing it from filling up and keeping the memory usage by the application at a minimum.

3.4 Fall detection

A fall is an event that can be divided into 5 stages [13] (see Fig. 3) in the following way:

1. *Pre-fall*: Activity of daily life performed by the person.
2. *Free fall*: The free fall of the subject to the ground caused by the loss of balance.
3. *Impact*: When the subject hits the ground.

4. *Post impact*: When the person is on the ground and remains inactive by the impact.

5. *Recovery*: When the subject stands or tries to do so.

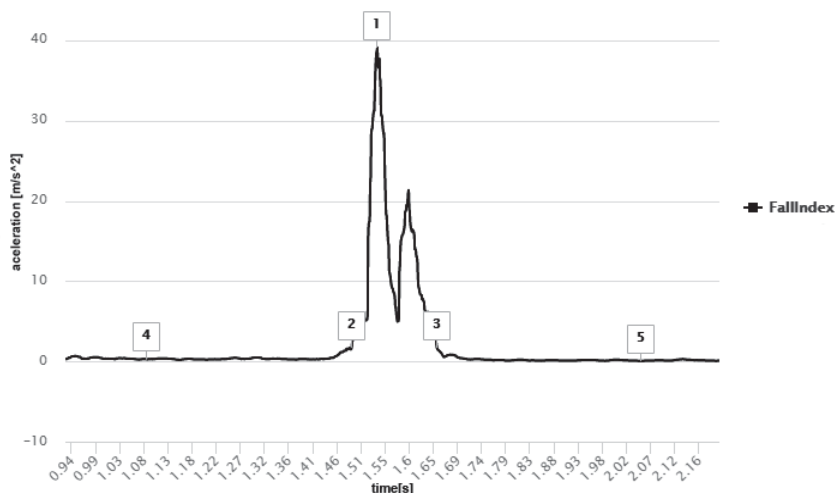


Fig. 3. Acceleration functions of a fall obtained by (1) and his characterization.

Figure 3 shows the fall index expressed in m/s^2 as a function of time in seconds. The falling index is obtained according to (1). As it depends on adjacent data, it is necessary to have a high sampling rate for efficient results.

$$FI_i = \sqrt{\sum_{k=x,y,z} \sum_{i-j}^i ((A_k)_i - (A_k)_{i-1})^2} \quad (1)$$

where A_k is the acceleration in the axis k , $k \in \{x,y,z\}$, i is the index of samples $k \in \{1, 2, \dots, j\}$ and j is the size of the window to be used to analyze de data.

The fall detection algorithm is performed in real time and is divided into the following three phases: 1) Smartphone Acceleration Threshold; 2) Smartphone Pattern Recognition; 3) Smartwatch Threshold and Pattern Recognition.

Smartphone Acceleration Threshold (fixed threshold). This is the first stage because it has the lowest computational cost. It consists of calculating the fall index taking a window of 20 measurements. The highest value in this window that also exceeds $24 m/s^2$ is selected as a candidate and the next step is carried out.

Smartwatch Threshold and Pattern Recognition. The last step is to check if the SmW also detected a fall. The procedure on this device is the same, only that the size of the window is 5 measurements for the calculation of the fall index, since the sampling rate is much lower than the one obtained in the SmP. When a fall detection takes place

on both devices, a message is sent to the designated emergency contact, along with the heartbeat and the recorded time&location data of the fall event. Additionally, all this information is sent to the server. Table 1 summarizes the values of the parameters for the detection of limits exceedance and the recognition of the acceleration pattern in both devices.

Table 1. Threshold values of the characterization parameters of a fall.

Value	Smartphone	Smartwatch
Peak of acceleration	24 m/s ²	2.10 m/s ²
Distance between point 2 and 3	1.95 Milliseconds	4.63 Milliseconds
Average of pre-fall	1.8 m/s ²	0.183 m/s ²
Average Post impact	0.9 m/s ²	0.16 m/s ²

The values in the table above were obtained from the analysis of the tests discriminated by type of fall (see section 4.3).

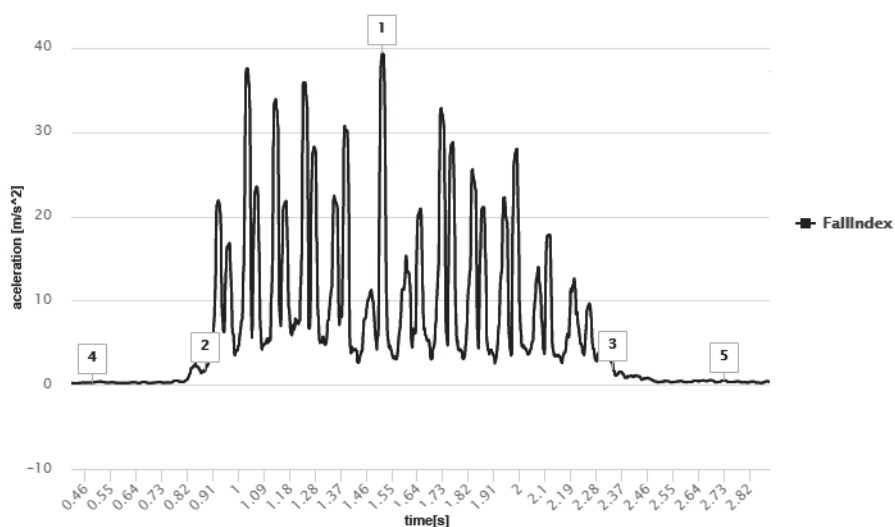


Fig. 4. Acceleration function of going down the stairs and his characterization.

Figure 4 shows the fall index of going down the stairs. Although it complies with the first part of the algorithm (point 1, the peak of acceleration greater than $24m/s^2$) the distance between point 2 and 3 is not correct, thus it does not meet the pattern and is not detected as a fall.

4 Performance evaluation of the system

4.1 Step Counter

Tests were made using our developed application, and contrasted against Google Fit.

For Google Fit the following tests were carried out:

1. The first test consisted of walking alone with the SmP resulting in a value of 1604 steps. Then it was synchronized with the SmW, which had not registered steps, and obtained a final value of 1604 steps.
2. The second test consisted of walking alone with the SmW resulting in a value of 2064 steps. Then it was synchronized with the SmP that had recorded steps from the first test and got a final value of 1604 steps. With the obtaining of this result, it was verified that the SmP disregards the results obtained by the SmW.

The same tests were carried out in our software, and the following results were obtained:

Table 2. Step count on each device for each test

Type	Smartphone	Smartwatch	Smartphone/ Smartwatch
Test 1	0	1150	1150
Test 2	1500	1150	1500

As shown, the algorithm always keeps the greatest number of detected steps, regardless of their origin.

4.2 Location detection

It was decided to make a journey of no more than 20 km of distance in which the application was marking the route made. As can be seen in Fig. 5 the path was delimited with the markers and red lines to obtain an estimate of the user's travel path. The application has obtained 6 different positions in which the device had a good signal quality to perform the marking of the place. After carrying out several tests of this type, it was concluded that the device was taking the information correctly.

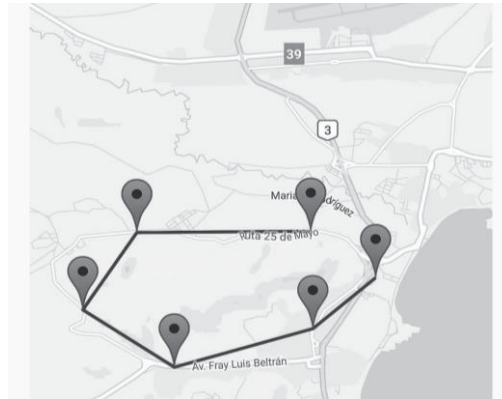


Fig. 5. Location detection interface.

4.3 Fall detection tests

In order to define the values used in the detection algorithms, a system was developed. This system records the acceleration in the devices and then sends them to a server for analysis. For that, 270 tests were performed, each one had associated the type of fall and the values 1 to 5 necessary for the detection algorithms (see 3.4), then the ranges of each parameter, according to the type of test, were analyzed resulting in those mentioned in table 1, these gave us the best percentage of detection for the tests performed.

Table 3 shows the values of fall detections according to the type of test and device that gives us the following result in percentages:

- Smartphone:
 - 96% of fall detections.
 - 0.47% false positives.
- Smartwatch:
 - 91,6% fall detections.
 - 3.8% false positives.

Those numbers must be interpreted as number of detection trials, e.g. in the test “falling forward” 29 of 30 trials were successfully detected as a fall in the SmP.

Table 3. Fall detection chart by type of test

Type	Smartphone	Smartwatch
walking 10 mts	0/30	0/30
Turning around	0/30	0/30
Tying shoes	0/30	0/30
Running	0/30	0/30
Stairs	0/30	0/30
Sit down slowly	0/30	1/30
Sit down quickly	1/30	7/30
Falling forward	29/30	26/30
Falling backward	29/30	29/30

5 Conclusions and future work

Mobile devices are widely used in daily human life and could be used to supervise people by means of GPS, pedometer, accelerometer, among other useful sensors. They also have the ability to send information by different communications ways such as text messages and calls over mobile data or Wi-Fi networks. In terms of detection of steps, mobile devices are very useful, although the sensors could be improved. Currently, the mere fact of moving the phone with the hand could detect false steps. This could be avoided with the use of other sensors working together, for example, GPS. Regarding the heart rate detection functionality, the sensors included in the SmW Moto 360 that we used for the tests work satisfactorily. A blood oxygen level detector could be added to further improve functionality. Regarding the detection of falls, the tests carried out show that by using the data from both a SmP and a SmW, the fall detection algorithm reduces the probability of a false positive event to a minimum. However, being a multipurpose gadget, sudden actions like taking the phone out of pocket or leaving it on a table might trigger an unwanted fall detection. This problem is further accentuated in the SmW since the wrist is not the ideal position to represent the stability of the person, as it is a device that is located in the torso or hip. That is why the ideal target group is the elderly or persons with reduced mobility, not only because of their health monitoring needs but also because of current device limitations.

Acknowledgments. This work was supported in part by National University of Patagonia San Juan Bosco (UNPSJB) (PI 1165 and PI 1189) and also by Engineering Faculty of UNPSJB.

References

1. Number of smartphone users worldwide from 2014 to 2020, <https://www.statista.com/aboutus/>
2. X. Su, H. Tong and P. Ji, "Activity recognition with smartphone sensors," in *Tsinghua Science and Technology*, vol. 19, no. 3, pp. 235-249, (2014).
3. Abbate, S., Avvenuti, M., Bonatesta, F., Cola, G., Corsini, P., & Vecchio, A. (2012). A smartphone-based fall detection system. *Pervasive and Mobile Computing*, 8(6), 883-899.
4. Tapia, E. M., Intille, S. S., Haskell, W., Larson, K., Wright, J., King, A., & Friedman, R. (2007, October). Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart rate monitor. In *Wearable Computers, 2007 11th IEEE International Symposium on* (pp. 37-40). IEEE.
5. Butte, N. F., Ekelund, U., & Westerterp, K. R. (2012). Assessing physical activity using wearable monitors: measures of physical activity. *Medicine & Science in Sports & Exercise*, 44(1S), S5-S12.
6. Chon, John, and Hojung Cha. "Lifemap: A smartphone-based context provider for location-based services." *IEEE Pervasive Computing* 10.2 (2011): 58-67.
7. OMS Enfermedades cardiovasculares, <http://www.who.int/mediacentre/factsheets/fs317/es/>
8. OMS | Caidas, <http://www.who.int/mediacentre/factsheets/fs344/es/>
9. El Alzheimer en el mundo, <http://www.alzheimers-support.com/es/alzheimers.html>
10. OMS, <http://expansion.mx/salud/2012/04/07/seis-de-cada-10-personas-no-tienen-suficiente-actividad-fisica-oms>
11. IDC: Smartphone OS market share 2016, 2015, <http://www.idc.com/promo/smartphone-market-share/os>
12. MDPI AG, http://www.mdpi.com/1424-8220/15/8/17827/htm*/
13. N. Noury, P. Rumeau, a. K. Bourke, G. O'Laighin, and J. E. Lundy: A proposal for the classification and evaluation of fall detectors, *Irbm*, vol. 29, no. 6, pp. 340-349 (2008).

Multimodal biometric recording architecture for the exploitation of applications in the context of affective computing

Enrique P. Calot and Jorge S. Ierache

Laboratorio de Sistemas de Información Avanzados,
Facultad de Ingeniería, Universidad de Buenos Aires,
Buenos Aires, Argentina

`ecalot@lsia.fi.uba.ar`, `jierache@yahoo.com.ar`

Abstract. In affective computing, it is important to design techniques that allow devices to acquire emotional states. To create and test these techniques it is necessary to have datasets that have several modalities namely, keystroke dynamics, electroencephalography, facial expressions, voice tone, heart rate, among others. This article presents a multimodal dataset that allowed us to detect the subjectivity that subsists in certain modalities —as are the surveys— and that is often overlooked, against objective modalities such as keystroke dynamics and electroencephalography. This article presents the creation of an environment in order to acquire a multimodal dataset. Work has also been done on the analysis of brain waves and their correspondence with other modalities.

Keywords: biosignal devices, affective computing, multimodal acquisition

1 Introduction

Affective computing is the study and development of systems and devices that can recognize, interpret, process and stimulate human emotions. Rosalind Picard defines that affective computing is “information technology that relates to emotions, not only those considered more important, such as joy or sadness, but also interest, boredom or frustration, which are those that are given in relation to computers” [28]. The emotional education of an individual in a multimodal context allows the simultaneous registration of different biometric information that allows deducing the emotional state of the individual. Affective computing and the development of applications in this area require emotion acquisition and learning. Future systems will interact with humans empathically with a diverse repertory in multimodal biometric contexts as well as the stimulation techniques.

Our goal is to create an empathic machine proficient in interpreting the human emotions and acting in consequence [29]. Controlling emotions of users shall increase productivity, minimize human errors and even prevent major disasters.

In order to study correlations between different acquisition modalities we created a dataset with 12 voluntary participants and recorded audio, video, heart

rate, keystroke dynamics and brain activity. To record different emotions we generate different emotions using visual and sonorous stimuli. Knowing how each modality corresponds to another will help predict emotions with only a keyboard or built-in microphone avoiding intrusive devices such as electroencephalographers (EEG).

2 Emotion modelization

In psychology, to model emotional states, there are different approaches, namely a dimensional and evaluation-based approach[11] and a categorical one. The former states that there is a small number of basic emotions, strongly connected in our brain and universally recognized[5]. This theory about the universality and interpretation of non-verbal affective expressions in terms of basic categories has been the most commonly adopted approach in research on automatic measurements of human affectivity. However, some researchers have shown that in everyday interactions, people have non-basic, subtle and quite complex affective states such as thinking, shaming or depression. Such states can be expressed through dozens of facial and body expressions, audio or anatomically possible physiological cues. Therefore, a single denomination (or any small number of discrete classes) may not reflect the complexity of the affective state conveyed by those rich sources of information[30]. For this reason, a group of researchers advocate the use of the dimensional description of human affectivity, where affective states are not independent of each other.

They are related to each other in a systematic way[8,30,21,32]. The most widely used dimensional model is a circular configuration called *Circumplex of Affectivity* developed by [30] (see Fig. 1(a)). This model is centered on the hypothesis that each basic emotion represents a bipolar entity that is part of the same emotional continuum. The proposed axes or poles are arousal (relaxed *vs.* excited) and valence (pleasure *vs.* dislike).

In this work we are using the dimensional approach employing the circumplex.

3 Stimulation

Each session took approximately one hour, although the time depended on the participant. The first step was to bring the subjects to a similar initial state considered neutral and free of any stimulus. This step consisted of a half minute on average of data entry, one minute relaxing with silence, 4 minutes and a half with a game and 3 minutes to obtain responses from his state (considered neutral or initial). The game consisted of using the mouse to grab certain colored dots on the screen while being played a relaxing background music. Then the four phases started in the order: happiness, calm, sadness, and Anger lasting approximately 13 minutes each. The experiment sweeps the four quadrants of the arousal–valence model[2] (see Fig. 1(b)) in a clockwise path moving over each phase in which the same stimulatory sequence is applied. They combined five

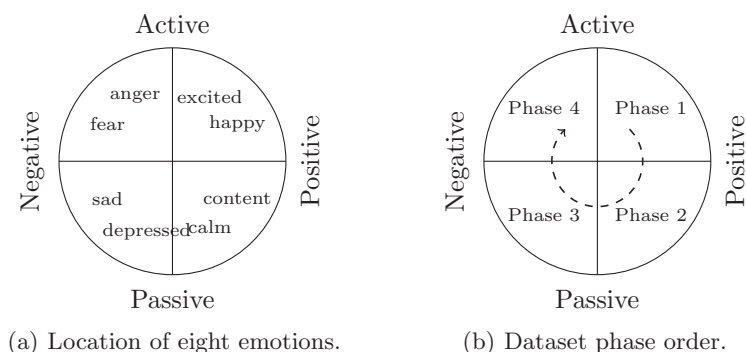


Fig. 1: Arousal-Valence Model.

minutes of visual stimulation (images), five minutes of audiovisual (videos and audio) and around three other for the responses to the current state (self-report, voice, and typing).

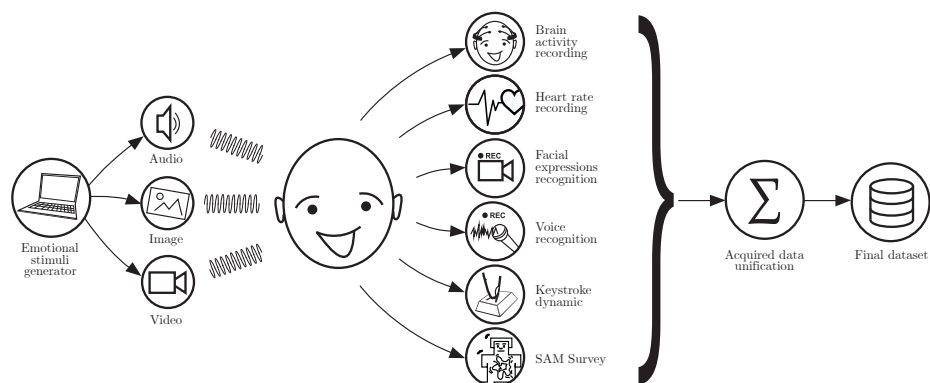


Fig. 2: Multimodal acquisition process.

To perform the visual stimulation, we requested a set of data compiled by [18] named IAPS. The set contains an image database used in various scientific studies related to emotions and is considered a de facto standard. We have also downloaded four videos with a high emotional content. Firstly, a subset of images taken from IAPS (100 images, 25 per phase) is shown, and then a video or audio, whose theme is related to the transient affective phase. For each phase, upon completion of image, video and audio stimulation, the subject is asked to complete a self-report using the Self-Assessment Manikin (SAM) survey proposed by Lang[17] (see Sect. 4.6). Then, we proceed to capture voice and keystroke dynamics. See Fig. 2.

4 Data acquisition and modalities

Several input interfaces had been used to create this dataset. The following sections summarize the main features of each modality.

4.1 Encephalography (EEG)

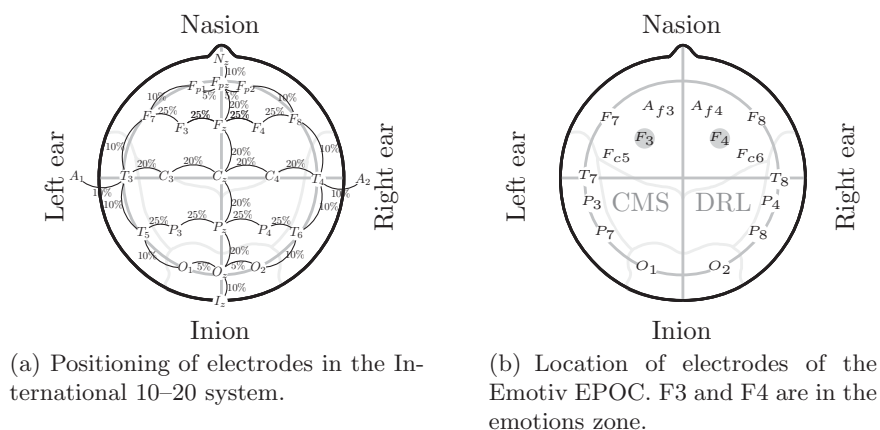


Fig. 3: Location of EEG electrodes.

A Brain-Machine Interface (BMI) facilitates communication between mental or cognitive functions created from a person's brain, by capturing electrical signals, to be processed, classified and communicated with specific applications or devices[15,13,14]. This dataset used the Emotiv EPOC headset for the acquisition of encephalographic signals by means of 14 electrodes, which were initially considered those located in the prefrontal cortex (F3, F4, AF3, AF4, according to the International 10-20 system[16], see Fig. 5). OpenViBE was employed to receive and store the raw samples and then apply filters to the electrodes and perform post-processing. This can be used to calculate the position of the emotional state in the bidimensional arousal-valence model[6,25].

4.2 Heart rate

The heart rate was acquired through a wristband placed on the participant (Xiaomi Mi Band 1s), which records the heart rate continuously over time. It is a polycarbonate band with an optical heart rate sensor that weights 14.5g.

4.3 Video

Video capture is performed continuously at all times. To analyze emotional gestures it is possible to use the tool `clmtrackr`[26], which detects faces and then

interprets emotions utilizing a training from a database of mapped points. The implementation uses constrained local models fitted by regularized landmark mean-shift developed by Audun Mathias Øygaard and based on [31]. To detect the emotional states, the MUCT database[22] was adopted, supplementing some extra images of Øygaard. The emotional model provides it with 18 coefficients[3]. The video can also be analyzed with any other tool.

4.4 Audio

All audio tracks are preserved in stereo for further analysis. The participant is required to read a text that is a different shuffle of the corresponding sentences employed for keystroke dynamics.

4.5 Keystroke dynamics

Keystroke dynamics are acquired after the stimulation, requesting the participant to enter a list of sentences of a neutral emotional load. Those passages had been randomly shuffled through the four phases. Both, the flight time and the hold time, are stored in order to reconstruct all keyboard events. In this regard, important advances were made in the laboratory, with a new method for classifying patterns and a replication of one of the most commonly used method[7]. A framework in Python has been coded in order to provide functions to analyze keystroke data. It includes a library compatible with standalone Python and Jupyter notebooks with a Python kernel.

4.6 Self-report (SAM)

A last-minute survey is conducted asking the participant about his current emotional state. It makes use of the widely adopted methodology to capture emotions named Self-Assessment Manikin (SAM) proposed by Lang[17]. As described by Bradley and Lang[2], SAM is an easy-to-administer non-verbal method for quickly assessing the pleasure, arousal, and dominance associated with a person's emotional reaction to an event.

It consists of 5 drawings that characterize if the person is happy or sad (valence) and how excited (arousal). For each domain, there are 9 points, one per figure and an intermediate point between them.

The Fig. 4 presents the paper and pencil version of SAM, illustrating its non-verbal representation, with figures of various points along each of the two dimensions. SAM extends from a smiling and happy figure to a frowned and unhappy figure when the valence dimension is represented and extends from an excited figure with eyes open to a relaxed sleepy figure for the arousal dimension[2]. We have omitted a dominance dimension to simplify the experiment.

SAM has been used effectively to measure emotional responses in a variety of situations, including reactions to photos [10,19], images[23], sounds[1], advertisements[24], painful stimuli[20], among others. In addition, SAM has been applied to children[9], patients with anxiety disorders[4], people with similar phobias[12], psychopaths[27], and other clinical conditions[2].

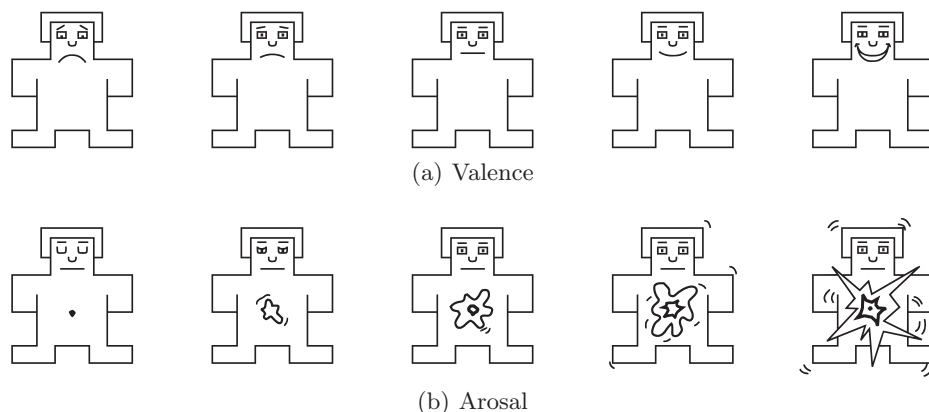


Fig. 4: SAM and its different states along the dimension axis

5 Environment description

In total 12 subjects performed all the four phases plus two researchers who performed an initial capturing test. The captures were made in December 2016 during the period of two weeks. The construction of the space, the programming of the testing interface and the devices setup took roughly six months of preparation.

To avoid environmental distractions, the acquisition was conducted in a conditioned room within the facilities of the Laboratory of Advanced Information Systems at the Faculty of Engineering of the UBA. This space was named as Emotional Immersion Chamber (EIC), consisting of approximately 4m², it is communicated with an adjacent control room where researchers enduringly supervise connectivity and captures. It has also a waiting room where the participants stay while the researchers prepare the equipment and begin the experiment. The room is isolated in both, sound and visibility, the windows are covered with green panels to facilitate the extraction of images from the video. Sound proofing had been performed to measure the level of noise and distraction in the space.

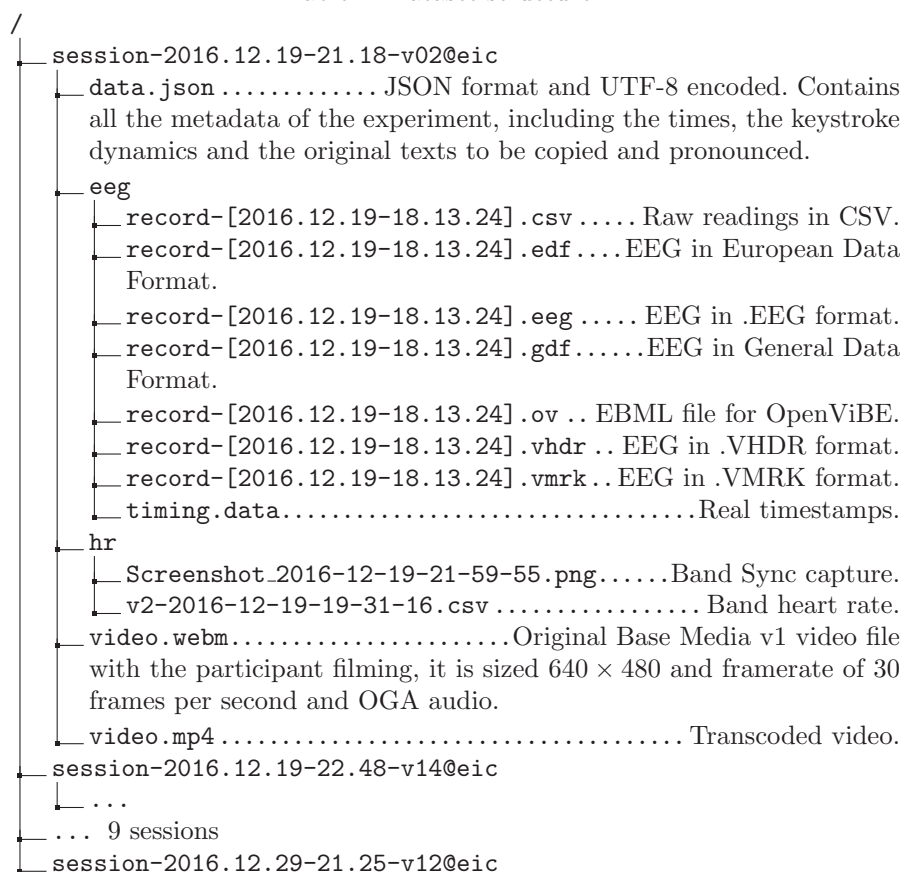
6 Data analysis

The resultant dataset consists of several files organized in directories as shown in Table 1.

There is an infinite number of opportunities to exploit the acquired data. Not only the analysis of the correlations between the different modalities but also it is possible to obtain emotional patterns that are uniquely evident in some of them. In the specific case of EEG interpretation, as a preliminary research, Fig. 5 displays a prediction of the dimensions of arousal and valence of subject

7. The first white band corresponds to the relaxation time, the four next bands represent each phase in the stimulation —dark gray for visual and light gray for audiovisual— and response —white— parts respectively. Dashed lines indicate phase transitions. It is projected to develop this line by replicating other available methods.

Table 1: Dataset structure



The raw information related to emotional states was successfully captured inside the EIC. Although some problems with the real time synchronization were detected at the beginning they were solved and a major dataset had been produced. The EIC is now available for further research on the topic. We created several Jupyter notebooks[33] to analyze all the raw information.

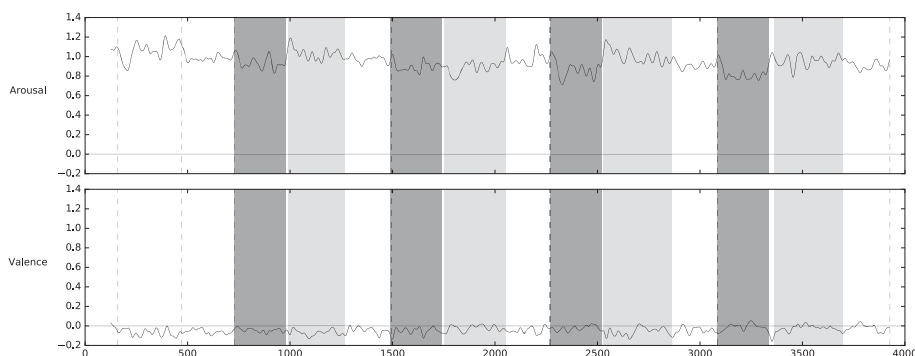


Fig. 5: Predicted arousal and valence from subject 7. Time is measured in seconds.

6.1 Future lines of research

From the collected data set, several research lines are drawn, which are intertwined in a single main objective: the generation of redundancy from the inference of emotional states through the use of multivariate biosignals.

In that sense, at first, we will seek to establish an emotional pattern for each individual through the interpretation of their respective brain activities. In addition, the generalization of this inference will be explored in order to determine the existence of a common framework that allows the emotional models of individuals to be independent.

To complete the studies on the inference of the emotional profile of an individual it will be used the heart rate as an indicator of the level of excitement that, together with keystroke dynamics, brain waves and facial recognition, will allow the construction of a first prototype of an empathic machine.

Acknowledgments. This paper acknowledges support from the University of Buenos Aires with the project under UBACyT 2014-2017 GEF localized in the Laboratorio de Sistemas de Información Avanzados, Engineering Faculty, University of Buenos Aires. The project recognizes the collaboration from the ISIER (FICCTE UM) assisting only in BMI application concerns, under the PID 01-001/12/14.

References

1. Bradley, M.M.: Emotional memory: A dimensional analysis, pp. 97–134. Taylor & Francis, Hillsdale, NJ: Erlbaum (1994), <https://books.google.com.ar/books?id=H0TrAgAAQBAJ>
2. Bradley, M.M., Lang, P.J.: Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry* 25(1), 49–59 (1994)

3. Calot, E.P., Rossi, F.M., Gonzalez, N., Hasperué, W., Ierache, J.S.: Avances en educación de dinámica de tecleo y el contexto emocional de un individuo aplicando interfaz cerebro computadora. pp. 872–876 (jun 2016), <http://hdl.handle.net/10915/53247>
4. Cook, III, E.W., Melamed, B.G., Cuthbert, B.N., McNeil, D.W., Lang, P.J.: Emotional imagery and the differential diagnosis of anxiety. *Journal of Consulting and Clinical Psychology* 56(5), 734–740 (1988), <http://dx.doi.org/10.1037/0022-006X.56.5.734>
5. Ekman, P., Friesen, W.V.: *Unmasking the Face: A Guide to Recognizing Emotions from Facial Clues*. Malor Books (2003), <https://books.google.com.ar/books?id=TukNoJDgMTUC>
6. Feldman, L.A.: Valence focus and arousal focus: Individual differences in the structure of affective experience. *Journal of personality and social psychology* 69(1), 153 (1995)
7. González, N., Calot, E.P., Ierache, J.S.: A replication of two free text keystroke dynamics experiments under harsher conditions. In: 2016 International Conference of the Biometrics Special Interest Group (BIOSIG). pp. 1–6 (Sep 2016), <http://ieeexplore.ieee.org/document/7736905/>
8. Grandjean, D., Sander, D., Scherer, K.R.: Conscious emotional experience emerges as a function of multilevel, appraisal-driven response synchronization. *Consciousness and cognition* 17(2), 484–495 (2008)
9. Greenbaum, P.E., Turner, C., Cook, III, E.W., Melamed, B.G.: Dentists' voice control: Effects on children's disruptive and affective behavior. *Health Psychology* 9(5), 546–558 (1990), <http://dx.doi.org/10.1037/0278-6133.9.5.546>
10. Greenwald, M.K., Cook, III, E.W., Lang, P.J.: Affective judgment and psychophysiological response: Dimensional covariation in the evaluation of pictorial stimuli. *Journal of psychophysiology* 3(1), 51–64 (1989)
11. Gunes, H., Schuller, B., Pantic, M., Cowie, R.: Emotion representation, analysis and synthesis in continuous space: A survey. In: *Automatic Face & Gesture Recognition and Workshops (FG 2011)*, 2011 IEEE International Conference on. pp. 827–834. IEEE (Mar 2011)
12. Hamm, A.O., Globisch, J., Cuthbert, B.N., Vaitl, D.: Startle reflex modulation in simple phobics and normals. *Psychophysiology* 28, 228 (1991)
13. Ierache, J.S., Pereira, G., Calot, E.P., Iribarren, J.: Framework for brain computer interface implemented to control devices in the context of home automation. In: *XIX Congreso Argentino de Ciencias de la Computación*. No. 5736 (2013), <http://lsia.fi.uba.ar/papers/ierache13b.pdf>
14. Ierache, J.S., Pereira, G., Nervo, F., Iribarren, J.: Emotional status focused on stimuli by applying brain-machine interface. In: Feierherd, G.E., Pesado, P., Sposito, O.M. (eds.) *Computer Science & Technology Series. XX Argentine Congress of Computer Science, Selected papers*, pp. 245–256. Editorial de la Universidad Nacional de La Plata (EDULP) (2015)
15. Ierache, J.S., Pereira, G., Sattolo, I., Iribarren, J.: Aplicación de interfases lectoras de biose textasciitilde nales en el contexto de la domótica. In: *XV Workshop de Investigadores en Ciencias de la Computación* (2013), <http://lsia.fi.uba.ar/papers/ierache13.pdf>
16. Jasper, H.H.: Report of the committee on methods of clinical examination in electroencephalography. *Electroencephalography and Clinical Neurophysiology* 10(2), 370–375 (2017/07/18 1958), [http://dx.doi.org/10.1016/0013-4694\(58\)90053-1](http://dx.doi.org/10.1016/0013-4694(58)90053-1)

17. Lang, P.J.: Behavioral treatment and bio-behavioral assessment: computer applications, pp. 119–137. Ablex, Norwood, NJ (1980)
18. Lang, P.J., Bradley, M.M., Cuthbert, B.N.: International affective picture system (iaps): Affective ratings of pictures and instruction manual. Technical report A-8 (2008)
19. Lang, P.J., Greenwald, M.K., Bradley, M.M., Hamm, A.O.: Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology* 30(3), 261–273 (1993)
20. McNeil, D.W., Brunetti, D.G.: Pain and fear: A bioinformational perspective on responsivity to imagery. *Behaviour research and therapy* 30(5), 513–520 (Sep 1992)
21. Mehrabian, A.: Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology* 14(4), 261–292 (1996)
22. Milborrow, S., Morkel, J., Nicolls, F.: The muct landmarked face database. Pattern Recognition Association of South Africa (0) (2010), <http://www.dip.ee.uct.ac.za/textasciitildenicolls/publish/sm10-prasa.pdf>, <http://www.milbo.org/muct>
23. Miller, G.A., Levin, D.N., Kozak, M.J., Cook, III, E.W., McLean Jr, A., Lang, P.J.: Individual differences in imagery and the psychophysiology of emotion. *Cognition and Emotion* 1(4), 367–390 (1987)
24. Morris, J.D., Bradley, M.M., Waine, C.A., Lang, J.B.: Assessing affective reactions to advertisements with the self-assessment manikin (sam). In: Southern Marketing Association Conference (1992)
25. Oude Bos, D.: Eeg-based emotion recognition-the influence of visual and auditory stimuli. *Capita Selecta (MSc course)* (2006)
26. Øygaard, A.M.: Emotion detection example. *clmtrackr*. (2015), <http://auduno.github.io/clmtrackr/examples/clm\emotiondetection.html>, accessed: 2017-05-23
27. Patrick, C.J., Bradley, M.M., Lang, P.J.: Emotion in the criminal psychopath: Startle reflex modulation. *Journal of Abnormal Psychology* 102(1), 82–92 (1993), <http://dx.doi.org/10.1037/0021-843X.102.1.82>
28. Picard, R.W.: *Affective Computing* (MIT Press). The MIT Press (2000)
29. Rossi, F.M., Calot, E.P., Ierache, J.S.: Educación emocional de un individuo en contextos multimodales en computación afectiva. In: XIX Workshop de Investigadores en Ciencias de la Computación (WICC 2017, Buenos Aires, Argentina) (2017)
30. Russell, J.A.: A circumplex model of affect. *Journal of Personality and Social Psychology* 39(6), 1161–1178 (1980), <http://dx.doi.org/10.1037/h0077714>
31. Saragih, J.M., Lucey, S., Cohn, J.F.: Deformable model fitting by regularized landmark mean-shift. *International Journal of Computer Vision* 91(2), 200–215 (Jan 2011), <http://dx.doi.org/10.1007/s11263-010-0380-4>
32. Scherer, K.R., Schorr, A., Johnstone, T.: *Appraisal Processes in Emotion: Theory, Methods, Research*. Series in Affective Science, Oxford University Press (2001), <https://books.google.com.ar/books?id=IWLnBwAAQBAJ>
33. Thomas, K., Benjamin, R.K., Fernando, P., Brian, G., Matthias, B., Jonathan, F., Kyle, K., Jessica, H., Jason, G., Sylvain, C., et al.: Jupyter notebooks – a publishing format for reproducible computational workflows. *Stand Alone 0(Positioning and Power in Academic Publishing: Players, Agents and Agendas)*, 87–90 (2016), <http://doi.org/10.3233/978-1-61499-649-1-87>

Reconocimiento de gestos aplicado al control de dispositivos

Una experiencia en control de TV

Carla Luna Gennari, César Estrebou, and Laura Lanzarini

Instituto de Investigación en Informática LIDI. Facultad de informática.
Universidad Nacional de La Plata
carla.lunagennari@gmail.com, {cesarest, laural}@lidi.info.unlp.edu.ar

Resumen El avance tecnológico ha permitido construir controladores cada vez más elaborados a la hora de desarrollar una interfaz hombre-máquina amigable. En esta dirección, el reconocimiento de gestos hechos con las manos ha recibido gran atención ya que su implementación sólo requiere de un hardware mínimo.

Este trabajo presenta el prototipo de un sistema reconocedor de gestos capaz de detectar las manos y realizar su seguimiento. Según el gesto del cual se trate se generan comandos que al transmitirlos de manera infrarroja permiten controlar un dispositivo electrónico. Durante el desarrollo se combinaron soluciones de distintas áreas entre las que se destacan: redes neuronales, procesamiento de imágenes y electrónica. Los resultados obtenidos han sido satisfactorios. Cabe aclarar que, si bien el desempeño de la solución obtenida ha sido medido al controlar las funciones básicas de un televisor, la misma puede ser aplicada en numerosas situaciones.

Keywords: Segmentación de Manos, Reconocimiento de Gestos, Interfaz Hombre-Máquina, Control de Dispositivos

1. Introducción

Una interfaz hombre-máquina juega un papel importante a la hora de transmitir una intención de un usuario a un dispositivo. A medida que la tecnología avanza se incorporan al mercado soluciones nuevas con distintos tipos de hardware y una forma de comunicación propia. Elementos convencionales como el mouse y teclado, hacen que la operatoria en ocasiones sea compleja. En cuanto a los comandos por voz, si bien han sido ampliamente utilizados, presentan serias limitaciones ante la diversidad de idiomas o cuando deben operar en ambientes ruidosos.

Por lo antes dicho, hoy en día existe mucho interés en el reconocimiento de gestos hechos con las manos. En la actualidad hay diferentes trabajos realizados con aplicación en áreas muy diferentes como por ejemplo realidad aumentada, control de dispositivos o reconocimiento del lenguaje de señas Argentino entre otras [6, 7, 11]. También existen trabajos relacionados donde se utilizan los gestos

hechos con las manos para controlar las funciones básicas de un equipo de audio ubicado dentro de un vehículo [9].

Para llevar a cabo el reconocimiento de gestos hechos con las manos deben resolverse dos partes perfectamente diferenciadas. La primera de ellas se refiere específicamente a la segmentación de las manos y la segunda se ocupa de la caracterización y reconocimiento del gesto. Ambas presentan diferentes niveles de complejidad según el tipo de cámara que se utilice así como la cantidad y el tipo de gestos a reconocer.

En lo que se refiere a la cámara de video, existen dispositivos que brindan información en 3D permitiendo captar no sólo el desplazamiento de los objetos sino la profundidad a la cual se encuentran con respecto al punto de observación [2]. Por ejemplo en [5] se utiliza un mapa de profundidad para ayudar a la segmentación. La elección de la cámara es una relación de compromiso entre el costo del equipamiento necesario para resolver el problema y la complejidad del algoritmo de segmentación a desarrollar. En este trabajo se ha decidido utilizar una cámara de video convencional e identificar la zona en la cual se encuentran las manos a través de un reconocedor de colores de piel basado en una red neuronal similar a la indicada en [12, 15]. Para llevar a cabo su entrenamiento se construyó una base de datos de colores de piel ya que al momento de diseñar este prototipo no se disponía de este tipo de información.

Para la segunda parte del problema es necesario tener en cuenta la diversidad de formas que puede tomar una mano humana a la hora de realizar un gesto. El reconocimiento está ligado a la elección específica de los gestos que se quieren procesar.

Este trabajo presenta el prototipo de un sistema reconocedor de gestos hechos con una mano. Según el gesto del cual se trate, genera comandos que al transmitirlos de manera infrarroja permiten controlar un dispositivo electrónico. Se trata de una solución de bajo costo para la que se describe tanto el diseño de hardware como el software de control. En este caso particular, fue utilizado para reemplazar el control remoto de un TV, posibilitando cambiar de canal y subir o bajar el volumen con sólo mover una mano frente al televisor. Su diseño modular permite incorporar el reconocimiento de nuevos gestos así como los comandos necesarios para operar con otros aparatos.

Utilizar gestos hechos con las manos para interactuar con dispositivos puede resultar sumamente útil para personas con movilidad reducida.

2. Hardware para Captura y Control de dispositivos

Una parte fundamental en la construcción de este prototipo fue la elección de los componentes de hardware para obtener un dispositivo embebido que respondiera tanto a los requerimientos de software como a los de futuras ampliaciones de hardware. Los requerimientos propuestos que debía cumplir son los siguientes: bajo costo económico, capacidad de cómputo para procesamiento de imágenes, capacidad de comunicación con otros dispositivos a bajo nivel, tamaño reducido,



Figura 1: Módulos que componen el hardware. (a) Raspberry Pi 3. (b) Cámara web. (c) Sensor de luminosidad. (d) Emisor y receptor infrarrojos.

bajo consumo de energía, capacidad de funcionar con baterías y capacidad de actualización de hardware y software.

Teniendo en cuenta los requerimientos antes mencionados se decidió utilizar una Raspberry Pi 3 (Fig. 1a) como corazón del dispositivo. Cuenta con 1Gb de memoria RAM y un procesador ARM Cortex-A53 de 1.2GHz con 4 núcleos lo que ofrece un nivel de cómputo aceptable para procesamiento de imágenes moderado. También tiene un tamaño reducido, capacidad de comunicación con otros dispositivos a través de su GPIO y de sus 4 puertos USB. Cuenta con comunicación inalámbrica vía WiFi y Bluetooth. Requiere una alimentación de 5V y 1.5A para nuestro tipo de aplicación, lo que hace posible el uso de baterías, como por ejemplo un powerbank. Todo esto a un costo razonable.

Para capturar los gestos de las manos se utilizó una cámara web convencional (Fig. 1b) con una tasa de adquisición de unos 20 cuadros por segundo a una resolución de 1280 x 1024 píxeles. Para controlar las funciones del TV se utilizó un módulo emisor de luz infrarroja (Fig. 1d) y un módulo receptor infrarrojo que permite incorporar al sistema los códigos de aquellos controles remotos que no tengan códigos conocidos. También se utilizó un módulo para sensar la luz ambiente (Fig. 1c) con el objetivo de realizar ajustes en la luminosidad de la imagen capturada con la cámara web.

3. Software para Reconocimiento de Gestos

Teniendo en cuenta que esta es la primera versión del dispositivo, se decidió limitar su funcionamiento a una aplicación relativamente simple que no demandara una gran cantidad de cómputo. De esta manera se construyó un prototipo funcional tanto de hardware como de software que permite reemplazar el control remoto de un televisor convencional por un controlador que reciba las instrucciones a través de gestos realizados con las manos. En este sentido se limita tanto el modelo de representación como la cantidad de gestos de la mano, aunque en futuras versiones se tiene planificado ir incorporando un modelo más complejo como el planteado aquí [10].

Respecto del software utilizado para el desarrollo de este prototipo, se utilizó la versión Jessie de Raspbian, que es el sistema operativo con soporte oficial

de Raspberry. El sistema fue implementado en Python por su facilidad para prototipado rápido y por la gran disponibilidad de bibliotecas que ofrecen tanto algoritmos tradicionales como de vanguardia. Para las partes que requirieron procesamiento de imágenes se utilizó la biblioteca OpenCV que tiene una sólida madurez y sus algoritmos están altamente optimizados, incluso para aprovechar las características de la GPU. Tanto para el entrenamiento y funcionamiento de la red neuronal RCE como para el algoritmo de seguimiento de la mano se utilizó una implementación propia optimizada para NumPy.

En las secciones que siguen a continuación se describen en detalle los pasos de cada una de las etapas en las que se divide todo el proceso de reconocimiento del gesto para controlar un TV. En el esquema de la figura 2 se puede observar todo el proceso que realiza el software del dispositivo.

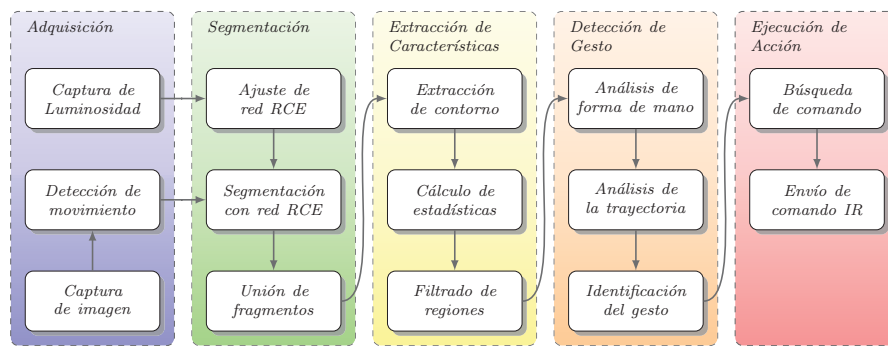


Figura 2: Proceso de reconocimiento del gesto y ejecución de comandos.

3.1. Adquisición

La etapa de adquisición se divide en dos partes. La primera consiste en la captura, mediante una cámara web convencional, de una imagen RGB con una resolución de 1280 x 1024 píxeles a una tasa de adquisición que puede variar de 10 a 20 cuadros por segundo. Una vez obtenida la imagen correspondiente al cuadro actual, se la compara con la del cuadro adquirido anteriormente para determinar si se produjo movimiento y así evitar el procesamiento innecesario.

La segunda parte consiste en la adquisición de la cantidad de luz ambiente a través de un transductor que mide esta variable en lux (lumen/m^2). Este valor obtenido permite ajustar la luminosidad de la imagen en la etapa de segmentación y de esta manera corregir las variaciones en la iluminación que afectan a los colores.

3.2. Segmentación

Para realizar la detección del color de piel en la imagen se utilizó una red neuronal RCE [8] (Restricted Coulomb Energy). Esta arquitectura de red presenta tres capas de neuronas, donde la primer capa y la segunda están completamente conectadas, y la segunda capa con la tercera solo estén conectadas parcialmente. La capa de entrada es decir la primera, es representada por los valores del espacio de color utilizado, en este caso RGB. La capa intermedia, llamada capa de prototipo, contiene información de color de los píxeles, los cuales corresponden a valores RGB que representan colores de piel que fueron incorporados durante el entrenamiento de la red o etapa de aprendizaje. La última capa o capa de salida contiene las clases a las que puede corresponder el color del pixel entrante. En este caso sólo hay una única clase que responde para determinar si un color corresponde o no a la piel humana.

Para realizar la segmentación de la piel humana se ingresa cada pixel de la imagen a la red RCE para determinar si éste se corresponde o no al color de la piel. Como resultado de esta operación se obtiene una máscara preliminar a la que luego se le aplica una operación morfológica de cierre para unir áreas que pudieran haber quedado desconectadas.

Es importante destacar que antes de realizar la segmentación, con el objetivo de subsanar el problema de inestabilidad que provocan las variaciones de iluminación, se aplica un ajuste a la red RCE según la luz ambiente captada por el sensor de luminosidad. Esta corrección se aplica a los valores RGB de las neuronas de la capa intermedia para neutralizar los cambios de intensidad de luz que están presentes en la imagen.

3.3. Extracción de características

En esta etapa, luego de obtenidas las áreas de la imagen donde se localiza la piel, se determina cuales de todas esas porciones pueden corresponderse con la mano de una persona.

El proceso inicia con la máscara que representa a todos los píxeles reconocidos como piel en la etapa anterior. A ésta máscara se le aplica un algoritmo de extracción de contornos para obtener un listado de estos. Cada contorno agrupa píxeles interconectados en la máscara que forman una región donde potencialmente podría haber una mano.

Una vez obtenido el listado, por cada contorno se calcula el área, el perímetro y los ejes principales que son propiedades que dan una pauta de las características geométricas generales de la región.

Finalmente, para filtrar el listado, se analiza cada contorno o región comparando sus propiedades con las propiedades de una región que contiene una mano, descartando aquellos que difieren mucho de lo esperado. Como resultado de esta comparación se obtiene un listado de contornos candidatos que representan regiones potenciales donde puede encontrarse una mano.

3.4. Reconocimiento del gesto

Para determinar si las regiones candidatas obtenidas en la etapa anterior se corresponden o no con una mano se aplica la técnica de *Template Matching*. En esta técnica se utiliza un listado de plantillas pre-definidas que representan las formas de las manos aceptadas o reconocidas por la aplicación. Cada región del listado de candidatas se compara con cada plantilla pre-definida para determinar el grado de correlación que existe entre ambas. La forma de la mano reconocida es la que mayor valor de correlación tenga, siempre que se supere un umbral establecido como parámetro de configuración.

Una vez que se obtiene una región candidata que coincide con alguna de las formas pre-definidas de mano, se inicia el proceso de seguimiento tomando como referencia la posición de la región como posición de inicio. Luego se analiza la secuencia posterior de imágenes localizando la mano para determinar la evolución de su posición. Cuando la distancia entre la posición inicial y la posición actual de la mano supera un valor determinado por un parámetro de configuración se establece la dirección del movimiento y se procede a determinar si el gesto coincide con los definidos en la aplicación.

3.5. Ejecución de Acción

Identificado el gesto de la mano, se determina el comando infrarrojo que éste tiene asociado para enviarlo al TV. Para ello, se utiliza un pequeño módulo de hardware que genera una señal infrarroja que es recibida por el TV, quien la decodifica e interpreta para ejecutar la función asociada a dicha señal.

Para codificar los comandos infrarrojos se utiliza la biblioteca LIRC (Linux Infrared Remote Control). Esta biblioteca permite decodificar y reproducir una secuencia de pulsos infrarrojos de la misma manera que lo hace un control remoto convencional. Cuenta con una gran cantidad de códigos predefinidos de controles remotos e incorpora un servicio con la capacidad de copiar, desde un control remoto, aquellos códigos que no están predefinidos.

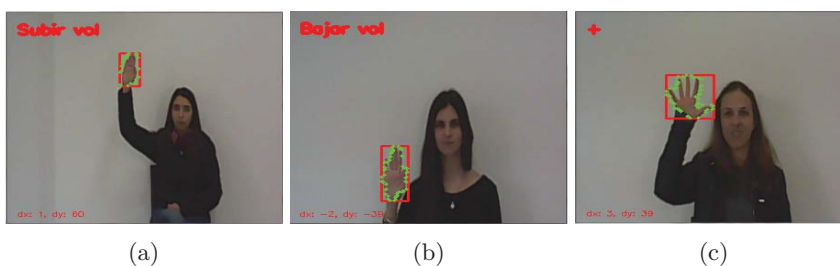


Figura 3: Reconocimiento del gesto y acción realizada. (a) Subir el volumen. (b) Bajar el volumen. (c) Siguiente canal.

4. Pruebas/Resultados

4.1. Sistemas de Color y Red Neuronal RCE

Para la segmentación de manos se utilizó una red neuronal RCE [8] que determina cuando un pixel de la imagen se corresponde con el color de la piel. En la revisión de la literatura sobre la segmentación de piel basada en el color del pixel [1, 4, 13, 14] se encuentra que distintos algoritmos aplicados a imágenes en diferentes sistemas de representación del color obtienen resultados aceptables. Por este motivo, se decidió realizar una serie de pruebas en los sistemas de representación RGB, HSL, HSV, YCbCr, Cie-LAB para determinar cual es el más conveniente. Para realizar las pruebas en los distintos sistemas se utilizó la base de datos MOHI [3] que contiene muestras de manos de 250 personas y una base de datos construida ad hoc con imágenes que no contienen piel.

En la figura 4 se muestran los resultados obtenidos en las distintas pruebas realizadas. Para cada sistema de color se muestran dos barras que expresan el promedio de píxeles clasificados como piel cuando la imagen contiene piel y cuando no contiene piel. En general se puede observar que no hay grandes diferencias entre los distintos sistemas. También se puede observar que una mejora en la detección de los píxeles de la BDD con piel incrementa la cantidad de píxeles detectados en la BDD sin piel y viceversa. En consecuencia, se decidió optar por utilizar el sistema RGB para la segmentación con la red RCE para evitar el costo del cómputo de la transformación a otro sistema.

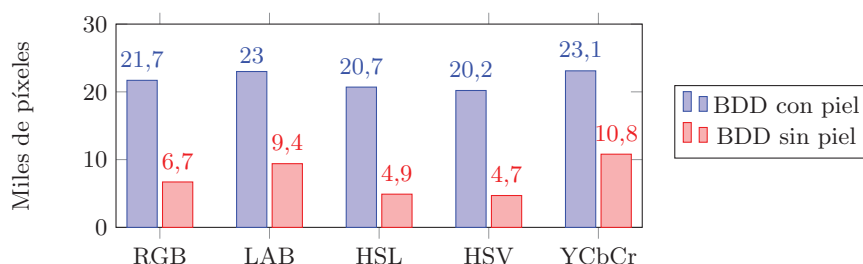


Figura 4: Red neuronal RCE en diferentes sistemas de representación del color.

4.2. Reconocimiento de Gestos

Con el prototipo terminado se realizaron pruebas para medir su precisión en un ambiente controlado. Para esto se utilizaron 11 sujetos en dos condiciones diferentes de iluminación. Una condición de iluminación es baja (30 lux) y la otra condición de iluminación es media (al menos unos 600 lux). De cada sujeto se tomaron 3 formas de la mano (figura 6): mano abierta con dedos separados, mano abierta con dedos juntos y mano cerrada con índice y pulgar separados.

Con cada forma se realizaron movimientos en dos direcciones opuestas (arriba y abajo). En la tabla de la figura 5 se pueden observar los resultados obtenidos.

Gesto	Distancia 180 cm		Distancia 300 cm	
	+600 lux	30 lux	+600 lux	30 lux
Mano Abierta - Dedos separados - Arriba	100 %	100 %	100 %	100 %
Mano Abierta - Dedos separados - Abajo	100 %	100 %	100 %	100 %
Mano Abierta - Dedos juntos - Arriba	100 %	45 %	100 %	33 %
Mano Abierta - Dedos juntos - Abajo	100 %	45 %	100 %	33 %
Mano con índice y pulgar - Arriba	40 %	0 %	30 %	0 %
Mano con índice y pulgar - Abajo	40 %	0 %	30 %	0 %

Figura 5: Resultados en porcentaje de la detección de los gestos.



Figura 6: Formas de manos procesadas. (a) Mano Abierta con dedos separados. (b) Mano abierta con dedos juntos. (c) Mano con índice y pulgar.

Para las pruebas con las formas de las manos se puede observar que para los casos de mano abierta se pueden detectar bien en condiciones de iluminación aceptable. Para los casos de mano abierta con dedos juntos se puede observar que falla cuando la iluminación es baja. Para los casos de la mano con índice y pulgar se observa que la detección falla de manera importante en las 2 condiciones de iluminación. La principal causa de este problema es la sombra que genera la flexión de los dedos. Esta sombra hace que la segmentación falle al no poder reconocerla como piel y se produzcan separaciones importantes que hacen que la región de la mano se extraiga parcialmente. Luego no se encuentran coincidencias o bien porque la región extraída se descarta porque no cumple con las propiedades geométricas esperadas o porque la técnica de *Template Matching* falla porque no es robusta para encontrar coincidencias parciales.

5. Conclusiones y Trabajos Futuros

En este trabajo se ha presentado la primera versión de un prototipo de un sistema reconocedor de gestos realizados con las manos para controlar las funciones de un TV.

Uno de los resultados positivos de este trabajo es que se ha obtenido una segmentación de piel aceptable utilizando una red neuronal RCE que funciona independientemente del sistema de color en que utiliza la imagen.

Un aspecto aún no resuelto es la sensibilidad del reconocedor propuesto a la variación de la luz ambiente. En este sentido, se encontró una solución parcial a esta dificultad utilizando lecturas de un sensor de luz para adaptar la red neuronal RCE a las condiciones de iluminación del ambiente. Cuando hay ausencia de luz ambiente o esta es notablemente baja, la corrección por intensidad de la luz no resulta suficiente para una cámara web convencional. En este caso es conveniente realizar una adaptación de la cámara para que funcione con iluminación infrarroja.

Debido a que este es un trabajo inicial, hay algunas líneas de trabajo que se están desarrollando y otras líneas que quedan por explorar y desarrollar. Actualmente se está trabajando para mejorar la segmentación de las áreas de la piel. Muchas veces debido a la combinación de iluminación con la posición de las manos, se producen sombras en las uniones de las falanges que provocan que la segmentación deje los dedos separados de las manos. Una solución que se está explorando es la incorporación al entrenamiento de la red neuronal RCE ejemplos de piel con sombra, lo que permitiría mejorar la segmentación en la unión de las falanges. Los resultados obtenidos en algunas pruebas preliminares que se realizaron al momento de la escritura de este documento son prometedores en este aspecto.

En lo que se refiere al hardware, aún resta explorar las maneras de controlar las funciones distintos dispositivos electrónicos mediante red cableada, red inalámbrica, bluetooth e infrarrojo para aprovechar las capacidades del dispositivo y expandirlas. Finalmente quedan por analizar algunas alternativas de reemplazo de la Raspberry Pi 3 para reducir el costo. Dispositivos como Orange Pi, en sus versiones Zero, One y Lite podrían reducir el costo entre un tercio y la mitad de una Raspberry. Estos dispositivos son compatibles y ofrecen características de hardware similares con algunas limitaciones que van desde una menor cantidad de memoria RAM, menos puertos USB hasta la ausencia de WiFi y bluetooth según el modelo.

Referencias

1. Chaves-González, J.M., Vega-Rodríguez, M.A., Gómez-Pulido, J.A., Sánchez-Pérez, J.M.: Detecting skin in face recognition systems: A colour spaces study. *Digit. Signal Process.* 20(3), 806–823 (May 2010)
2. Dinh, D.L., Kim, J.T., Kim, T.S.: Hand gesture recognition and interface via a depth imaging sensor for smart home appliances. *Energy Procedia* 62(Complete), 576–582 (2014)

3. Hassanat, A., Al-Awadi, M., Btoush, E., Al-Btoush, A., Alhasanat, E., Altarawneh, G.: New mobile phone and webcam hand images databases for personal authentication and identification. *Procedia Manufacturing* 3, 4060 – 4067 (2015), 6th International Conference on Applied Human Factors and Ergonomics and the Affiliated Conferences, AHFE 2015
4. Kakumanu, P., Makrogiannis, S., Bourbakis, N.: A survey of skin-color modeling and detection methods. *Pattern Recogn.* 40(3), 1106–1122 (Mar 2007)
5. Kang, B., Tan, K., Tai, H., Tretter, D., Nguyen, T.Q.: Hand segmentation for hand-object interaction from depth map. *CoRR abs/1603.02345* (2016)
6. Piumsomboon, T., Clark, A., Billinghamurst, M., Cockburn, A.: User-defined gestures for augmented reality. In: *CHI '13 Extended Abstracts on Human Factors in Computing Systems*. pp. 955–960. CHI EA '13, ACM, New York, NY, USA (2013)
7. Reifinger, S., Wallhoff, F., Ablassmeier, M., Poitschke, T., Rigoll, G.: Static and dynamic hand-gesture recognition for augmented reality applications. In: *HCI Intelligent Multimodal Interaction Environments: 12th International Conference, Part III*. pp. 728–737. Springer, Berlin, Heidelberg (2007)
8. Reilly, D.L., Cooper, L.N., Elbaum, C.: A neural model for category learning. *Biological Cybernetics* 45(1), 35–41 (Aug 1982)
9. Riener, A.: Gestural interaction in vehicular applications. In: *Computer*. vol. 45, pp. 42–47. IEEE Computer Society, Los Alamitos, CA, USA (2012)
10. Ronchetti, F., Quiroga, F., Estrebou, C., Lanzarini, L.: Clasificación de configuraciones de manos del lenguaje de señas argentino con probsom. In: *XXI Congreso Argentino de Ciencias de la Computación* (2015)
11. Ronchetti, F., Quiroga, F., Estrebou, C., Lanzarini, L., Rosete, A.: Sign Language Recognition Without Frame-Sequencing Constraints: A Proof of Concept on the Argentinian Sign Language, pp. 338–349. Springer, Cham (2016)
12. Sui, C., Kwok, N.M., Ren, T.: A restricted coulomb energy (rce) neural network system for hand image segmentation. In: *Computer and Robot Vision (CRV), 2011 Canadian Conference on*. pp. 270–277. IEEE (2011)
13. Vezhnevets, V., Sazonov, V., Andreeva, A.: A survey on pixel-based skin color detection techniques. In: *GraphiCon*. pp. 85–92 (2003)
14. Xu, Z., Zhu, M.: Color-based skin detection: survey and evaluation. In: *2006 12th International Multi-Media Modelling Conference*. pp. 10 pp.– (2006)
15. Yin, X., Guo, D., Xie, M.: Hand image segmentation using color and rce neural network. *Robotics and Autonomous Systems* 34(4), 235–250 (2001)

VIII Workshop Procesamiento de Señales y Sistemas de Tiempo Real (WPSTR)

Coordinadores

Sposito, Osvaldo - Villagarcia Wanza, Horacio – Frati,
Emmanuel – Ierache, Jorge

Comité Científico

Acosta, Nelson (Argentina)	Ramón, Hugo (Argentina)
Bría, Oscar (Argentina)	Riba, Alberto Eduardo (Argentina)
Encinas, Diego (Argentina)	Romero, Fernando (Argentina)
Lugo, Jorge (Argentina)	Runco, Jorge (Argentina)
Maidana, Carlos (Argentina)	Suppi, Remo (España)

Análisis de Experimentos de Comunicaciones de Tiempo Real en el Bus CAN

Fernando G. Tinetti¹, Fernando L. Romero, Alejandro D. Pérez

Instituto de Investigación en Informática LIDI (III-LIDI)
Facultad de Informática, UNLP, 50 y 120, La Plata, Argentina
{fernando, fromero}@lidi.info.unlp.edu.ar, perez.alejandrodaniel@gmail.com

Resumen. Este trabajo presenta el desarrollo de un análisis de tiempos de comunicaciones en CAN, inicialmente desde la perspectiva de planificación de tareas con prioridad fija apropiativa para sistemas monoprocesador. Este método sería adaptado para la programación de mensajes CAN, que no incluye la posibilidad de apropiación del canal de comunicaciones compartido. A su vez, también se complementa con un algoritmo de asignación de prioridades óptimo aplicable a los sistemas no apropiativos. Se muestran resultados sobre una implementación de una red real para verificar experimentalmente estos tiempos. Este análisis garantiza tiempos conocidos y correctos, los mensajes y señales serán distribuidos dentro de los plazos, requisito fundamental para un sistema de tiempo real.

Palabras Clave: Bus CAN, Comunicaciones de Tiempo Real, Sistemas de Tiempo Real.

1 Introducción

El desarrollo inicial de un análisis de tiempos CAN se basa en [5], directamente sobre la planificación de tareas con prioridad fija apropiativa para sistemas de un único procesador. Este método es adaptado para la programación de mensajes CAN y se complementa con lo investigado en [3], que mostró cómo un algoritmo de asignación de prioridades óptimo introducido en [1] es aplicable a los sistemas no apropiativos. Todos estos desarrollos se pueden verificar de manera experimental al menos en parte construyendo una red CAN real sobre la cual tomar tiempos de comunicaciones, evaluando los plazos de tiempo real que se pueden cumplir/asegurar.

Los buses definidos en CAN se arbitran utilizando prioridades, y una vez que se obtiene el acceso al bus, el mensaje se transmite completo, sin posibilidad de apropiatividad en el caso de la llegada de un mensaje de mayor prioridad mientras se está transmitiendo el mensaje para el que se accedió al canal de comunicaciones. Para garantizar que se transmitan los mensajes de más alta prioridad con baja latencia, se debe tratar la interferencia, o en realidad el retraso producido por parte de señales menos urgentes, con menor prioridad. Por otro lado, también se debe tener en cuenta

¹Comisión de Investigaciones Científicas de la Prov. de Buenos Aires.

en general que no solamente se tiene la posibilidad de tiempo de espera en estos casos de prioridades invertidas (un mensaje de menor prioridad está usando el bus mientras uno de mayor prioridad está esperando para acceder), sino que todos los mensajes deben esperar el acceso al medio de comunicación cuando hay mensajes de mayor prioridad, que accederán previamente. Todo el desarrollo del trabajo se enfocará en el comportamiento de la transmisión de Data Frames del estándar CAN.

Los mensajes de todos los nodos en la red que acceden al medio de transmisión por un controlador comparten una única cola basada en prioridades y son enviados al bus acorde a una planificación no apropiativa de prioridad fija. CAN no tiene un reloj único/global para todos los controladores, por lo tanto cada controlador CAN tiene su propia señal de reloj, que con un grado de tolerancia especificado por el protocolo, puede comunicarse con el resto de los nodos. Por ello, los nodos se resincronizan al momento de la transmisión de un mensaje en el flanco ascendente del bit de comienzo de una transmisión, causado por quien haya comenzado a transmitir primero. Los nodos pueden comenzar una transmisión cuando el bus se encuentra desocupado por un espacio de 3 bits interframe. El nodo comienza a transmitir con un bit dominante ('0') lo que provoca que el resto de los nodos se sincronicen con el flanco ascendente de este bit y se vuelvan receptores.

Cualquier mensaje que esté listo para ser transmitido deberá esperar a que se inicie otro ciclo de arbitraje en el momento en el cual el bus quede nuevamente desocupado. Por definición, el protocolo CAN necesita que el transmisor inserte un bit con polaridad opuesta cada 5 bits de la misma polaridad. Este proceso es conocido como "bit stuffing" y es luego invertido por el receptor. En los primeros trabajos, como en [5] no se tuvo en cuenta este proceso, pero luego en [2], se agrega este proceso al modelado de tiempos para la transmisión de los mensajes. En el peor caso, se debe agregar un bit de stuffing cada 4 bits del mensaje y, por lo tanto, el tiempo máximo de transmisión estaría dado por C_m en la Eq. (1)

$$C_m = \left(g + 8s_m + 13 + \left\lfloor \frac{g + 8s_m - 1}{4} \right\rfloor \right) \tau_{bit} \quad (1)$$

donde s_m es la cantidad de bytes de datos del mensaje, g es una constante, 34 para el formato estándar de identificador o 54 para el formato extendido, y τ_{bit} es el tiempo de transmisión de un bit.

2 Tiempos: Esperas, Máximo, Errores

Si bien los tipos de mensajes son dependientes de la aplicación, los tipos de mensajes son predefinidos para la simulación y para la implementación. De hecho, cada mensaje m tiene un identificador y prioridad únicos y fijos, se puede decir tanto mensaje m como mensaje de prioridad m . Cada mensaje se encola por una rutina de software, por el flujo normal del programa o por ejecución de una interrupción en el host. Esta tarea necesita de un tiempo para ejecutarse entre 0 y J_m , queuing jitter del mensaje, y es heredado desde el tiempo de respuesta de la tarea, incluyendo la demora por polling [5]. Los eventos que disparan el encolamiento ocurren con un mínimo de periodicidad de T_m , el periodo del mensaje, siendo soportados los tipos de eventos: a)

Los que ocurren estrictamente con un periodo de T_m , b) Los que ocurren esporádicamente con un mínimo de separación de T_m , c) Los que ocurren solo cuando el sistema es reiniciado. Cada mensaje tiene un *hard deadline* D_m : tiempo máximo desde el evento inicial hasta la transmisión del mensaje.

Se puede definir el peor caso de tiempo de respuesta R_m de un mensaje como el mayor tiempo desde que ocurre el evento de inicio hasta que el mensaje comienza a ser recibido por los nodos que lo requieran. Un mensaje se puede llamar “planificable” (schedulable) si y sólo si su peor caso de tiempo de respuesta es menor o igual a D_m ($R_m \leq D_m$). El sistema es “planificable” (schedulable) si y sólo si todos los mensajes son planificables. R_m es, básicamente, la suma de dos tiempos: 1) El retardo W_m (también llamado *Window time*), que es el mayor tiempo de permanencia del mensaje en el controlador CAN, y 2) El peor tiempo de transmisión, C_m . A su vez, el retardo del tiempo de ventana W_m se compone de: 1) Tiempo de bloqueo, B_m , dado por mensajes de menor prioridad en proceso de transmisión, y 2) Tiempo de interferencia, donde un mensaje con prioridad más alta gana el arbitraje y se transmite en vez del mensaje m .

El tiempo de bloqueo es relativamente sencillo de calcular, como $B_m = \max(C_k)$ con $k \in lp(m)$, donde $lp(m)$ es el conjunto de mensajes con menor prioridad que m . El tiempo de interferencia está muy relacionado al concepto de “período ocupado” (busy period) introducido en [4], donde se definen a estos períodos ocupados de nivel i como el intervalo de tiempo $[a, b]$ dentro del cual las tareas con una prioridad i o mayor son procesados dentro de este intervalo y no se tienen tareas de esta prioridad en los intervalos $(a - t, a)$ y $(b, b + t)$ siendo un tiempo $t > 0$. Adaptando esta definición al protocolo CAN, se tiene que un período ocupado de prioridad m . El período se inicia en un tiempo T_s donde un mensaje de prioridad m o mayor es encolado listo para transmitir, y no hay mensajes de prioridad m o mayor que estén esperando a ser transmitidos, encolados estrictamente antes de T_s . Es un intervalo continuo de tiempo, durante el cual cualquier mensaje de prioridad menor que m está imposibilitado de comenzar una transmisión y ganar el arbitraje. Finalmente, el período concluye en un tiempo T_e , cuando el bus pasa a estar listo para el siguiente ciclo de transmisión y arbitraje y aún no hay mensajes de prioridad m o superior esperando a ser transmitidos que hayan sido encolados estrictamente antes del tiempo T_e .

La característica clave del periodo ocupado es que todos los mensajes de prioridad m o superior encolados estrictamente antes de la finalización son transmitidos mientras dure el periodo. En términos matemáticos, los períodos ocupados pueden ser vistos como un intervalo $[t_s, T_e)$. En consecuencia, el final de un periodo ocupado puede coincidir con el comienzo de otro periodo ocupado [4]. Se puede calcular tiempo ocupado como el máximo de

$$t_m^{n+1} = B + \sum_{\forall j \in hp(m)} \left\lceil \frac{t_m^n + J_j + \tau_{bit}}{T_j} \right\rceil C_j \quad (2)$$

donde $hp(m)$ es el conjunto de mensajes con prioridad más alta que m , y J_j y T_j se relacionan con el jitter y la periodicidad de los mensajes respectivamente, tal como se explica anteriormente. Se puede observar que el tiempo ocupado toma una forma monótona creciente, iniciando con el valor B_m e incrementándose hasta alcanzar uno

de los siguientes casos: a) $t_m^n + C_m > D_m$, en el que se tiene un mensaje que no es válido por superar su deadline, o b) $t_m^{n+1} = t_m^n$, donde se tiene el peor tiempo de respuesta de la primera instancia del mensaje en el período ocupado que está dado por $t_m^n + C_m$.

Al considerar los errores y los tiempos de recuperación relacionado, un nodo al detectar un error comienza el proceso de recuperación transmitiendo un indicador de error. El proceso de resincronización toma al menos 29 tiempos de transmisión de 1 bit. Se define una función que represente este error, $E(t)$, como el mayor tiempo que necesario para la recuperación de los errores en cualquier intervalo de duración t [4]. La sobrecarga que involucran los errores en cualquier intervalo de duración t sería

$$\left(n_{burst} + \left\lceil \frac{t}{T_{error}} \right\rceil \right) C_{overhead}$$

Donde $C_{overhead}$ es el mayor tiempo que demora el manejo de un error y el reenvío del mensaje más largo transmitido en un intervalo, el peor caso de latencia, dado por R_m . El mayor tiempo de teniendo en cuenta los errores entonces sería

$$C_{overhead,m} = \max_{\forall j \in (hp(i) \cup \{i\})} (C_j) + 29\tau_{bit}$$

Por lo que finalmente se puede definir E_m como

$$E_m = \left(n_{burst} + \left\lceil \frac{t_m + C_m}{T_{error}} \right\rceil \right) C_{overhead,m}$$

Teniendo en cuenta esta componente aportada por los errores se llega a una ecuación con la cual analizar los peores casos en ambientes con ruido:

$$t_m = B_m + E_m + \sum_{\forall j \in hp(m)} \left\lceil \frac{t_m + J_j + \tau_{bit}}{T_j} \right\rceil C_j \quad (3)$$

3 Benchmark Especifico

La Sociedad de Ingenieros Automotrices (SAE: Society of Automotive Engineers) proporciona un banco de datos de prueba utilizado para evaluar diferentes tecnologías de comunicación de señales que son enviadas entre siete subsistemas diferentes en un prototipo de automóvil eléctrico [5]. Este banco de pruebas es usado para mostrar un caso de uso “real” para las ecuaciones y modelos que anteriores. Entre los siete subsistemas que generan señales están, por ejemplo, los frenos, el propio conductor, el motor, etc. En total se tienen 53 señales, caracterizadas no solo por el subsistema que las generan sino por la cantidad de bits de información, si son periódicas o esporádicas, jitter, etc. Un mensaje periódico tiene un periodo fijo de tiempo entre invocaciones y el esporádico un requisito de latencia impuesto por la aplicación. Se considera como válido el mensaje que cumple con esta latencia.

El análisis presentado es válido para cualquier orden del conjunto de mensajes. Es importante seleccionar el orden y las prioridades correctas a la hora de obtener un sistema planificable, y de esta manera maximizar la robustez del sistema frente a

errores. En [5] se propone una asignación de prioridades “por deadline” (*deadline monotonic priority assignment*), asignando a los de menor diferencia la mayor prioridad que solo es óptima para sistemas cuya programación de tareas sea de prioridad fija apropiativa, asumiendo plazos no mayores a sus periodos. Que no sea óptimo no significa que no sea utilizable, solo que en caso de tener tiempos más críticos se debería elegir otra opción. Por ejemplo, según la Tabla 1, y asumiendo jitter 0 y tasa de transferencia de 125 Kb/s, el cálculo cumple con lo planteado en el análisis, donde se tenía $R_m < D_m$. Sin embargo, se puede ver un contraejemplo usando

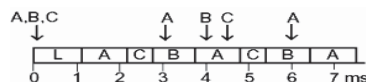
Tabla 1. Ejemplo de Caracterización de Mensajes.

Mens.	Periodo.	Deadline	Bytes	T transm.	T transm.
A	3.0 ms	3.0 ms	8	1.08 ms	2.16 ms
B	4.0 ms	4.0 ms	8	1.08 ms	2.16 ms
C	4.5 ms	4.5 ms	1	0.53 ms	3.24 ms

la misma Tabla 1, en el que no siempre todos los mensajes pueden ser atendidos antes del deadline:



en el inicio, los tres mensajes A, B, y C esperan que el mensaje L de prioridad más baja termine su transmisión. Luego de esto, A y B ganan sucesivamente el arbitraje frente a C que no llega a transmitirse antes de su deadline de 4.5 ms. Esta situación pretende ser la óptima, y según esto no debería haber otro orden de prioridad que cumpla con lo buscado. Sin embargo, si se tiene en cuenta el orden A, C y B se obtienen los peores tiempos de respuesta, con $R_A = 2.16$, $R_C = 2.68$, y $R_B = 3.76$. Con este orden de prioridad y estos tiempos de respuesta, todos los mensajes alcanzan sus deadlines:



Es decir que dando la mayor prioridad al mensaje más corto, los tres mensajes comienzan su transmisión dentro de los 3ms de ser encolados. Ninguno de ellos estará sujeto a interferencia de una segunda instancia del mensaje A y consecuentemente a una segunda instancia del mensaje B. Para este tipo de situaciones, en [3] se muestra que es aplicable a los sistemas no apropiativos un algoritmo de asignación de prioridades óptimo que se describe en [1]. En general, el algoritmo de [1] es aplicable siempre y cuando el peor caso de tiempo de respuesta de un mensaje: a) no dependa de un ordenamiento de los mensajes de más alta prioridad, y b) no es de mayor tamaño si se le da una prioridad mayor.

La longitud de la cola de demoras y la longitud del periodo ocupado no dependen de un orden específico de los mensajes de alta prioridad. El término bloqueante B_m puede hacerse mayor al incrementar la prioridad, pero a su vez es contrarrestado por un decremento de la interferencia. Es posible encontrar un ordenamiento óptimo

siguiendo el algoritmo de dado en [1]:

```

Algoritmo de asignación óptima de prioridades( )
{
  Para cada nivel de prioridad, comenzando desde el más bajo
  {
    Para cada mensaje no asignado m
    {
      Si m es planificable con esta prioridad
      asignar a m esta prioridad
      else
        return El sistema no es planificable
    }
  }
  return El sistema es planificable
}

```

Para n mensajes, este algoritmo realiza como máximo $n(n-1)/2$ evaluaciones y garantiza una asignación de prioridades planificables si es que existe. Se debe tener en cuenta que el algoritmo no especifica un orden en el que los mensajes deberían ser analizados en cada nivel de prioridad. Este orden influye en gran medida en la asignación de la prioridad elegida si hay más de un orden planificable. Una mala elección de pedido inicial puede resultar en una asignación de prioridades que deje al sistema como planificable, sin ninguna seguridad de que se haya alcanzado un grado óptimo. Teniendo los datos presentados por la SAE y los modelos presentados, se pueden aplicar los cálculos propuestos por el análisis. Sin embargo, estos pasos pueden ser aplicados a cualquier conjunto de mensajes CAN siguiendo los siguientes pasos:

1. Obtención de las características de los mensajes: tamaño, identificador, velocidad del bus, etc.
2. Ordenar los mensajes por algún criterio, sea el de dado en [5] o [1].
3. Aplicar el cálculo propuesto a cada mensaje.

Para caso de la SAE se usará como guía la propuesta dada en [5], el método “D - J” en los que asigna mayor prioridad a los que tienen menor margen de espera (es decir, un valor de D - J más pequeño). Los valores de la Tabla 2 muestran el resultado.

Un punto importante a tener en cuenta es la periodicidad al momento de ordenar los mensajes. En este caso, el mensaje 14 es esporádico, y se puede dar como mínimo una vez cada 50ms, pero su deadline es de 5ms. Teniendo en cuenta esto, y que tal como lo indica la Tabla 2 es una señal correspondiente a la acción del freno, es natural que aparezca primero en la tabla. Para los mensajes periódicos, se tiene que el deadline es igual al tiempo del periodo, ya que de no ser atendido antes de llegar este punto el mensaje es sobrescrito inevitablemente. a partir del quinto mensaje analizado (el número de señal 8) el tiempo se dispara y no llegamos a transmitir el mensaje antes de alcanzar el deadline de 5ms. Para estos casos no nos quedan más opciones que subir la velocidad en la que operamos sobre el bus, o utilizar la técnica de “piggyback”.

Utilizar piggyback en CAN consiste en agrupar los mensajes que provienen del mismo origen, en un mensaje único. Por ejemplo, el subsistema que representa a la batería envía cuatro señales de un byte cada 100ms (las señales 1, 2, 4 y 6). Estas pueden ser representadas como un único mensaje de 4 bytes. Esto reduce la

utilización del bus ahorrando la sobrecarga de tres mensajes. La transmisión de un byte de datos puede llegar a requerir la transmisión de hasta 63 bits. Esta técnica se puede extender a las señales que poseen distintos periodos. Por ejemplo, las señales 29, 30 y 32 pueden ser transmitidas en un mensaje con un periodo de 5ms, mientras que dos de las señales tienen un periodo de 10ms (por cada segundo mensaje de la primera, se envían las dos señales de las restantes, cuando no son enviadas se puede enviar un valor nulo en su lugar).

Tabla 2. Cálculo Aplicado al Benchmark SAE.

Signal Number	Signal Description	Size/bits	J/ms	T/ms	Periodic / Sporadic	D/ms	From	Rm para 125Kbs
14	Hi&Lo Contactor Open/Close	4	0,1	50	S	5	Battery	1,776
9	Brake Pressure, Line	8	0,2	5	P	5	Brakes	2,816
49	Processed Motor Speed	8	0,2	5	P	5	I/M C	3,856
42	Torque Command	8	0,2	5	P	5	V/C	4,896
8	Brake Pressure, Master Cylinder	8	0,1	5	P	5	Brakes	5,936
7	Accelerator Position	8	0,1	5	P	5	Driver	11,136
43	Torque Measured	1	0,1	5	P	5	I/M C	17,880
11	Transaction Clutch Line Pressure	8	0,1	5	P	5	Trans	25,632
32	Clutch Pressure Control	8	0,1	5	P	5	V/C	36,536
29	High Contactor Control	8	0,3	10	P	10	V/C	65,088
30	Low Contactor Control	8	0,4	10	P	10	V/C	112,024
53	Main Contactor Acknowledge	1	1,5	50	S	20	V/C	207,744
48	Shift in Progress	1	1,4	50	S	20	V/C	378,352
46	Idle	1	1,3	50	S	20	V/C	686,120
44	FWD/REV	1	1,2	50	S	20	V/C	1250,728
40	Key Switch	1	1,1	50	S	20	V/C	2288,960
39	Warning Lights	1	1	50	S	20	V/C	4192,712
27	SOC Reset	1	0,9	50	S	20	Driver	7721,760
38	Backup Alarm	7	0,9	50	S	20	V/C	14297,048
37	Brake Solenoid	1	0,8	50	S	20	V/C	26734,536
52	Status/Malfunction (TBD)	8	0,8	50	S	20	I/M C	50253,808

Tabla 3. Análisis de Piggybacking.

Signal N°s	Size /bytes	J /ms	T /ms	D /ms	R (125Kb/s)
14	1	0.1	50.0	5.0	1.544
8,9	2	0.1	5.0	5.0	2.128
7	1	0.1	5.0	5.0	2.632
43,49	2	0.1	5.0	5.0	3.216
11	1	0.1	5.0	5.0	3.720
32,41	2	0.1	5.0	5.0	4.304
31,34,35,37,38,39,40,44,46,48,53	6	0.2	10.0	10.0	5.192
23,24,25,28	1	0.2	10.0	10.0	8.456
15,16,17,19,20,22,26,27	2	0.2	10.0	10.0	9.040
41,43,45,47,49,50,51,52	3	0.2	10.0	10.0	9.696
18	1	0.2	50.0	20.0	10.200
1,2,4,6	4	0.3	100.0	100.0	19.088
12	1	0.3	100.0	100.0	19.592
10	1	0.2	100.0	100.0	20.096
3,5,13	3	0.4	1000.0	1000.0	28.904
21	1	0.3	1000.0	1000.0	29.408
33,36	1	0.3	1000.0	1000.0	29.912

La técnica piggyback también se puede aplicar a señales que no son generadas al mismo tiempo (como puede ser el caso de señales esporádicas). Dichas señales pueden ser enviadas en un mensaje “servidor” donde la estación que envía el mensaje verifica constantemente la aparición de la señal esporádica antes de encolar el mensaje. Con este enfoque, una señal esporádica puede ser demorada por no más de un tiempo dado por el periodo de polling, más el peor tiempo de latencia del mensaje “servidor”. Si se tuviera que aplicar la técnica en un mensaje que tiene como deadline

Si bien el segmento presentado en la Fig. 3 corresponde a un lapso de tiempo muy reducido, permite verificar que en la red real construida el envío de los mensajes se da como en “escalera”. De esta manera, se corrobora la ecuación en la que se basa el análisis de tiempos de espera, basado en el valor conocido como “periodo de ventana”. Tal como se puede derivar de la ecuación, que utiliza los peores tiempos de los mensajes con mayor prioridad.

	A	B	C	D	E	F	G	H	I	J	K
1		0	1	2	3	4	5	6	7	8	9
2	0	2,576	0	0	0	0	0	0	0	0	0
3	1	2,579	0	0	0	0	0	0	0	0	0
4	2	2,584	0	0	0	0	0	0	0	0	0
5	3	0	2,66	0	0	0	0	0	0	0	0
6	4	0	0	5,248	0	0	0	0	0	0	0
7	5	0	0	0	7,84	0	0	0	0	0	0
8	6	0	0	0	0	10,432	0	0	0	0	0
9	7	0	0	0	0	0	13,27	0	0	0	0
10	8	0	0	0	0	0	0	15,564	0	0	0
11	9	0	0	0	0	0	0	0	18,163	0	0
12	10	0	0	0	0	0	0	0	0	0	20,752

Fig. 3. Tiempos de Esperas de los Mensajes Medidos en la Red CAN.

En la Fig. 4, se puede observar el gráfico resultado a partir de las métricas obtenidas a través del benchmark sobre la red real construida. En este gráfico se puede identificar la cadencia con la que los mensajes van llegando al controlador CAN, y cómo el tiempo de cada mensaje depende de los mensajes de mayor prioridad que se procesan con anterioridad.

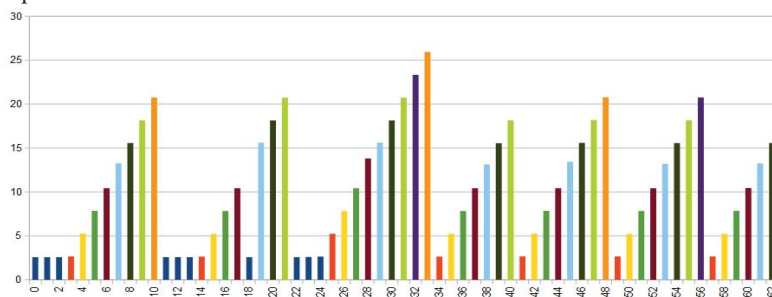


Fig. 4. Periodos Ocupados Medidos en la Red CAN.

Cada barra vertical de la Fig. 4 identifica un mensaje o instancia de mensaje en particular (el tipo de señal se identifica con un color), numerándolos desde 0 en adelante (sobre el eje x) y la altura de la cada barra es directamente el tiempo del mensaje desde que llega al controlador CAN hasta que es recibido en destino. Interpretando las primeras 10 barras de la Fig. 3 se puede identificar que:

- Los mensajes CAN 0 a 2 pertenecen a una misma señal 0, que tiene periodo de 30ms, por lo que el emisor puede enviar los 3 mensajes antes de que llegue la señal 1 que tiene un periodo de 100ms.
- Luego de esto, en un mismo instante (a los 100ms) llegan datos de las señales 1 a la 8. Estas entran todas al mismo tiempo, por lo que se tienen tiempos

incrementales de los mensajes 3 y 10 del gráfico, de acuerdo a la forma en que se envían los mensajes CAN teniendo en cuenta las prioridades.

- El envío de los mensajes que llegaron en el mismo instante suceden en menos de 30ms y luego llega un nuevo mensaje de la señal 0. Es importante notar que ninguno de los mensajes tuvo una latencia superior a la peor calculada anteriormente por medio de los valores de la tabla, por lo que se puede decir que se prueba el cálculo y puesta en marcha de este tipo de análisis de tiempos de manera satisfactoria.

5 Conclusiones y Trabajo Futuro

Se pudo llevar a cabo la corroboración experimental del análisis planteado en trabajos anteriores sobre los tiempos de latencia y respuesta en la transmisión de mensajes utilizando el protocolo CAN. Se ha explorado la representación matemática de cada uno de los fenómenos presentes en el bus buscando las herramientas necesarias para el análisis de un conjunto de señales de un sistema hipotético, tomando como base el benchmark SAE. Este enfoque fue y es utilizado por los fabricantes de automóviles y maquinaria a la hora de plantear sus sistemas basados en CAN, por lo que se deja presente en este trabajo un análisis que si bien es útil por sí mismo, a su vez también permite nuevas investigaciones y enfoques. Se implementó y validó en una prueba sobre un sistema real, construyendo no solamente el hardware sobre el cual llevar a cabo la experimentación sino incluyendo también la instrumentación necesaria para la recolección de los datos a contrastar con el modelo matemático.

Entre los trabajos futuros sin lugar a dudas se debe considerar un estudio cuidadoso para incrementar el ambiente de hardware de experimentación, principalmente en lo que se refiere a contar con una red CAN más cercana a las reales en cuanto a cantidad de controladores. De manera análoga se debe determinar cuidadosamente el ancho de banda del bus y la cantidad de fuentes de mensajes (tipos de señales involucradas) que cada controlador CAN debe administrar/manejar en el acceso al bus CAN.

Referencias

1. Audsley, N. C., "Optimal priority assignment and feasibility of static priority tasks with arbitrary start times", 1991, <http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.41.6673>.
2. Davis, R., Burns, A., Brill, R., Lukkien, J. J. "Controller Area Network (CAN) schedulability analysis: Refuted, revisited and revised", Springer, Real-Time Syst 35: 239, 2007. doi:10.1007/s11241-007-9012-7.
3. George, L., Rivierre, N., Spuri, M., "Preemptive and Non-Preemptive Real-Time Uni-Processor Scheduling", Rapport de recherche n°2966, INRIA Rocquencourt, 1996.
4. Lehoczky, J. P., "Fixed Priority Scheduling of Periodic Task Sets with Arbitrary Deadlines", Proceedings., 11th Real-Time Systems Symposium, pp. 201-209, 1990.
5. Tindell, K., Burns, A., Wellings, A. J., "Calculating controller area network (CAN) messages response times", Control Eng. Practice, Vol.3 No. 8, 00. 1163-1169, Elsevier Science Ltd., 1995.

Análisis de Performance y Consumo sobre Sistemas Embebidos Multinúcleo

Dell'Oso Matías¹, Medina Santiago¹, Romero Fernando¹, De Giusti Armando^{1,2}.

¹Instituto de Investigación en Informática LIDI (III-LIDI), Facultad de Informática, Universidad Nacional de La Plata, 50 y 120 2do piso, La Plata, Argentina.

²CONICET – Consejo Nacional de Investigaciones Científicas y Técnicas
{mdelloso, smedina, fromero, degiusti}@lidi.info.unlp.edu.ar

Resumen. El avance tecnológico de los últimos años y la cantidad de datos que actualmente un sistema debe procesar, les exigen a los sistemas de tiempo real actuales una mayor capacidad de cómputo. Para hacer frente a esta demanda, surgieron los procesadores multinúcleo orientados a aplicaciones embebidas. En el siguiente trabajo se estudiarán puntualmente los sistemas multinúcleos asimétricos, y se presentará un análisis comparativo entre dos procesadores pertenecientes a la familia Cortex-M. Los puntos de análisis incluyen los tiempos de latencia entre núcleos, utilizando distintas técnicas de comunicación y el consumo de energía.

Palabras clave: Sistemas Embebidos, Multinúcleo, Consumo de Energía, Sistemas Heterogéneos, Cortex-M, Tiempo Real, Proyecto CIAA.

1. Introducción

En su mayoría, los Sistemas de Tiempo Real [1][2][3] están basados en microcontroladores, pequeñas unidades de procesamiento dedicadas de bajo consumo y con gran cantidad de periféricos integrados para el control de sensores y actuadores.

Hoy en día, con el auge del Internet de las Cosas y el avance tecnológico, la cantidad de datos que un sistema debe procesar ha crecido exponencialmente en los últimos años [4]. Las tecnologías actuales les exigen a los sistemas en general y, especialmente a los sistemas de tiempo real, una mayor capacidad de procesamiento. Esto llevó a que, en enero de 2014, NXP¹ desarrollara el primer microcontrolador multinúcleo asimétrico basado en la familia Cortex-M [5][6][7]. El objetivo principal de este tipo de procesadores es contar con un procesador de alto rendimiento y un coprocesador con un poder de cómputo más reducido. La principal ventaja es poder reducir la carga de trabajo del núcleo principal. Por ejemplo, mientras el procesador de alto rendimiento se encuentra procesando datos, el coprocesador podría estar

¹ NXP (Next eXPerience) Semiconductors es una empresa fabricante de semiconductores que se creó en el año 2006 a partir de la división de semiconductores de la empresa Philips. En 2015, NXP adquirió la empresa estadounidense Freescale Semiconductor, uno de los mayores fabricantes de semiconductores de ese momento.

manejando la entrada y salida del sistema. Otra gran ventaja de este tipo de sistemas es el ahorro de energía, donde podría utilizarse el coprocesador para realizar los cálculos habituales y, eventualmente, despertar al procesador principal para realizar alguna acción compleja, como por ejemplo manejar la pila TCP/IP o hacer procesamiento de señales. De esta manera tendríamos un sistema de bajo consumo que, de ser requerido, puede realizar operaciones que requieran de un gran poder de cómputo.

Los enfoques a tener en cuenta en el análisis de los sistemas embebidos multinúcleo pueden ser varios: el manejo de la entrada/salida, la planificación de tareas entre núcleos, el manejo de memoria compartida, el pasaje de mensajes entre núcleos y, como principal punto, lograr la mejor relación consumo/rendimiento.

Con el objetivo de estudiar algunos de los enfoques mencionados anteriormente, en este trabajo se lleva a cabo una comparación de dos microcontroladores multinúcleo asimétricos desarrollados por NXP (ver Sección 2.1).

Este artículo está organizado de la siguiente manera: la sección 2 detalla el hardware utilizado para realizar las pruebas; en la sección 3 se hará mención al modelo de comunicación entre núcleos; la sección 4 muestra las metodologías utilizadas en los experimentos; la sección 5 explica las pruebas realizadas y los resultados obtenidos y, la sección 6, contiene las conclusiones y líneas de trabajo futuras.

2. Hardware utilizado

2.1. Microcontroladores

Con el fin de ubicar el trabajo en el marco del proyecto CIAA (Computadora Industrial Abierta Argentina [8]), las pruebas fueron realizadas sobre las placas de desarrollo EDU-CIAA-NXP [9] y picoCIAA [10] (ver Figura 2.1 y 2.2 respectivamente).



Figura 2.1: EDU-CIAA-NXP



Figura 2.2: picoCIAA

Ambas placas cuentan con microcontroladores multinúcleo asimétricos. La primera posee el procesador LPC4337, perteneciente a la familia LPC43xx, primera serie de microcontroladores multinúcleo basado en la familia Cortex-M que fue lanzada al mercado. La segunda está basada en el procesador LPC54102, perteneciente a la nueva línea de procesadores multinúcleo asimétricos desarrollada por NXP. Las

características principales de dichos microcontroladores se pueden observar en la Tabla 2.1.

Tabla 2.1: Características principales del microcontrolador LPC4337 y LPC54102

Características	LPC4337	LPC54102
Procesador de alto rendimiento	ARM Cortex-M4F	ARM Cortex-M4F
Procesador de bajo rendimiento	ARM Cortex-M0	ARM Cortex-M0+
Frecuencia de reloj	204 MHz (compartido)	100 MHz (compartido)
Soporte para instrucciones SIMD	Si	Si
Unidad de punto flotante por hardware	Si	Si
SRAM	136 kB	104 kB
Flash	1 MB	512 kB

Ambos microcontroladores tienen el mismo modelo en lo que respecta al núcleo de alto rendimiento (Cortex-M4F), pero difieren en el coprocesador. El Cortex-M0+ es el procesador más eficiente en lo referido a consumo energético de la familia Cortex [11]. Está construido sobre la base de su antecesor, el Cortex-M0, manteniendo el mismo repertorio de instrucciones y siendo compatible con las mismas herramientas de software (ver Tabla 2.2). Gracias a esto, el LPC54102 reduce considerablemente el consumo con respecto al LPC4337, convirtiéndolo en un procesador orientado a las aplicaciones donde el consumo es una variable central.

Tabla 2.2: Características principales de los Cortex-M4F, Cortex-M0 y Cortex-M0+

Características	Cortex-M4F	Cortex-M0	Cortex-M0+
Arquitectura	ARMv7E-M – Harvard	ARMv6-M – Von Neumann	ARMv6-M – Von Neumann
Repertorio de instrucciones	Thumb/Thumb-2	Thumb/Thumb-2 (subconjunto)	Thumb/Thumb-2 (subconjunto)
Pipeline	3 etapas	3 etapas	2 etapas
DSP	Si	No	No
FPU	Si	No	No
MPU	Opcional	No	Opcional

En cuanto a la eficiencia en el manejo del sistema multinúcleo, el LPC54102 también saca ventaja. Gracias a la experiencia obtenida con la serie LPC43xx, no sólo el hardware ha sido mejorado, sino que también las librerías de software brindadas por el fabricante simplifican en gran medida la comunicación entre ambos núcleos.

2.2. Sensor de corriente

Para realizar las pruebas de consumo se utilizó un módulo de bajo costo. En particular, se escogió el sensor de corriente INA219 [12] debido a su facilidad de uso y a que posee una resolución adecuada para realizar las mediciones.

A continuación, se detallan sus principales características:

- Voltaje de alimentación: 3 V – 5 V
- Rango de medición de voltaje: 26 V
- Rango de medición de corriente: $\pm 3,2$ A
- Resolución: 0,8 mA
- Resistencia de sensado: 0,1 ohm 1% 2W
- Protocolo de comunicación: I2C

3. Modelo de Comunicación entre núcleos (IPC)

Tanto en el microcontrolador LPC4337 como en el LPC54102, el Cortex-M4 funciona como procesador maestro, mientras que los Cortex-M0 y Cortex-M0+ funcionan como procesadores esclavos. A su vez, en ambos casos, los dos núcleos tienen acceso completo² al mapa de memoria. Luego de un *reset* del sistema, o al salir del modo de ahorro de energía *Deep power-down*, únicamente el procesador maestro comienza la ejecución de código, mientras que el coprocesador, o procesador esclavo, se mantiene en estado de *reset* hasta que el procesador principal lo despierte. Una vez que ambos procesadores se encuentran ejecutando sus respectivos programas, es posible intercambiar mensajes entre ellos mediante la lectura/escritura de datos en una zona de memoria a la que ambos núcleos tengan acceso. Si bien es posible que cada procesador trabaje por encuesta (polling) de esa zona de memoria para saber si el otro núcleo dejó un mensaje, una mejor aproximación sería utilizar las *interrupciones cruzadas* que cada procesador posee.

En el caso del LPC4337, la posición IRQ#1 del vector de interrupciones del procesador maestro corresponde a la solicitud de interrupción del procesador esclavo, y la posición IRQ#1 del vector de interrupciones del procesador esclavo corresponde a la solicitud de interrupción del procesador maestro. Tanto en el Cortex-M4 como en el Cortex-M0, la solicitud de interrupciones se genera usando la instrucción Assembler *SEV* (Send Event: Enviar Evento). Dicha instrucción genera un pulso de un clico de reloj en la línea TXEV (Transmit Event: Transmitir Evento) disponible en cada uno de los núcleos (ver Figura 3.1).

El LPC54102, en cambio, como se mencionó en el apartado 2.1, posee ciertas mejoras que hacen de la comunicación entre núcleos un proceso mucho más sencillo. Cada procesador puede producir hasta 32 interrupciones definidas por el usuario hacia

² A no ser que la Unidad de Protección de Memoria (MPU) de alguno de los procesadores (Cortex-M4 o Cortex-M0+) se haya configurado para limitarle el acceso a alguna región. El Cortex-M0 no dispone de MPU, por lo tanto, no existe forma de limitar su acceso al mapa de memoria del sistema.

el otro núcleo. Para simplificar este proceso, el LPC54102 define (vía software) una estructura de memoria llamada *Mailbox* (Buzón [14]), la cual cuenta con una serie de herramientas para manejar, de una manera simple y con librerías de alto nivel, la comunicación entre el procesador maestro y el esclavo. De esta forma, es posible enviar un mensaje y al mismo tiempo generar una interrupción en la posición IRQ#31 del vector de interrupciones del otro núcleo. A su vez, el LPC54102 posee un registro de exclusión mutua que le permite a los núcleos realizar un acceso seguro a la zona de memoria compartida.

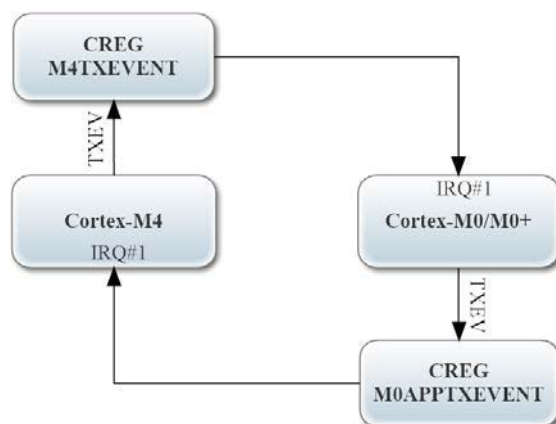


Figura 3.1: Esquema de interrupciones entre núcleos del microcontrolador LPC4337 [13].

De esta manera se evita que los núcleos tengan que encuestar permanentemente una zona de memoria. El procesador que requiera enviar un dato al otro, coloca el mensaje en la memoria compartida y genera un pulso de interrupción. El otro procesador podrá entonces, en la rutina de atención de interrupciones, acceder a la zona de memoria en la que se encuentra el dato y retirarlo.

4. Metodología

4.1. Latencia de comunicación entre núcleos

Para realizar las mediciones de tiempo, se utilizó el registro DWT_CYCCNT perteneciente a la unidad DWT (Data Watchpoint and Trace [15]) implementada en los Cortex-M4F de ambos microcontrolares. Dicho registro cuenta la cantidad de ciclos de ejecución del procesador.

Mediante dicho registro, se obtuvo la cantidad de ciclos que tardó cada prueba en llevarse a cabo y, mediante un cálculo adecuado, se convirtió esa cantidad de ciclos a unidades de tiempo. Este proceso puede observarse en el Listado 3.1.

Listado 3.1: Método para medir latencia entre núcleos.

```
*DWT_CTRL |= 1; // Activa el contador de ciclos
*DWT_CYCCNT = 0; // Resetea el contador de ciclos
/*
TODO: Comunicación entre núcleos
*/
ciclos = (float)(*DWT_CYCCNT); // Latencia en ciclos
TiempoUs = CyclesToUs(ciclos); // Latencia en Us
```

4.2. Consumo de energía

Para realizar las mediciones de consumo energético, se utilizó un método de medición de consumo invasivo, poniendo en serie a la alimentación de las placas de desarrollo el sensor INA219 mencionado en el inciso 2.2. Este sensor devuelve una magnitud analógica proporcional al flujo de corriente que ingresa a la placa que se está queriendo medir. Dicho valor de salida se capturó utilizando el convertidor analógico/digital de otro microcontrolador y que hizo las veces de monitor, mostrando un promedio del consumo durante un tiempo específico.

Cabe aclarar que, para esta primera instancia, no fueron desactivados completamente todos los periféricos de las placas de desarrollo, ni tampoco el procesador de depuración. Por dichas razones, los valores obtenidos no son meramente resultado del consumo de los procesadores.

5. Ensayos y Resultados

5.1. Latencia de comunicación entre núcleos

Con el fin de medir la latencia de comunicación entre núcleos utilizando diferentes métodos de comunicación, se plantearon los siguientes escenarios:

- Escenario A: tiempo que tarda el procesador principal en despertar al coprocesador.
- Escenario B: tiempo de ida y vuelta (RTT) entre núcleos utilizando interrupciones. Para generar las señales de interrupciones se utilizó directamente la instrucción Assembler *SEV* con el fin de reducir al mínimo el tiempo de comunicación. Cabe destacar que en este caso no se realizaron transferencias de datos por lo que no hubo accesos a memoria. Esta prueba sólo fue realizada para el LPC4337, debido a que el manejador de interrupción del LPC54102 no cuenta con ninguna posición que responda ante esta señal.

- Escenario C: este escenario es similar al escenario B, pero en vez de utilizar directamente la instrucción *SEV*, se utilizaron funciones de LPCOpen³ con el fin de evaluar el overhead que agrega utilizar dicha biblioteca.
- Escenario D: en este escenario se midió el tiempo de RTT entre núcleos utilizando la memoria compartida. La comunicación se realizó haciendo polling sobre una dirección de memoria hasta que el otro núcleo escriba un dato.
- Escenario E: este escenario intenta reflejar lo que sería un caso de comunicación real. Se midió el tiempo de RTT entre núcleos utilizando todas las funciones de LPCOpen y, a diferencia del escenario B, sí se realizaron transferencia de datos por lo que este tiempo también refleja varios accesos a memoria.

Tabla 5.1: Tiempos de latencia de las comunicaciones entre núcleos expresados en microsegundos (us).

Escenario	EDU-CIAA-NXP (204MHz)	picoCIAA (96MHz)
A	24,0409	11,0614
B	0,392156	-
C	3,1911	2,0937
D	0,2442	0,2916
E	6,1323	2,7943

La tabla 5.1 muestra que, en la mayoría de los escenarios planteados, con el procesador LPC54102 se obtuvieron mejores tiempos de respuesta, a pesar de estar trabajando a menor velocidad de reloj. Estos resultados responden a lo esperado debido a las mejoras (descriptas en las secciones 2.1 y 3) que posee este procesador frente a su antecesor.

5.2. Consumo de energía

Para medir el consumo energético de ambas placas de desarrollo, también se plantearon diferentes escenarios:

- Escenario A: ambos procesadores se encuentran en reposo.
- Escenario B: procesador maestro realizando procesamiento y procesador esclavo en reposo.
- Escenario C: procesador maestro en reposo y procesador esclavo realizando procesamiento.
- Escenario D: ambos procesadores realizando procesamiento.

³ LPCOpen es una extensa colección de bibliotecas de software (manejadores y middleware) y programas de ejemplo que permiten a los desarrolladores crear productos multifuncionales para la familia de microcontroladores LPC [16].

En las pruebas realizadas sobre la picoCIAA, se estableció la frecuencia de reloj del LPC54102 a su valor máximo, 100 MHz. En lo que respecta a la EDU-CIAA-NXP, las pruebas también se realizaron a su frecuencia de reloj máxima, 204 MHz y, luego, a 100 MHz con el objetivo de comparar los resultados con los obtenidos para la picoCIAA. Con el fin de constatar que los resultados obtenidos sean fehacientes, se tomó como referencia el trabajo realizado por uno de los ingenieros responsables del Proyecto CIAA [17].

Tabla 5.2: Consumo energético de las placas de desarrollo expresado en miliamperios (mA).

Escenario	EDU-CIAA-NXP (204 MHz)	EDU-CIAA-NXP (96 MHz)	picoCIAA (96 MHz)
A	278,16	187,51	20,01
B	299,19	197,71	27,05
C	291,78	194,06	23,9
D	312,25	204,2	31,1

La Tabla 5.2 muestra una amplia diferencia de consumo entre ambas placas de desarrollo. Uno de los motivos de estos resultados es la diferencia de tecnologías que hay entre placas, El LPC54102 tiene mejoras de hardware orientadas a disminuir el consumo. A su vez, otro factor que marca la diferencia es la cantidad de hardware extra que contiene la placa EDU-CIAA-NXP en su conjunto frente a la picoCIAA que, como se mencionó anteriormente, no fueron desactivados en su totalidad para este análisis.

6. Conclusiones y trabajo futuro

6.1. Conclusiones

Este trabajo muestra el funcionamiento de dos microcontroladores con procesadores multinúcleo asimétricos. Se realizaron distintos estudios sobre ambos con el fin de observar cómo respondía cada uno de ellos ante los mismos escenarios de prueba.

Teniendo en cuenta los tiempos de respuesta obtenidos en las pruebas de comunicación entre núcleos, se puede concluir que, dichos tiempos, en ambos microcontroladores, están dentro de los límites correctos y esperados para este tipo de sistemas [18]. Esto da la posibilidad de usar estos microcontroladores en sistemas de tiempo real, en los cuales se necesita tener un alto nivel de determinismo y en donde los tiempos de respuesta son críticos.

En lo que respecta a las pruebas de consumo energético, los resultados obtenidos para la EDU-CIAA-NXP, aunque estén dentro de los valores esperados, no satisfacen los requerimientos de una aplicación de bajo consumo. Sin embargo, las pruebas realizadas sobre la picoCIAA verifican las mejoras tecnológicas del procesador LPC54102, las cuales se ven reflejadas en un menor consumo ante mismos escenarios.

Para finalizar, podemos concluir que la utilización de microcontroladores multinúcleo en sistemas de tiempo real otorga muchas ventajas y amplía significativamente el panorama de aplicaciones a resolver con este tipo de sistemas. Sin embargo, esto conlleva un mayor esfuerzo en el desarrollo. Trabajar sobre este tipo de procesadores implica un nuevo paradigma de programación que no es el habitual en este tipo de sistemas. Para alcanzar un rendimiento adecuado, se deben tener en cuenta todas las condiciones que impone la programación de sistemas paralelos.

6.2. Trabajo futuro

Quedan como pendientes a este trabajo, la medición de consumo al aislar los distintos periféricos y el debugger de cada placa de desarrollo, lo que permitirá hacer foco solamente en el consumo de cada uno de los núcleos. A su vez, será de interés volver a realizar las pruebas utilizando los distintos modos de bajo consumo de ambos procesadores.

Otro de los principales puntos de estudio para el futuro es la utilización de sistemas operativos de tiempo real, estáticos o dinámicos, en microcontroladores multinúcleo. Esto implica el manejo de tareas para cada núcleo, la posibilidad de contar con uno o varios planificadores, sincronización, definición memoria compartida para crear tareas dinámicas, entre otras características importantes para el análisis.

Por último, es de interés el desarrollo de un algoritmo inteligente que pueda decidir en tiempo real qué núcleo debe ejecutar determinada operación, en función a la carga de procesamiento que se requiera. Con esto, se obtendría el mejor rendimiento del procesador, considerando tiempo de respuesta y consumo energético.

Bibliografía

- [1] Fan, X. *Real-Time Embedded Systems: Design Principles and Engineering Practices*. Newnes, 2015. ISBN – 13: 978-0128015070
- [2] Kopetz, H. *Real-Time Systems: Design Principles for Distributed Embedded Applications*. Springer, 2011. ISBN – 13: 978-1441982360
- [3] Laplante, P. *Real-Time Systems Design and Analysis: Tools for the Practitioner*. Wiley-IEEE Press, 2011. ISBN – 13: 978-0470768648
- [4] Raj, P., Raman, A. *The Internet of Things: Enabling Technologies, Platforms, and Use Cases*. Auerbach Publications, 2017. ISBN – 13: 978-1498761284
- [5] Yiu, J. *The Definitive Guide to ARM R Cortex R-M3 and Cortex R-M4 Processors*. Newnes, 2014. ISBN – 13: 978-0124080829
- [6] NXP (2017, Jul 1). *LPC Dual-Core microcontrollers based on ARM® Cortex® -M0 & M4F* [Online]. Disponible en: http://www.nxp.com/pages/lpc-dual-core-cortex-m0-m4f:MC_1423136028688
- [7] NXP (2017, Jul 2). *Product Longevity* [Online]. Disponible en: http://www.nxp.com/about/about-nxp/technology-programs/product-longevity:PRDCT_LONGEVITY_HM
- [8] Proyecto CIAA (2017, Jul 5). *Computadora Industrial Abierta Argentina* [Online]. Disponible en: <http://www.proyecto-ciaa.com.ar>

- [9] Proyecto CIAA (2017, Jul 5). *EDU-CIAA-NXP* [Online]. Disponible en: http://www.proyecto-ciaa.com.ar/index_comprar_educiaanxp.html
- [10] Proyecto CIAA (2017, Jul 5). *picoCIAA* [Online]. Disponible en: http://www.proyecto-ciaa.com.ar/index_comprar_picociaa.html
- [11] ARM (2017, Jul 16). Cortex-M0+ [Online]. Disponible en: <https://developer.arm.com/products/processors/cortex-m/cortex-m0-plus>
- [12] Texas Instruments (2017, Jul 7). *Hoja de datos del sensor de corriente INA219* [Online]. Disponible en: <https://cdn-shop.adafruit.com/datasheets/ina219.pdf>
- [13] *LPC43xx User Manual*, pp. 29, sec. 2.2.
- [14] *LPC5410x User Manual*. Chapter 29.
- [15] ARM (2017, Jul 16). *Cortex-M4 Technical Reference Manual*. Chapter 9. Disponible en: <http://infocenter.arm.com/help/index.jsp?topic=/com.arm.doc.ddi0439b/BABJFFGJ.html>
- [16] NXP (2017, Jul 18). *LPCOpen Libraries and Examples* [Online]. Disponible en: <http://www.nxp.com/products/microcontrollers-and-processors/arm-processors/lpc-cortex-m-mcus/software-tools/lpcopen-libraries-and-examples:LPC-OPEN-LIBRARIE>
- [17] Ridolfi, P. *Extensión del Sistema Operativo FreeOSEK para multiprocesadores asimétricos*. 2015.
- [18] Medina, S., Pi Puig, M., Dell'Oso, M., Romero, F., De Giusti, A., Tinetti, F.: *Comparación de Sistemas Operativos Embebidos sobre la Computadora Industrial Abierta Argentina*. In: Libro de Actas XXI Congreso Argentino de Ciencias de la Computación, pp. 1083 - 1092, 2016. <http://sedici.unlp.edu.ar/handle/10915/56385>

Control PID Robusto de Sistemas de Fase no Mínima Combinando QFT y GSA: Aplicación al Control Carga- Frecuencia

Patricia N. Baldini¹,

¹ Departamento de Electrónica, Facultad Regional Bahía Blanca,
Universidad Tecnológica Nacional, 11 de abril 461, Bahía Blanca, Argentina
pnbaldi@frbb.utn.edu.ar

Abstract. En este trabajo se presenta una metodología que automatiza la sintonización de un controlador robusto de estructura fija aplicable a sistemas de fase no mínima con parámetros variables. El diseño se realiza en el dominio de la frecuencia mediante un procedimiento de *loop shaping* que combina el algoritmo de optimización heurística de búsqueda gravitacional con la Teoría de Realimentación Cuantitativa. El procedimiento implica una optimización multiobjetivo, con restricciones, ya que se busca lograr un buen compromiso entre respuesta dinámica, estabilidad robusta y rechazo de perturbaciones con mínimo esfuerzo de control. El controlador adoptado es del tipo proporcional-integrativo-derivativo (PID) por su difundido uso industrial. El algoritmo seleccionado resulta de fácil comprensión e implementación, con buena eficiencia computacional y características de convergencia estable. La eficacia del método propuesto se comprueba mediante simulación para la regulación de frecuencia de un área de sistema de generación de potencia hidroeléctrica.

Keywords: Optimización Heurística, GSA, Control PID, Sistemas de fase no-mínima, control robusto, QFT.

1 Introduction

Los sistemas denominados de fase no mínima (FNM) por la característica dominante de su respuesta en frecuencia, se modelan en el dominio de Laplace mediante Funciones Tránsito que presentan ceros en el semiplano derecho. Estos ceros originan limitaciones en las características de la dinámica alcanzable con el control de la planta e imponen compromisos de diseño que son independientes del tipo de controlador de orden entero adoptado. La complejidad se incrementa cuando los parámetros del sistema son no estacionarios o inciertos. Como consecuencia, resulta un desafío lograr un diseño que consiga el mejor desempeño teniendo en cuenta las restricciones estructurales.

El objetivo del control es el de cumplir en forma simultánea con un conjunto de especificaciones de comportamiento temporal, cualquiera sea el valor de los parámetros inciertos variables dentro de un rango finito que se incluyen en el modelo como incertidumbre paramétrica.

En estos casos, por la complejidad del problema que resulta altamente no lineal con un espacio factible de soluciones no convexo, las técnicas de optimización tradicionales no resultan aplicables para la sintonización eficiente de un controlador predeterminado. Una alternativa que ha demostrado ser efectiva y de uso cada vez más frecuente, la representan los métodos heurísticos de optimización, de búsqueda inteligente.

En este trabajo se propone la utilización del algoritmo estocástico, inspirado en la física, conocido como búsqueda gravitacional (GSA) para la determinación automática de los parámetros del controlador para SFNM. La función objetivo a minimizar y las restricciones de estabilidad se plantean de acuerdo a la Teoría de Realimentación Cuantitativa (QFT) para garantizar un diseño robusto [1],[2]. El objetivo que propone QFT es la síntesis de un controlador lo más simple posible, que satisfaga las especificaciones con el menor costo de realimentación y teniendo en cuenta la incertidumbre del modelo. Mediante el algoritmo de GSA se automatiza la etapa denominada *loop shaping*, es decir, la determinación sin sobrediseño de los parámetros del controlador en el dominio de la frecuencia. Esto permite lograr las mayores prestaciones del controlador sin aumentar su complejidad.

El algoritmo GSA [3] ha sido elegido debido a la simplicidad para su comprensión e implementación y la característica de crecimiento lineal del tiempo de procesamiento con respecto a las dimensiones del problema. Por otro lado, la convergencia ha sido comprobada teóricamente [4]-[6] y se ha verificado una buena capacidad de exploración del espacio de búsqueda y de encontrar soluciones óptimas [7],[8] contando con la ventaja adicional de un reducido número de parámetros de ajuste. Por otra parte, numerosos estudios experimentales sugieren que permite conseguir iguales o mejores resultados en términos de tiempo de procesamiento, calidad de la solución, recursos de cómputo y convergencia, que otros algoritmos de igual filosofía [9]-[11].

La metodología QFT fue seleccionada siguiendo el criterio de mantener la claridad, simplicidad y versatilidad del procedimiento de diseño. El método de diseño reportado por algunos autores, basado en la resolución de un problema equivalente de fase mínima [12], [13], se considera una complicación de cálculo actualmente innecesaria debido a que las restricciones impuestas por los sistemas de FNM son estructurales y a la efectividad de los algoritmos heurísticos disponibles.

Los controladores con acción proporcional-integrativa-derivativa (PID) son los de mayor aplicación industrial. Esto se debe a que, a pesar de su simplicidad, permiten resolver satisfactoriamente una gran variedad de problemas complejos cuando está sintonizado apropiadamente [14].

El método presentado se aplica al diseño del regulador de frecuencia para un área de un sistema de generación de potencia hidroeléctrica. El comportamiento de estos sistemas es inherentemente de FNM por efecto de la inercia del agua. El objetivo final, desde el punto de vista de la teoría de control, es lograr en cada área el balance entre la potencia generada y la demanda, de modo de mantener la frecuencia y el intercambio de potencia entre áreas en sus valores nominales. La dinámica no-lineal, no-estacionaria y las variaciones permanentes y aleatorias características de la carga hacen que el desarrollo de técnicas de control eficientes sea todavía un problema de interés para garantizar la calidad del servicio y la protección del sistema. Los resultados obtenidos y la robustez se comprueban mediante simulación.

2 Descripción del Problema

El diseño de sistemas de fase no mínima es siempre una tarea compleja debido a que este tipo de plantas imponen limitaciones estructurales en el diseño. Las características dinámicas alcanzables quedan restringidas por la naturaleza de la planta, independientemente del tipo de controlador a usar. Estas restricciones estructurales se manifiestan en la respuesta temporal a entradas de tipo escalón por un comportamiento inicial típico en sentido opuesto al de la entrada. El consecuente *undershoot* se caracteriza por una amplitud que varía en forma inversa al tiempo de respuesta. Para reducir el *undershoot* la acción de control deberá ser tanto más lenta cuanto más significativo sea el comportamiento de fase no mínima (ceros positivos de menor valor). Por otro lado, en el dominio de la frecuencia se ven reducidos tanto el margen de ganancia como el ancho de banda posibles para el sistema. De modo que resulta un desafío lograr un diseño que permita el mejor comportamiento dinámico sobre todo cuando se añade la complejidad de parámetros variables con las condiciones de operación [12].

Estas variaciones son incorporadas en el modelo como incertidumbre paramétrica, de modo que determinan una familia de FTs. El diseño se resuelve automatizando el procedimiento de *loop shaping* sobre la FT nominal, como un problema de optimización basado en GSA, donde la función objetivo y las restricciones de estabilidad se plantean en el marco de QFT para garantizar robustez.

2.1 Conceptos Básicos sobre GSA

El algoritmo de búsqueda gravitacional es un método estocástico de búsqueda inteligente basado en las leyes físicas de gravedad y movimiento de Newton, introducido por Rashedi *et al.* [3]. Se trata de realizar una búsqueda eficiente en el espacio D -dimensional de soluciones factibles, mediante partículas de masa variable cuyas posiciones representan un conjunto posible de parámetros del controlador que se desea sintonizar. Las partículas se mueven por efecto de fuerzas de atracción de tipo gravitacionales. Los desplazamientos se producen en direcciones que tienen en cuenta la mejor y la peor solución global en cada iteración, en el intento de encontrar la óptima. La masa de cada agente varía según el grado de éxito obtenido en la búsqueda, de modo que las fuerzas provocan un movimiento del conjunto hacia las partículas con mayor suceso y, por lo tanto, de mayores masas. Para lograr un buen balance entre exploración y explotación solo el subgrupo de N_{best} agentes con mayor masa ejercen fuerzas sobre los otros. Las masas inerciales y gravitacionales se presuponen iguales.

Las expresiones de actualización de posiciones $\mathbf{X}_k^i = [x_k^{i1} \ x_k^{i2} \ \dots \ x_k^{in} \ \dots \ x_k^{iD}]^T$, y velocidades $\mathbf{V}_k^i = [v_k^{i1} \ v_k^{i2} \ \dots \ v_k^{in} \ \dots \ v_k^{iD}]^T$ de cada partícula i de las N de la población, en la k -ésima iteración, son las siguientes:

$$\mathbf{V}_{k+1}^i = R_l \mathbf{V}_k^i + \mathbf{a}_k^i, \quad (1)$$

$$\mathbf{X}_{k+1}^i = \mathbf{X}_k^i + \mathbf{V}_k^i. \quad (2)$$

Fuerzas \mathbf{F}^i , masas M^i y aceleraciones se calculan en cada iteración según :

$$M_k^i = \frac{1}{\sum_{j=1}^N m_k^j} \left(\frac{f_k^i(\mathbf{X}_k^i) - f_{worst\ k}}{f_{best\ k} - f_{worst\ k}} \right), \quad (3)$$

$$\mathbf{F}_k^i = \sum_{\substack{j \in N_{best} \\ j \neq i}}^N R_2 G_k M_k^j M_k^i \frac{(\mathbf{X}_k^j - \mathbf{X}_k^i)}{\|\mathbf{X}_k^j - \mathbf{X}_k^i\|_2 + \varepsilon}, \quad (4)$$

$$\mathbf{d}_k^i = \frac{\mathbf{F}_k^i}{M_k^i}. \quad (5)$$

donde: $1 \leq i \leq N$; R_1, R_2 son factores de peso aleatorios con distribución uniforme en el intervalo $[0,1]$, $G_k = G_0 e^{-(\beta k/k_{max})}$ es la gravedad, G_0 y β son parámetros de ajuste del método, $\varepsilon \ll 1$ es un factor de seguridad, f es el valor de la función objetivo correspondiente, $f_{best\ k} = \min_{1 \leq j \leq N} f(\mathbf{X}_k^j)$ y $f_{worst\ k} = \max_{1 \leq j \leq N} f(\mathbf{X}_k^j)$ representan el mejor y el peor valor de la función objetivo del conjunto.

3 Optimización Multiobjetivo Mediante GSA en el Marco de QFT

La estructura del control adoptada es la de realimentación con dos posibles grados de libertad, según la figura.

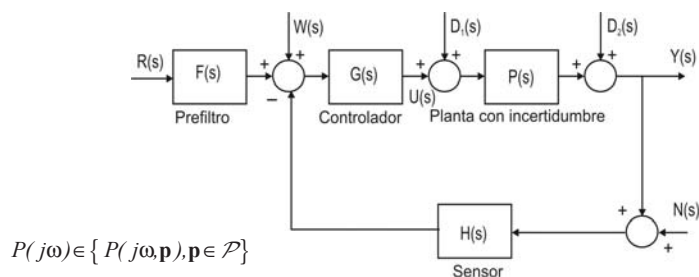


Fig. 1: Esquemas del sistema de control

En esta configuración, basada en el modelo complejo de FT, R es la señal de referencia, Y es la variable controlada, $P(s)$ y $H(s)$ representan la planta y la dinámica del sensor, respectivamente. Los procesos dinámicos W , D_1 y D_2 modelan perturbaciones y N representa el ruido de sensado.

La estructura general del controlador a considerar incluye términos proporcional, integrativo y derivativo, PID, al que se le incorpora un polo de alta frecuencia para evitar amplificación de ruido. Su FT resulta:

$$G_{PID}(s) = K_p + \frac{K_i}{s} + \frac{s K_d}{s \frac{K_d}{T} + 1} \quad (6)$$

El problema de optimización en el marco de QFT provee una técnica eficiente de diseño robusto con objetivos múltiples en el dominio de la frecuencia. Los parámetros de la planta $P(s)$ que no pueden determinarse en forma precisa definen, dentro del espacio de incertidumbre, una región que la cuantifica, Θ , y una familia de FTs asociada. Para la planta, su *template* es definido como el conjunto de respuestas en frecuencia posibles asociadas al subespacio Θ , $\{P(j\omega, \theta), \theta \in \Theta\}$. Una FT de la familia, $P_\theta(j\omega)$, se adopta como modelo nominal y se utiliza como base para el diseño. Las especificaciones de comportamiento robusto para el rango de frecuencias de operación, se expresan analíticamente mediante restricciones sobre el módulo de las FTs que relacionan las distintas variables del lazo [1],[2]. Estas restricciones se representan como regiones admisibles para la función de lazo abierto nominal buscada, $L_\theta(j\omega) = G(j\omega) P_\theta(j\omega) H_\theta(j\omega)$, en un plano módulo-fase conocido como carta de Nichols. Los límites de estas regiones, cada frecuencia de trabajo ($\omega \in \Omega$), quedan definidos por un conjunto de curvas denominadas *bounds* que incluyen la incertidumbre.

El diseño consiste en determinar un controlador, $G(j\omega)$, tal que $L_\theta(j\omega)$ satisfaga las restricciones establecidas por los *bounds* lo más ajustadamente posible para evitar el sobrediseño, manteniendo la ganancia de alta frecuencia lo menor posible. Si esto se logra, se garantiza que también lo harán todas las funciones correspondientes al *template*. Esto implica un proceso de optimización en un espacio no convexo, con objetivos múltiples y restricciones, con más de una solución posible. En este proceso, los *bounds* se calculan al inicio y se mantienen constantes durante las iteraciones.

Tabla 1. Especificaciones en el Diseño Robusto con QFT.

Especificación	Restricción
Estabilidad Robusta	$\left \frac{L(j\omega+\theta)}{1+L(j\omega+\theta)} \right \leq \delta_1, \omega \in \Omega_1$
Rechazo de perturbaciones D_1	$\left \frac{P(j\omega+\theta)}{1+L(j\omega+\theta)} \right \leq \delta_2(\omega), \omega \in \Omega_2$
Restricción sobre el esfuerzo de control	$\left \frac{G(j\omega)}{1+L(j\omega+\theta)} \right \leq \delta_M(\omega), \omega \in \Omega_3$

La dimensión del espacio paramétrico es $D = 4$ y la posición de la partícula i -ésima, $\mathbf{X}_i = [K \ T_i \ T_d \ T]^T$, es un vector de parámetros posibles. La función a minimizar o función objetivo incluye propiedades de estabilidad robusta y rechazo de perturbaciones y reducción del costo de realimentación, según la Tabla 1. Su expresión resulta:

$$f(\mathbf{X}) = \gamma_1 20 \log(K_{HF}) + \gamma_2 \sum_{k=1}^{n_f} f_{bnd}(\omega_k) + \gamma_3 \sum_{k=n_{LIM}}^{n_f} f_{UHF}(\omega_k) \tag{7}$$

$$f_{bnd}(\omega_k) = d_k \quad , \quad k = 1, 2, \dots, n_f \quad , \quad f_{UHF}(\omega_k) = d_{Lk} \delta \quad ,$$

$$k = n_{lim}, n_{lim} + 1, \dots, n_f \quad \delta = \begin{cases} 0 & \text{si se cumple con la condición} \\ 1 & \text{si no se cumple} \end{cases}$$

donde d_k y d_{Lk} denotan las distancias entre la función transferencia de lazo abierto nominal, $L_0(j\omega_k)$, y el *bound* correspondiente a ω_k , y entre esa función y el denominado *bound* de alta para frecuencia universal (UHB), para $\omega_k \geq \omega_{Lim}$ ($k \geq k_{lim}$), respectivamente. Esta condición garantiza un buen comportamiento a alta frecuencia. K_{HF} es la ganancia de alta frecuencia o costo de realimentación. Para una diferencia r entre polos y ceros de $L(s)$, esta ganancia se define como

$$\lim_{\omega \rightarrow \infty} |L(j\omega)| \approx \lim_{s \rightarrow \infty} \frac{K_{HF}}{\omega^r}. \quad (8)$$

Las constantes γ_i son factores de peso y n_f es el número de frecuencias consideradas dentro del rango de trabajo obtenido a partir de la respuesta en frecuencia del sistema. Se incluyen una condición de módulo decreciente y una restricción, g_1 , que garantiza la estabilidad durante el proceso. Teniendo en cuenta que un sistema lineal e invariante en el tiempo es estable si las raíces de su polinomio característico ($L_0(p_i) + 1 = 0$), son reales y negativas

$$g_1(\mathbf{X}) = \max_{1 \leq i \leq m+2-r} (Re(p_i), 0) \leq 0 \quad (9)$$

El tratamiento de las restricciones y los límites de búsqueda se realiza mediante el método de penalización y el criterio para la convergencia tiene en cuenta la invarianza del mejor valor de la función objetivo, dentro de la tolerancia considerada, durante un número determinado de iteraciones.

4 Regulación de Frecuencia en Sistemas de Potencia

El control de frecuencia es un mecanismo fundamental de los sistemas de generación de potencia eléctrica que contribuye a mantener el balance entre la potencia generada y la demanda en cada área de control. En estos sistemas se debe asegurar que la frecuencia y el intercambio de potencia eléctrica entre áreas se mantengan dentro de límites aceptables, de forma tal que los consumos abastecidos tengan una buena calidad de suministro eléctrico. Esto se ha resuelto normalmente mediante la aplicación de técnicas tradicionales de control, a menudo eficaces, pero no necesariamente eficientes [15]. Considerando el desarrollo de potentes técnicas de control de sistemas y el gran avance de la electrónica de los últimos años, resulta interesante investigar estrategias que maximicen la calidad del control [16]-[19].

En particular, en el caso de potencia hidroeléctrica, el sistema resulta de fase no mínima pues, por efecto de la inercia del agua: una variación en la apertura de la compuerta produce un efecto inicial contrario en la presión a través de la turbina y, en consecuencia, en la potencia mecánica, antes de conseguirse el resultado esperado.

4.1 Modelo Dinámico y Especificaciones

El modelo del sistema de potencia para estudios de regulación carga-frecuencia está integrado por el sistema gobernador, la turbina, la carga y el generador, como se

muestra esquemáticamente en la Fig. 2.

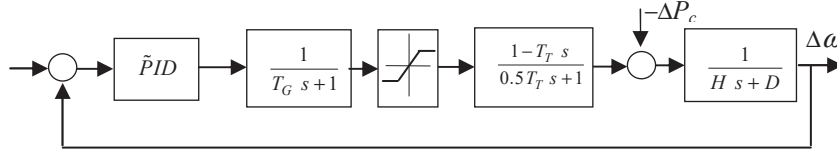


Fig. 2 : Diagrama en bloques con FT de cada componente y controlador

El objetivo desde el punto de vista de control es modificar la posición de la compuerta para regular el flujo de agua a presión a través de la turbina, y por lo tanto controlar la potencia del generador, independientemente de las variaciones de carga que resultan aleatorias e impredecibles. Estas variaciones naturales presentan valores generalmente pequeños en comparación con la capacidad del sistema, de modo que es razonable el uso de modelos linealizados alrededor del punto de operación para diseñar el sistema de control. Las limitaciones físicas tales como restricciones en la tasa de generación (GRC) deben ser tenidas en cuenta para evitar inestabilidad. Los cambios bruscos en la demanda de potencia se incorporan al modelo como perturbaciones de tipo escalón.

Siguiendo la referencia [17] se adoptó el modelo de turbina linealizado despreciando la fricción hidráulica. Como se trata de un sistema de regulación, se consideraron especificaciones de estabilidad robusta, rechazo de perturbaciones a la entrada del generador, costo de realimentación y características de módulo decreciente. La constante de *droop* o estatismo, $R=20$, se incluye en el controlador, de modo que $K_{p\text{final}} = K_p - 1/R$.

Según la Tabla 1, con $H_o(j\omega) = 1$ y $M(j\omega\theta)$ la FT del grupo generador-carga:

$$\Omega_1 = \{\omega: 0.01 \leq \omega \leq 20\}, \quad \Omega_2 = \{\omega \in \Omega_1 : \omega \geq \omega_{\text{LIM}} = 5\}, \quad (10)$$

$$\left| \frac{L(j\omega\theta)}{1+L(j\omega\theta)} \right| < 1.2, \quad \left| \frac{\Delta\omega}{\Delta P_m} \right| = \left| \frac{M(j\omega\theta)}{1+L(j\omega\theta)} \right| < 0.4 \left| \frac{j\omega}{(j\omega)^2 + 0.4j\omega + 0.035} \right|.$$

Los parámetros del modelo son: la constante de tiempo del agua asociada a la turbina, T_T , el factor de amortiguamiento y la inercia del conjunto generador-carga, D y H , respectivamente, y la constante de tiempo del gobernador, T_G . Los rangos de variación, en unidades relativas, p.u., fueron: $2 \leq T_T \leq 4$, $2.5 \leq H \leq 6$, $1 \leq D \leq 1.5$, $T_G = 0.2$, [17]-[19].

Los parámetros de GSA adoptados fueron [8]: tamaño de la población, $N=20$, $\alpha=20$, $G_0=100$. El número de iteraciones promedio para la convergencia resultó de 50 para una tolerancia de 10^{-5} .

Los parámetros del controlador resultantes fueron $PID = [K_p \ K_i \ K_d \ T]^T = [0.889 \ 0.062 \ 0.989 \ 98.9]^T$; $K_{p\text{final}} = -18.892$ para $R=0.05$; los márgenes de fase y ganancia y la frecuencia de cruce: $MF=90.6^\circ$, $MG=11.4$ dB, $\omega_0=0.085$ rad/seg.

En la Fig. 3 (a) los *templates* muestran el grado de dispersión en las respuestas en frecuencia de lazo abierto nominal producto de la incertidumbre, y en (b) la respuesta en frecuencia sin el controlador y después del diseño, sobre el conjunto de *bounds*

mostrando la eficiencia del *loop shaping*.

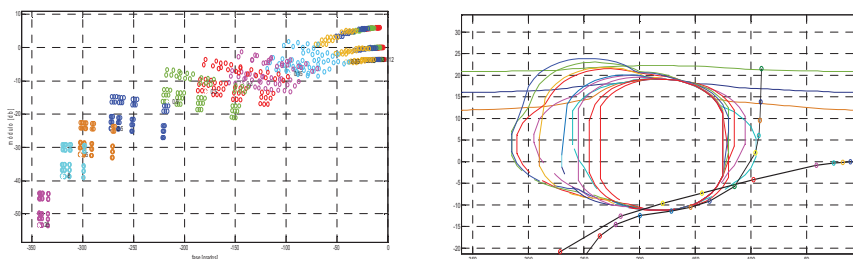


Fig. 3: a) *Templates* b) *Bounds y Respuesta en frecuencia nominal*

En las Fig. 4 y 5 se observan las respuestas temporales producida por una variación de la carga $\Delta P_C = 0.02$ en p.u. de la frecuencia $\Delta\omega$, de la salida del gobernador y de la señal de control u . Se muestran los resultados para los peores casos de operación o casos extremos, y para el considerado como nominal.

Puede verse que se cumplen las especificaciones, manteniendo valores bajos de la energía de control y evitando problemas con la tasa de generación (GRC).

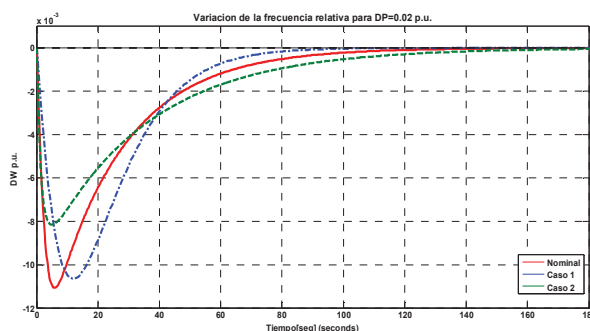


Fig. 4: Variación temporal de la frecuencia angular relativa para los peores casos y el nominal

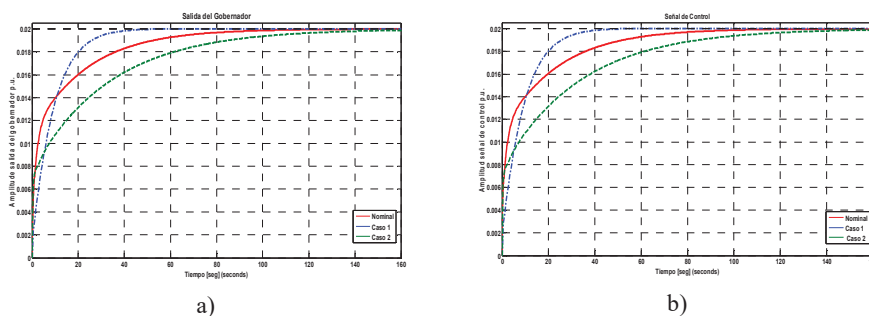


Fig. 5: Variaciones temporales para los peores casos y el nominal de a) la salida del gobernador b) la señal de control.

5 Conclusiones

En este trabajo se propuso una metodología robusta de diseño de controladores PID para sistemas de fase no mínima con parámetros inciertos. Se integran la Teoría de Realimentación Cuantitativa con el algoritmo de optimización heurística GSA para automatizar la sintonización del controlador. Esta combinación permite obtener las mejores prestaciones del controlador adoptado, cumpliendo las especificaciones de diseño aún en los peores casos determinados por la incertidumbre, sin necesidad de modificar la estructura *online*. El algoritmo de optimización heurística converge a una de las posibles soluciones óptimas con una dispersión aceptable medida por la varianza del valor de convergencia del vector de parámetros. La aplicación a un esquema de regulación carga-frecuencia de sistemas de potencia hidroeléctrica condujo a buenos resultados frente a las variaciones de carga y los consecuentes cambios de las condiciones de operación. Se verifica la flexibilidad de los controladores PID para cumplir especificaciones múltiples en sistemas complejos. El comportamiento fue verificado por simulación. La comparación con resultados publicados [17]-[19] permite confirmar la efectividad de la metodología propuesta que puede fácilmente ser extendida a otro tipo de controladores o de especificaciones. Actualmente se está trabajando en la aplicación de esta metodología para el control descentralizado de áreas interconectadas de sistemas de generación de potencia.

References

1. Houpis, H., Rasmussen, S., García Sanz, M.: Quantitative Feedback Theory: Fundamentals and Applications. 2da. Ed, CRC Press, Florida (2006)
2. García Sanz, M.: Quantitative Robust Control Engineering: Theory and Applications. RTO-En-SCI- vol.166, pp. 1-44 (2006)
3. Rashedi, E.; Nezamabadi-pour, H.; Saryazdi, S.: GSA: A Gravitational Search Algorithm. Information Sciences, vol. 179, pp. 2232-2248 (2009)
4. Ghorbani F., Nezamabadi, H., On the Convergence Analysis of Gravitational Search Algorithm, Journal of Advances in Computer Research, vol.3, no.2, pp. 45-51,(2012)
5. Tian, D.P.: A Review of Convergence Analysis of Particle Swarm Optimization. Int. Jou. Grid and Distributed Computing, vol.6, no.6, pp.117-128 (2013)
6. Sabir, N.M.; Puteh, M.; Mahmood, M. R.: A Review of Gravitational Search Algorithm. Int. J. Advance. Soft Comput. Appl., vol. 5, no. 3, pp. 1-38 (2013).
7. Kumar, Y., Sahoo, G.: A Review on Gravitational Search Algorithm and its Applications to Data Clustering & Classification, Int. Jou. on Intelligent Systems and Applications, vol. 6, pp.79-93, (2014)
8. Eldos, T., Al Qasim, R.: On The Performance of the Gravitational Search Algorithm (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 4, No. 8, (2013)
9. Wang, J-S., Song, J-D.: Function Optimization and Parameter Performance Analysis Based on Gravitation Search Algorithm, Algorithms, vol. 9, 3 (2016)
10. Can, U., Alatas, B.: Physics Based Metaheuristic Algorithms for Global Optimization American Journal of Information Science and Computer Engineering, vol. 1, No. 3, pp. 94-106, (2015)

- 11.Chen, W.H.y Balance, D.J.: QFT Design for Uncertain Non-minimum Phase and Unstable Plants Revisited. *Int. Journal on Control*, vol. 4, no. 9 (2001)
- 12.Khodabakhshian, A., Rahimi, H., Golbon, N.: QFT for Load Frequency Control of Non-Minimum Phase Hydro Power Plant. En: *proceeding of Int. Conference on Control Applications*, Munich, IEEE Press. (2006)
- 13.Ämtrom, K.; Hågglund, T.: *Control PID Avanzado*. Prentice Hall, Madrid (2009)
- 14.Kundur, P.: *Power Systems Stability and Control*. Cap. 9.1 y 11.1, McGraw Hill, NY(1994)
- 15.Hussein, T.: A Genetic Algorithm for Optimum Design of PID Controller in Load Frequency Control. *W. Academy of Science, Eng. and Tech.*, vol.70, pp. 956-960 (2012)
- 16.Kishor, N., Singh, S.P. y Raghuvanshi, A.S.: Dynamic simulations of hydro turbine and its state estimation based LQ control. En: *Conversion and Manag.*, vol.47, pp.3119-3137,(2006)
- 17.Khodabakhshian, A., Golbon, N.: Robust Load Frequency Controller Design for Hydro Power Systems. En: *Proc. Int. Conf. on Control Application*, Toronto, Canada, IEEE Press (2005)
- 18.Meng, L., Diao, F.: QFT Fractional Order Controller for Non-minimum Phase Hydro Power Plant. En: *Proc. of the 31st Chinese Control Conference*, Hefei, China, pp.2743-2748 (2012)
- 19.Satpati, B., Bandyopadhyay, I., Das, G., Koley, C.: Robust controller design for load frequency control of non-minimum phase Hydro power plant using PSO enabled automated Quantitative Feedback Theory. En: *Proc. India Conference. INDICON 2008*. IEEE Press, Kanpur, pp. 349 - 354 (2008)

Excluding Ionospherically Unsafe Satellite Geometries in GBAS CAT-I

Oscar Bria, Javier Giacomantone, and Luciano Lorenti

Research Institute in Computer Science (III-LIDI) - School of Computer Science
National University of La Plata - Argentina
`onb@info.unlp.edu.ar`

Abstract. We show the results of the implementation of a preliminary algorithm for excluding ionospherically unsafe satellite geometries in Ground-Based Augmentation Systems Category I. Minimum knowledge of the ionospheric treat model is assumed and the assistance of the code-carrier divergence monitor is not considered. All the satellites in view above 5° in elevation are included in the computations. The inflation of the standard deviation of the vertical ionospheric gradient implements the exclusion. Full availability remains for a typical day in the site of La Plata Airport.

Keywords: GBAS Category I, Ionospheric Threat, Parameter Inflation.

1 Introduction to GBAS

Ground-Based Augmentation System (GBAS) is a system that provides differential corrections and integrity monitoring of Global Navigation Satellite Systems (GNSS) signals for navigation and precision approach service in the vicinity of the host airport. GBAS yields the accuracy, integrity, continuity and availability necessary for Category I, and eventually Category II, and III precision approach operations [1, 2].

The system consists of three parts: 1) the satellite signal in space (SIS) coming from one or several GNSS constellations; 2) the local Ground Facility (GGF) equipped normally with four satellite receivers, processing equipment and a VHF data broadcast (VDB) transmitter; and 3) the aircraft devices related to the multi-mode receiver (MMR).

The GGF provides the aircraft with approach path data and, for each satellite in view, geo-referenced range corrections and integrity information. The corrections enable the aircraft to determine its position relative to the approach path more accurately. Under nominal conditions these corrections are considered practically the same for the ground station and for the aircraft. That is, the basic assumption of GBAS is that errors have a high local correlation.

Integrity is the function of a system that warns users in a timely manner when the system or some part of it should not be used. The integrity function prevents severe risks, eventually affecting system continuity and availability [3]. In GBAS Category I (CAT-I) the responsibility for integrity resides exclusively

in the GGF. Verification and certification of GBAS integrity is based on the analysis of extremely rare events which can lead to large positioning errors. These events are generally only dealt theoretically or by simulation [4, 5]. The most important of these rare effects are strong ionospheric gradients, which are not directly recognized by the GGF by any intended means. Hence the importance of being able to foresee and mitigate the effects of this anomalies in some way [6]. Excluding unsafe geometries is one of such ways; ionospheric field monitor is other alternative [7].

2 Protection Levels, Alert Limits and Position Errors

As mentioned, for GBAS CAT-I the GGF is designed to guarantee the integrity of each broadcasted correction by monitoring, in diverse ways, each related satellite measure to ensure that the correction error is bounded. If that is not the case the satellite is declared unsafe. The nominal correction error standard deviation for each satellite $\sigma_{\text{pr_gnd},i}$ [m] as seen from ground, and the nominal standard deviation of the vertical ionospheric gradient σ_{vig} [m/m] coming from site studies, are also broadcasted.

Based on the information received from the GGF and also based in own data, the aircraft computes horizontal and vertical protection levels (HPL and VPL) for the measurements for each epoch (every 0.5 second). Particularly VPL are safe if they are bellow the vertical alert limit (VAL). For GBAS CAT-I precision approach operations VAL = 10 m at the minimum decision height of 200 ft that occurs at 6 km from the runway during landing.

If VPL is larger than VAL an integrity alert is generated. VPL is supposed to exceed the real unknown vertical position error (VPE). If VPE is larger than VPL but smaller than VAL the information is misleading. If VPE is larger than VPL and also larger than VAL the information is hazardously misleading.

The aircraft computes VPL as follow (see [8] for details):

$$\text{VPL} = K_{\text{ffmd}} \sqrt{\sum_i^N S_{\text{vert},i}^2 \sigma_i^2} . \quad (1)$$

$$\sigma_i^2 = \sigma_{\text{pr_gnd},i}^2 + \sigma_{\text{iono},i}^2 + \sigma_{\text{air},i}^2 + \sigma_{\text{tropo},i}^2 , \quad (2)$$

$$\sigma_{\text{iono},i} = F_i \sigma_{\text{vig}} (x_{\text{aircraft}} + 2\tau v_{\text{aircraft}}) , \quad (3)$$

$$\sigma_{\text{air},i}^2 = \sigma_{\text{multipath},i}^2 + \sigma_{\text{noise},i}^2 . \quad (4)$$

$S_{\text{vert},i}$ is the vertical position component of the weighted-least-squares projections matrix for satellite i (see equation (8)). N is the number of satellite in-view for the present epoch.

K_{ffmd} is the multiplier that determines the required probability of fault-free missed detection.

σ_i^2 is the variance of a normal distribution overbounding the range domain error distribution for satellite i for the fault-free hypothesis H_0 (see [9]).

$\sigma_{\text{pr}_{\text{gnd},i}}^2$ is the fault-free variance of the ground error term associated with the correction for satellite i .

$\sigma_{\text{iono},i}$ is the ground residual ionospheric uncertainty for satellite i . F_i is the vertical-to-slam obliquity factor for satellite i . σ_{vig} is the standard deviation of a normal distribution of the residual ionospheric uncertainty due to nominal spatial decorrelation (a typical value is 4 mm/km). The following factor has two components, the horizontal distance from GGF to aircraft (x_{aircraft}), and a synthetic distance produced by a smoothing filter used to mitigate multipath and noise with a time constant τ of 100 s. v_{aircraft} is the velocity of the aircraft in the direction of the airport.

For details on $\sigma_{\text{air},i}^2$ and $\sigma_{\text{tropo},i}^2$ see the above last references.

In the above set of equations is not included the vertical protection level under a single-satellite (satellite i) ephemeris fault, however it is considered in the computations. Failures of the hypothesis H_1 [10] are not relevant to excluding geometries [11].

3 Ionospheric Threat and Tolerable Error Limit

GPS satellites fly in medium Earth orbits (MEO) at an altitude of approximately 20200 km. The ionosphere is a region of the atmosphere located about 50 – 1000 km above the Earth's surface. In this region, solar radiation produces free electrons and ions that cause phase advances and group delays to GPS radio waves.

Ionospheric fronts (also called ionospheric storms) pose a significant threat to single frequency ground based augmentation systems (GBAS) for airplane precision approach because they can produce differential delays between the aircrafts and the GGF that are not detected in time to generate an integrity alert [12]. That is, uncorrelated delays could produce misleading and hazardously misleading situations. Ionospheric fronts are recommended [13] to be modeled by a moving wedge form with four parameters¹, including the maximum ionospheric spatial gradient for a particular region [14].

The first wedge model has been parameterized for the conterminous U.S. (CONUS) where ionospheric delay gradients as large as 435 mm/km have been observed [15]. In the Brazil ionosphere the largest gradient of about 850 mm/km has been registered [16]. There are not known developments of ionospheric threat wedge model for Argentina.

The development and utilization of the ionospheric threat model occurs in two stages [17]. The first stage is observation, in which data accumulated over a lengthy period (usually including the greatest activity of a solar cycle) is collected to describe and cover the features of ionosphere impact on GBAS [18]. The

¹ The four parameters are: spatial gradient, front moving speed, width and maximum delay [11].

results of observation are used to estimate the bounding parameters of the threat model. The second stage is simulation, in which the completed threat model is used in a simulation including the GGF and user operation. The simulations provide estimations of the integrity risk (particularly the hazardously misleading information (HMI) analyses [19]), the availability, and the impact to the ground and airborne monitors.

Equation (3) gives the magnitude of ionospheric range error in slant direction under steady-state conditions. The distance for decision height (DH) for CAT-I is 6 km away from the GGF. For a nominal σ_{vig} of 4 mm/km the error is bounded by 0.24 m, in about 99.8 % of the cases. Meanwhile an anomalous large ionospheric delay gradient of 425 mm/km could produce certainly a range error as large as 8.5 m.

The maximum vertical position error due to the worst-case user error induced by an ionosphere anomaly (as estimated by the ground subsystem) need only be bounded by a tolerable error limit (TEL) that luckily is greater than VAL [11, 15, 20, 21].

Several criteria have been proposed to establish a value for TEL. In [21] the obstacle clearance surface (OCS) concept is used to assess the safety of a CAT-I approach if an ionospheric anomaly produces vertical navigation errors. A maximum vertical error of about 29 m would be allowable at the nominal DH of 200 ft.

4 Algorithm Description

We followed [11, 15, 22, 23] as references for the algorithm implementation. The algorithm produces an inflation factor that depends of current satellite geometry. Inflation factors larger than 1 are applied to σ_{vig} when its nominal value do not produce a VPL that exceeds VAL when the possible error is not bounded by TEL. This mechanism excludes unsafe geometries for been considered by the aircraft.

The algorithm has four principal computational steps: 1. *GGF and Subset Geometries*, 2. *Ionosphere-Induced Range Error*, 3. *Ionosphere-Induced Vertical Error* and 4. *Parameter Inflation*.

4.1 GGF and Subset Geometries

Satellites that are visible to the GGF may not be included in the positioning solution of an approaching airplane. It is assumed that up to two satellites from the all-in-view satellite at the ground facility are not used by the airborne ². If there are N satellites visible to the GBAS ground facility, there are

$$\sum_{k=N-2}^N \binom{N}{k} \quad (5)$$

² The minimum number of satellites to be used by the airborne is four.

subset satellite geometries from the set of N satellites visible to the GGF. Figure 1 shows the evolution for the number of satellites in view. Figure 2 shows an example with nine satellites in view at a given epoch, the total number of subset geometries to consider in this case is, $\sum_{k=7}^9 \binom{9}{k} = 46$.

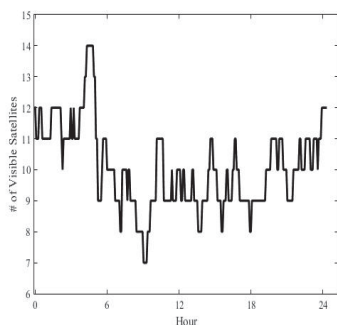


Fig. 1. Evolution of the number of visible satellites above 5° during a 24 hours typical period for a GBAS Ground Facility as if it was located at La Plata Airport (34.9655° S, 57.8954° W).

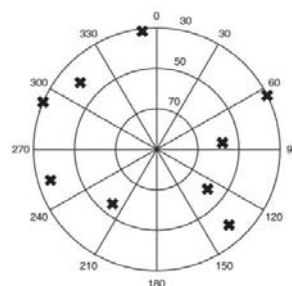


Fig. 2. Diagram of satellite positions over the sky above 5° (actually above 30° in this case) from the GBAS Ground Facility point of view. Nine satellite are in view for the current epoch.

4.2 Ionosphere-Induced Range Error

The closed-form approximation of the ionosphere-induced differential range error is:

$$\varepsilon = g \times (x_{\text{aircraft}} + 2\tau v_{\text{aircraft}}). \quad (6)$$

ε is the ionosphere-induced differential range error [m].

g is the ionospheric gradient [mm/km].

x_{aircraft} is the separation between the GGF and the approaching airplane [km].

τ is the time constant of the carrier-smoother filter [s].

v_{aircraft} is the velocity of the approaching airplane [km/s].

Given an airplane at a certain distance and with certain velocity equation (6) represents the worse case for a given g . In the bibliography (e.g., [24]) considerations are taken about the relative velocity between the ionospheric front and the particular satellite, and the preventive action of the code-carrier divergence (CCD) monitor is considered [25]. Neither of those advantages are taken here to possible reduce the range error ε for every satellite for a particular geometry.

4.3 Ionosphere-Induced Vertical Error

Given airplane and ionosphere front movement geometries, a large ionospheric gradient may be unobservable to the GGF and meanwhile affecting the airplane

that is not conveniently prevented of the integrity risk. While an airplane is approaching a runway, an ionospheric front can impact the SIS of two satellites ($k1$ and $k2$) simultaneously. Hence the worst case ionospheric error in vertical (IEV) for any pair of satellite of a given geometry for a particular epoch can be expressed as follows ³,

$$\text{IEV}_{k1,k2} = |S_{\text{vert},k1}\varepsilon_{k1}| + |S_{\text{vert},k2}\varepsilon_{k2}| . \quad (7)$$

$$S = (G^T W G)^{-1} G^T W . \quad (8)$$

$$G_i = [-\cos \text{El}_i \cos \text{Az}_i \quad -\cos \text{El}_i \sin \text{Az}_i \quad -\sin \text{El}_i \quad 1] . \quad (9)$$

$$W^{-1} = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_N^2 \end{bmatrix} . \quad (10)$$

Where $S_{\text{vert},ki}$ is the vertical position component of the weighted-least-squares projection matrix S for satellite ki (see [8]), this component is dependent of satellite geometry (number of satellite in view and their spatial distribution referred to the GBAS local ground station) at each epoch; G_i is the row of the observation matrix G for satellite i ; Az_i and El_i are the azimuth and elevation of satellite i (see Figure 2); W^{-1} is the inverse of the least square weighting matrix; and σ_i is the variance of a normal distribution that overbounds the true VPE distribution for satellite i under the fault-free hypothesis.

The IEV for every pair of each subset geometry considered for the approaching airplane is calculated and compared to obtain the maximum IEV (MIEV) for each geometry in a particular epoch. The number of pairs to be compared for each subset geometry with k satellites is

$$\binom{k}{2} . \quad (11)$$

Figure 3 shows an example of MIEVs compared with TEL.

Figure 4 shows the nominal VPLs at the same epoch as Figure 3. The subset geometries with VPL exceeding VAL are not approved by the airplane and they can not cause integrity failures even though their MIEVs are greater than TEL.

4.4 Parameter Inflation

From Figures 3 and 4 can be observed that some geometries exceeding TEL does not exceed VAL. In those cases exist potentially hazardous geometries not excluded by the integrity mechanism supported by the VPL computed with the nominal σ_{vig} broadcasted (see equations (1) and (3)).

³ A less constrained expression for IEV is presented in [22].

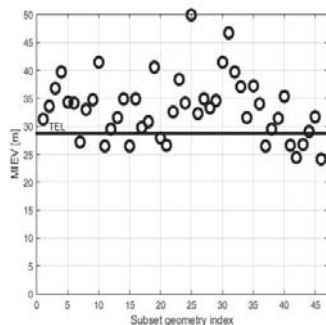


Fig. 3. MIEV for 29 subset geometries given 7 satellites in view. The 11 subset geometries exceeding TEL are potentially hazardous for this epoch.

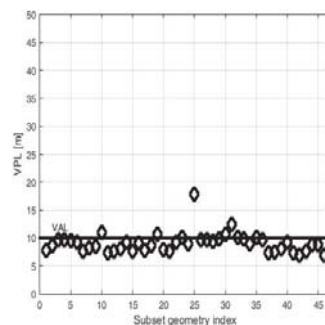


Fig. 4. Nominal VPL at the same epoch as Figure 3. The subset geometries with VPL exceeding VAL can not cause integrity failures.

If the GGF inflates the broadcasted parameters above their nominal values the airborne VPL increase as well. The most effective parameter to inflate is σ_{vig} which increase σ_i^2 .

To obtain the minimum σ_{vig} needed for an epoch, VPLs are pre-computed by the GGF iteratively increasing a inflation factor I_{vig} until all hazardous geometries are properly removed. Some particular considerations have to be made to make the computation but exceed the limits of this explanation (see [22, 24]). The new inflated σ_{vig} is broadcasted to the airplane. As Figure 5 shows, all hazardous geometries exceeding TEL are excluded after inflation.

5 Algorithm Test

The algorithm implementation was tested for the location coordinates of La Plata Airport taken samples of recorded ephemeris every minute of a typical day. The assumed position of the aircraft is the decision point at 200 *ft* of elevation. An effective separation of 20 *km* results from the sum of the actual distance of 6 *km* between the reference station and the user, and the 14 *km* of synthetic separation generated by the memory of the code-carrier smoothing filter used to mitigate multipath error and code noise [8]⁴. The nominal value of $\sigma_{\text{vig}} = 4.0 \text{ mm/km}$. VAL = 10 *m* for approaching operations. The only parameter of the threat model to be considered is the maximum ionospheric gradient $g = 400 \text{ mm/km}$.

Figure 6 shows the inflation factor I_{vig} . The inflated σ_{vig} can not exceed the maximum allowed broadcast value of 25.5 *mm/km* [26], meaning that in this case the maximum permitted value of $I_{\text{vig}} = 6.375$.

⁴ A constant velocity of 0.07 *km/s* is assumed during the approach. The time constant of the carrier-smoothing filter is $\tau = 100 \text{ s}$.

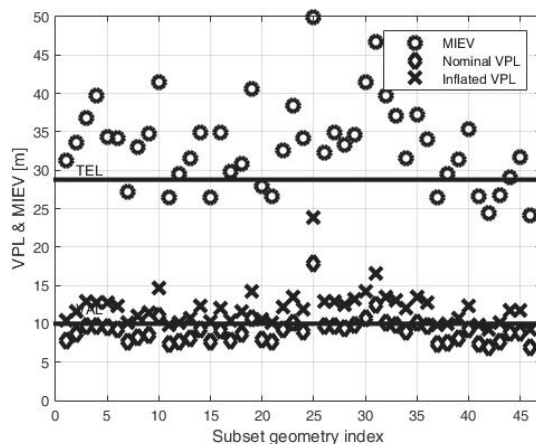


Fig. 5. Satellite geometry exclusion by σ_{vig} inflation. All hazardous geometries exceeding TEL are excluded after inflation.

By applying inflation factors to σ_{vig} all potentially hazardous geometries (and many more acceptable geometries) are eliminated from the approved set of geometries. Hence, system integrity is guaranteed by reducing system availability. Figure 7 shows the inflated VPL for the conditions of the test. The availability in this case remains 100 %.

6 Conclusions

It is difficult to improve integrity into existing systems while retaining sufficient availability to make the applications viable. As Pullen stated, these challenges are more than mathematical and require more than simply adding redundancy to mitigate the effects of individual failures [27].

Most of the anomalies affecting GBAS can be detected by the monitors founded in the CAT-I ground facility. Even moderate ionospheric anomalies are detected by the CCD and the clock acceleration monitors [12]. However, the worst-case ionospheric anomalies require additional mitigation techniques. The technique used here excludes possible hazardous geometries.

Equations (1) and (7) generate bounds too conservative to represent the wedge threat model and to accurately model the more realistic geometry and monitoring conditions found in practice. Nevertheless, the implementation shows that the availability is optimal for the site of La Plata Airport. This is due to the well suited geometry conditions, particularly the relatively large number of satellite always in view above 5° in elevation, as is observed in Figure 1.

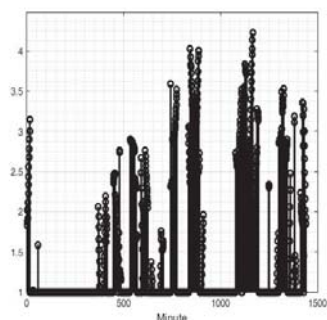


Fig. 6. Inflation factor I_{vig} for a 24 hours period. The broadcasted σ_{vig} is the inflation factor times the nominal $\sigma_{vig} = 4 \text{ mm/km}$.

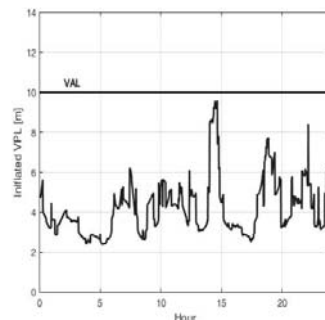


Fig. 7. Inflated VPL at La Plata airport for a typical day. Availability is 100 % for a DH at 6 km using all the satellite in view.

References

1. International Civil Aviation Organization (ICAO): International Standards, Recommended Practices and Procedures for Air Navigation Services Annex 10 (1985)
2. U S Federal Aviation Administration (FAA): Specification Performance Type One Local Area Augmentation System Ground Facility. FAA-E-2937A (2002)
3. Rinnan, A., Gundersen, N., Sigmond, M., Nilsen, J.: Operational GNSS Integrity. In: Dynamic Positioning Conference (2011)
4. Dautermann, T., Sgammini, M., Pullen, S.: GBAS Ionospheric Threat Analysis using DLRs Hardware Signal Simulator. In...
5. SANJEEV Gunawardena, S., Zhu, Z., De Haag, M., Graas, F.: Remote-Controlled, Continuously Operating GPS Anomalous Event Monitor. In: Journal of the Institute of Navigation (2009)
6. Park, Y., Pullen, S., Enge P.: Enabling LAAS Airport Surface Movement: Mitigating the Anomalous Ionospheric Threat. In: IEEE/ION, Position, Location, And Navigation Symposium (2010)
7. Suzuki, S., Nozaki, Y., Ono, T., Yosihara, T., Saitoh, S., Sonosuke Fukushima, S.: CAT-I GBAS Availability Improvement through Ionosphere Field Monitor (IFM). In Proceedings of the 24th International Technical Meeting of The Satellite Division of the Institute of Navigation (2011)
8. Radio Technical Commission for Aeronautics (RTCA): Minimum Operational Performance Standards for GPS Local Area Augmentation System Airborne Equipment. Technical Report DO253C (2008)
9. Radio Technical Commission for Aeronautics (RTCA): Minimum Aviation System Performance for the Local Area Augmentation System. Technical Report DO245A (2004)
10. Elias, P., Saotome, O.: System Architecture-based Design Methodology for Monitoring the Ground-based Augmentation System, Category I – Integrity Risk. In: J. Aerosp. Technol. Manag., São José dos Campos (2012)
11. Lee, J., Pullen, S., Park, Y., Enge P., Brenner, M.: Position-Domain Geometry Screening to Maximize LAAS Availability in the Presence of Ionosphere Anomalies.

- In: Proceedings of the 19th International Technical Meeting of the Satellite Division of The Institute of Navigation ION GNSS (2006)
12. Luo, M., Pullen, S., Akos, D., Xie, G., Datta-Barua, S., Walter, T., Enge P.: Assessment of Ionospheric Impact on LAAS Using WAAS Supertruth Data. In: Proceedings of the ION 58 th Annual Meeting (2002)
 13. International Civil Aviation Organization South American Office (ICAO-SAM): Guide for Ground Based Augmentation System Implementation (2013)
 14. Pullen, S. Park Y., Enge, P.: The Impact and Mitigation of Ionospheric Anomalies on Ground-Based Augmentation of GNSS. In: Radio Science (2009)
 15. Pullen, S., Enge, P.: An Overview of GBAS Integrity Monitoring with a Focus on Ionospheric Spatial Anomalies. In: Indian Journal of Radio & Space Physics (2007)
 16. Lee, J., Yoon, M., Pullen, S., Gillespie J., Mather, N., Cole, R., Rodrigues de Souza, J., Doherty, P., Pradipta, R.: Preliminary Results from Ionospheric Threat Model Development to Support GBAS Operations in the Brazilian Region. In: Proceedings of the 28th International Technical Meeting of The Satellite Division of the Institute of Navigation (2015)
 17. International Civil Aviation Organization Asia and Pacific Office (ICAO-APAC): GBAS Safety Assessment Guidance Related to Anomalous Ionospheric Conditions (2016)
 18. Kim, M, Lee, J. Pullen, S., Gillespie, J.: Data Quality Improvement and Applications of Long-Term Monitoring of Ionospheric Anomalies for GBAS. In: Proceedings of ION GNSS (2012)
 19. U.S. Federal Aviation Administration (FAA): Ground Based Augmentation System Performance Analysis and Activities Report. First Quarter Report (2017)
 20. Murphy, T., Harris, M., Park, Y., Pullen, S.: GBAS Differentially Corrected Positioning Service Ionospheric Anomaly Errors Evaluated in an Operational Context. In: Proceedings of the 2010 International Technical Meeting of The Institute of Navigation (2010)
 21. Shively, C., Niles, R.: Safety Concepts for Mitigation of Ionospheric Anomaly Errors in GBAS. In: Proceedings of the 2008 International Technical Meeting of The Institute of Navigation (2008)
 22. Lee, J., Seo, J., Park, Y., Pullen, S., Enge, P.: Ionospheric Threat Mitigation by Geometry Screening in Ground-Based Augmentation Systems. In: Journal of Aircraft (2011)
 23. Vemuri, S., Sarma, A., Redd, A., Reddy D.: Investigation of the Effect of Ionospheric Gradient on GPS Signals in the context of LAAS. In: Progress in Electromagnetics Research B (2014)
 24. Seo, J., Lee, J., Pullen, S., Enge, P., Close, S.: Targeted Parameter Inflation Within Ground-Based Augmentation Systems to Minimize Anomalous Ionospheric Impact. In Journal of Aircraft (2012)
 25. Simili, D., Pervan, B.: Code-Carrier Divergence Monitor for the GPS Local Area Augmentation System. In: IEEE/ION, Position, Location, And Navigation Symposium (2006)
 26. Radio Technical Commission for Aeronautics (RTCA): Minimum Operational Performance Standards for GPS Local Area Augmentation System Airborne Equipment. Technical Report DO246D (2008)
 27. Pullen, S.: Lessons Learned from the Development of GNSS Integrity Augmentations. In: Coordinates, <http://mycoordinates.org> (2016)

IoT for smart home energy planning

Emilio Orsi and Sergio Nesmachnow

Centro de Cálculo, Facultad de Ingeniería, Universidad de la República
Herrera y Reissig 565, 11300 Montevideo, Uruguay
Email: {eorsi,sergion}@fing.edu.uy

Abstract. This article presents a platform combining hardware and software for smart power consumption monitoring and planning in urban scenarios. The system integrates a hardware controller for energy efficiency, a communication protocol to improve data transmission, and a software module for planning and managing home devices. The proposed solution is implemented applying the Internet of Things paradigm, allowing the integration of computational intelligence techniques. A greedy algorithm is proposed for planning, according to user preferences and a maximum allowed power consumption. Results show that the power consumption of a water heater is reduced up to 38.9%, and two water heaters and one air conditioning can be optimized simultaneously without reducing the quality of service. These results suggest that the proposed approach is useful for home power consumption planning.

Keywords: Smart grid; energy; demand management

1 Introduction

The *smart grid* concept proposes using information and communication technologies for improving the quality and the performance of electrical networks to increase the efficiency, robustness, and adaptability of the network as a whole [6] [5]. This paradigm is usually applied when electricity is generated by means of unpredictable renewable energies, such as wind and solar. Uncertainty poses one of the biggest challenges in terms of the system operation to guarantee appropriate quality of service (QoS) to users of the electrical system.

Systems including mathematical models and algorithms are applied to control electrical devices to properly address issues related to QoS and energy distribution. Specific hardware support is also needed for measurement and management using sensors, data communication, and control devices. These pieces of hardware must be able to communicate and cooperate to properly react to dynamic situations (e.g., peaks of power consumption, valleys on renewable energy generation, price changes). Control devices allow implementing planning techniques to better use the available resources [10].

This article reports the advances on building a system for smart home energy management that includes an energy controller and computational intelligence methods for planning the power consumption on houses, buildings, and

neighborhoods. The main goal of the proposed system is to provide an efficient and intelligent management of electric devices, reducing the power consumption and the monetary cost, while satisfying QoS requirements by users. We plan to implement the system in a distributed platform applying Internet of Things (IoT) and cloud computing paradigms, providing great flexibility on hardware resources and communications. The research is developed within the STIC-AmSud project Cloud Computing for Smart-City Energy Management (CC-SEM) [16].

The article is organized as follows. Next section introduces the home energy management problem and reviews related works. The proposed system for energy management is described in Section 3. Section 4 evaluates the algorithm for home energy planning over two case studies. Finally, Section 5 presents the conclusions and the main lines for future work.

2 The home energy management problem

This section presents the home energy management problem and reviews relevant related works.

2.1 Problem description

Load management allows users to adjust power consumption according to expected QoS, different energy prices, etc. Planning techniques allow leveling the power consumption in homes. This way, users can apply strategies to shift the load of electric devices from peak hours (having high demand and expensive energy prices) to low demand and cheaper energy price periods [9]. *Demand response* [11] changes the way of managing energy flow for homes and cities to better match the power demand with the power supply, according to the energy production capabilities. Demand response techniques allow reducing peak demand, thus improving the quality of service of the electric grid as a whole.

The home energy management problem proposes using information systems and algorithms to find appropriate energy plannings, i.e., assigning a shared resource (energy) to several tasks or loads (electric devices) that consume energy.

Home devices have different features and power consumption profiles, which affect possible actions to take within a plan to minimize negative effects on the QoS provided to users. *Interruptible* loads, which can be interrupted during periods of time without causing a significant negative impact in the QoS, are very important for energy planning without affecting the user experience. Usually, interruptible loads preserve energy in form of heat (e.g., water heater and air conditioner). Taking into account the load features and the availability of both real-time and historical data (patterns/preferences of utilization for each home device), the proposed system uses a home controller, a specific protocol for load monitoring and controlling, and a planning system for home energy efficiency.

In order to handle the communication of power consumption and control messages in real time, data flow must be optimized to provide a robust and efficient demand management [2], thus the need of an optimized protocol for

communication. A controller is needed because most current home devices do not allow digital/remote controlling. The planning system allows defining specific actions to reduce the energy cost without affecting the QoS provided to the users.

2.2 Related work

There are a few examples of home automation devices for energy efficiency. Most of them are set to control specific environment parameters (e.g., heating/cooling), like Nest devices [1] or perform basic automation (e.g., turning devices on/off according to a user schedule). These devices lack a global view on the consumer profile and devices. Consumption meters are also becoming smart, like *ElectriSense* [4] that can identify the families of devices, or *Linky* [18] a connected device deployed by *Électricité de France*. None of these systems are open-source and their capabilities are often restricted to data acquisition and basic automation.

Gomes et al. [3] studied a scenario using several micro generation units (MGU) in a low tension distribution network. MGUs cause voltage fluctuations when switching on/off home devices, thus demand response is useful to preserve the quality of the network. Specific techniques were proposed for shifting the power consumption of washing/drying machines and dish washers; reducing the temperature of water heaters in peak hours; and shifting the use of water heaters. A simple scenario was studied, considering 10 MGUs with and without demand management. The analysis showed that the quality variations due to MGUs and demand management are important, but can be kept under acceptable values.

Soares et al. studied demand management for homes and categorized home devices, according to their behavior and the different management strategies to be applied [13] and presented a model and an algorithm for planning the power utilization to reduce the cost [14]. A non-linear integer programming problem formulation was introduced and Evolutionary Algorithms (EA) were applied to minimize the bills and the violations to the maximum contracted power. The proposed EA allowed reducing up to 40% the energy cost for users with respect to a reference scenario without demand management. Later, a multiobjective EA was applied to find a set of plans for minimizing cost and maximizing user satisfaction [12]. Plans are evaluated according to user-defined preferences for different times of the year. The main results showed that the cost reduction is between 22% and 24%, depending on the satisfaction levels. However, the experiments are based on synthetic data and no real system for implementing demand management in homes was proposed.

The analysis of the related works indicates that there is still room to contribute with solutions focused on developing a system to implement home demand management. The system proposed in this article combines a hardware controller and computational intelligence methods for power consumption management and planning.

3 A system for smart home energy management

The proposed system integrates three components, implemented in independent modules: EFEN for defining/storing user actions and preferences, and compute plannings; the home CONTROLLERS, and KHIMO for communication with the home controllers. These modules allow changing the electric flow in real time, controlling the devices according to specific rules. The modules and their interactions are shown in Fig. 1 and described next.

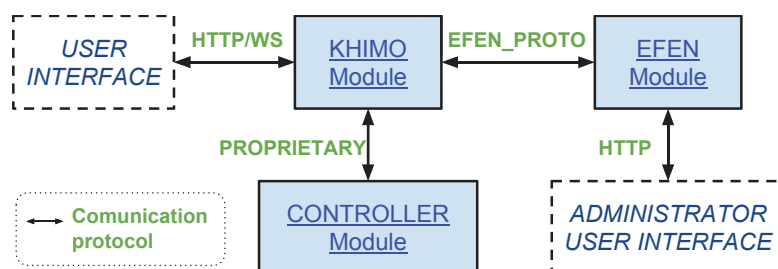


Fig. 1: Architecture of the proposed system.

EFEN. EFEN is the central module of the system, which defines the devices, stores the power consumption data, provides an interface for device control, and computes the planning. EFEN groups homes in buildings and buildings in neighborhoods, allowing energy planning at different levels.

EFEN also provides an Application Programming Interface (API) to integrate computational intelligence algorithms for big data analysis/pattern recognition and energy planning. Further details can be found in <https://www.fing.edu.uy/inco/grupos/cecal/hpc/EFEHO>.

Big data algorithms allow extracting useful information from raw power consumption data, to be used by machine learning and planning tools. This approach ensures scalability and portability of the developed solution, allowing a smooth transition from the prototyping/validation phases to a production phase. Basic ETL (Extraction, Transformation, Load) of data are implemented using MapReduce, while more advanced operations can be conducted with high-level tools (e.g., Pig or Apache Spark). Storage and access to the data are improved by using NoSQL databases, which allow storing data from sensors and other data sources (e.g., power generation, historical consumption charts, weather forecast). Due to its lightweight and efficient communication protocol, EFEN is capable of controlling over 50 devices per building within a neighbor context. Computational intelligence methods allow identifying correlated patterns and situations to be used on the optimization algorithms [17]. In EFEN, pattern analysis is performed to determine device utilization and cycle of power consumption at home/building/neighborhood levels. EFEN also allows the user to define preferences for using each device in a day or in a week.

Planning defines specific actions to be taken on the devices (an *agenda*), considering power consumption, pricing, user preferences, and user satisfaction. Other features included in EFEN are: feasibility check of agendas, evaluation of the energy stored in form of heat (for water heaters and air conditioning systems), cost evaluation, and user satisfaction estimation. User satisfaction is evaluated according to the features and services provided by each load. For example, for a water heater, satisfaction is related to the deviation from the water temperature the user expects when using it. These features are critical when evaluating the impact of interrupting a device that generates heat. For verification purposes, a tool to simulate historical power consumption time series is also included in EFEN.

CONTROLLER. The controller includes a module for power measuring, a relay for power supply control, and a Single Board Computer (SBC) that controls the measures and the relay, and communicates with KHIMO.

A raspberry Pi (rPi) is used as SBC, providing a flexible and portable solution. Additional hardware include a relay and a STPM01/10 module [15] for power measuring using the SPI protocol. Figure 2 shows the connections between the rPi and the relay. The red wire controls the state, purple is the ground wire and blue connects to the power source (5V).

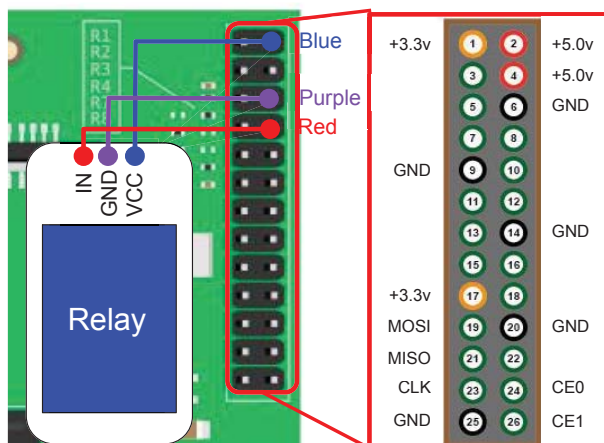


Fig. 2: Connections between the rPi and the relay.

The STPM01/10 chip was integrated using the steval-ipe016v1 board. Figure 3 shows the connections between the rPi and the SPI module, including MOSI, MISO (purple), and CLK (grey), the signal for read/write the channel (pin 22, green), the ChipSelect signal (pin 18, blue), the ground wire (pin 9, pink) and the VCC wire (pin 17, mustard).

The controller is presented in Fig. 4 (a male connector is on the rear side). Further details can be found in <https://www.fing.edu.uy/inco/grupos/cecal/hpc/EFEHO>.

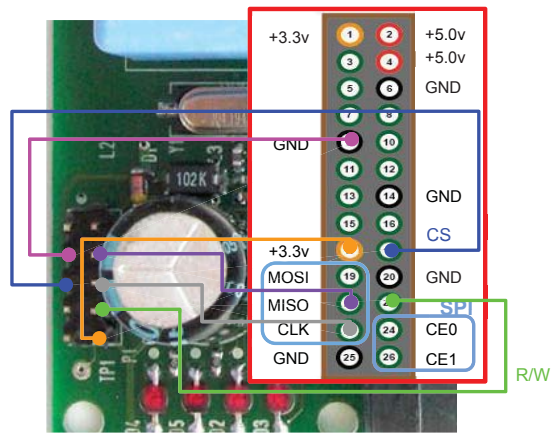


Fig. 3: Connections between the rPi and the STPM01/10.

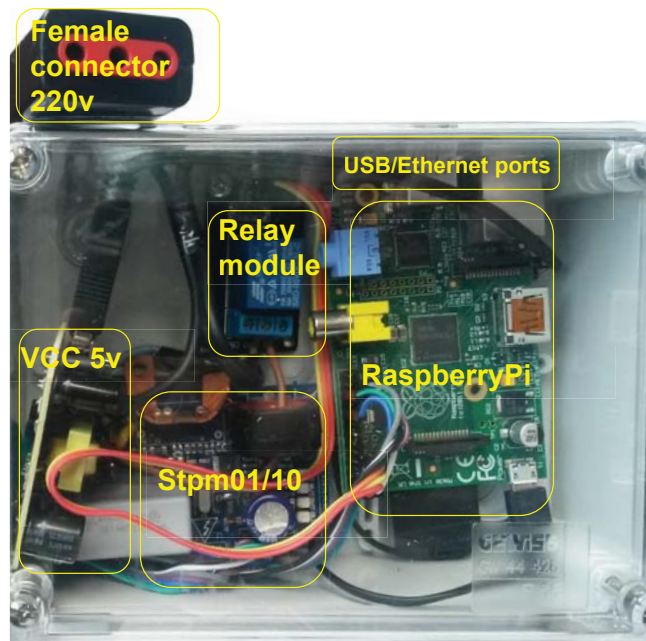


Fig. 4: Overview of the smart home energy controller.

KHIMO. The system is controlled by communications applying the IoT paradigm [8] to guarantee ubiquitous access using a wide range of communication devices (smartphones, tablets, web interface, etc.). The Khimo platform [7] is used to implement the communications between the home controller and the management system. Khimo provides the user interface that allows remote monitoring and controlling of several devices in real time.

The communication between KHIMO and EFEN is performed using EFEN_PROTO, a protocol specifically developed to guarantee efficiency and different functionalities for device control. EFEN_PROTO provides a channel for bidirectional communication to define actions to be performed on the loads and gathering information about the state and power consumption of the loads. Efficiency is achieved by enabling event subscription and avoiding intermittent queries (*polling*) to each device. Possible events to subscribe are state changes, power consumption changes, and connection events. In addition, message delivering (i.e., all pending actions) is performed simultaneously for all required devices via action grouping. EFEN_PROTO groups the loads according to the actions to perform and not viceversa. This mechanism avoids redundancy, thus improving the communication efficiency.

4 Experimental evaluation

This section presents a sample application of the proposed system for managing two scenarios including water heaters and an air conditioner. We focus the analysis on relevant interruptible loads that generate energy in the form of heat, as they are frequently used (especially in Uruguayan homes) and have a significant impact on domestic power utilization [13].

4.1 A greedy algorithm for defining agendas

A greedy algorithm is proposed to minimize power consumption and maximize user satisfaction by taking local decisions to build a global agenda. The maximum power available is a hard restriction of the planning problem.

The greedy algorithm applies the following steps:

1. Create an action considering the utilization periods defined in EFEN. The first action intends storing the total expected heat, and the following actions intend maintaining the heat in acceptable ranges. Each action is assigned a priority according to the probability of utilization of the corresponding load.
2. If the defined action generates a conflict with previously defined actions, priorities are taken into account to shift (anticipate or delay) the actions in conflict.
3. If there are time steps in the utilization periods that are not covered by an action, go to step 1
4. Store the actions in EFEN, to be applied when appropriate.

The proposed greedy algorithm is a simple but reasonable method for defining agendas. It is based on keeping the expected heat (and thus, the expected QoS) in a reasonable range for all time steps and shifting loads when needed, according to the user preferences. Furthermore, the proposed greedy algorithm can be used as an initial solution for applying more powerful optimization techniques (machine learning, EAs, etc).

4.2 Experimental analysis

The evaluation of the proposed algorithm for load planning focus on shifting power consumption of interruptible loads trying to fulfill the user satisfaction and not exceed a maximum power limit. The evaluation scenarios involve two simulated water heaters and one simulated air conditioner. Table 1 presents the features of the devices defined in EFEN. The tolerance for temperature is 10%.

Table 1: Features of the devices defined in EFEN.

	<i>water heater</i>	<i>air conditioner</i>
Power consumption (W)	1500	2000
Time to reach target temperature	40 m	25 m
Time to lose all heat	480 m	80 m
Maintenance frequency	48 m	14 m
Time for maintenance	4 m	2 m

Scenario 1: power consumption optimization for a water heater. According to the defined device, the water heater must be on for 4 minutes each period of 48 minutes to keep the temperature in the accepted range. According to the agenda computed for the water heater between 8:00 and 12:00, it is necessary to turn on the water heater for 56 minutes to guarantee the required QoS. The agenda allows reducing up to **38.9%** (88 minutes less) the time that the water heater is on, when compared with the non-optimized scenario.

Scenario 2: simultaneous optimization of power consumption for two water heaters and one air conditioner. This scenario sets a maximum instant power of 2200 W, to analyze a situation when a low power value is available, thus devices cannot be on at the same time. Devices are set to be used between 8:00 (minute 480) and 12:00 (minute 720). Water heaters have higher priority than the air conditioner.

Figure 5 shows the agenda computed by the greedy algorithm. Air conditioner is shifted to the beginning of the period, and water heater actions are shifted after that, according to the defined priorities. Water heaters interleave between 7:00–8:20, and no overlapping is observed. The graphic also shows short periods of power consumption to maintain the temperature.

Some agendas do not fulfill the QoS requirements at 100%. However, power consumption is less than when no automatic load management is applied and QoS is within the tolerance.

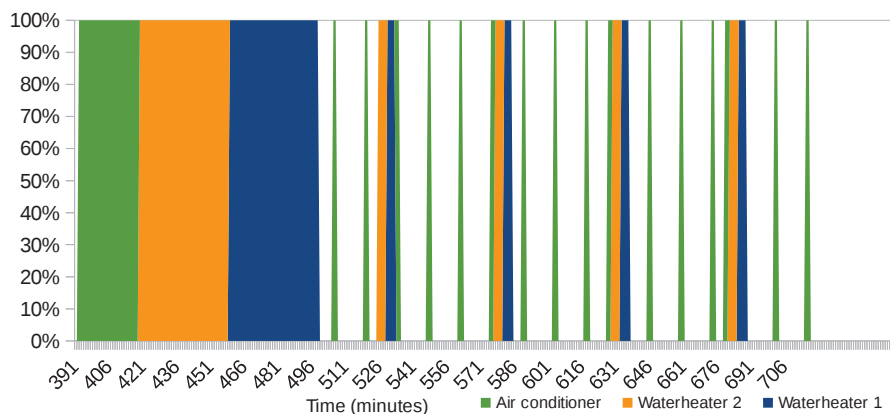


Fig. 5: Power consumption (agenda defined by EFEN for scenario 2)

5 Conclusions and future work

This article presented a hardware and software system for smart power consumption monitoring, controlling, and planning in urban scenarios. The system integrates a hardware controller for energy efficiency in homes, a protocol for efficient data communication, and a software module for planning and managing home devices and controllers.

The home energy management problem was studied by integrating cutting-edge technologies (computational intelligence, IoT, and cloud computing). A greedy algorithm was introduced for planning, according to user preferences and maximum allowed power consumption. The main results show that power consumption of a water heater can be reduced up to 38.9%. In addition, two water heaters and one air conditioning can be optimized simultaneously without reducing the QoS.

Results suggest that the proposed system is useful for home power consumption planning. It provides a robust platform that can be integrated to the Uruguayan power network for demand response and power consumption leveling.

The main lines for future work are related to improving the automation features of the proposed system, integrating computational intelligence/machine learning for pattern detection, and developing more powerful planning algorithms. The proposed system can also be extended to other public services.

References

1. Nest. <https://nest.com>. May 2017.
2. D. Balčiūnas and A. Kajackas. Transmission issues of the billing and dynamic power quality measures from the domestic sector. In *11th Int. Conf. on Electrical Power Quality and Utilisation*, pages 1–6, 2011.

3. A. Gomes, A. Soares, and C. Antunes. Impacts of demand side management and micro-generation units on low voltage distribution radial networks. In *11th Int. Conf. on Electrical Power Quality and Utilisation*, pages 1–7, 2011.
4. S. Gupta, M. Reynolds, and S. Patel. Electrisense: singlepoint sensing using emi for electrical event detection and classification in the home. In *12th ACM Int. Conf. on Ubiquitous Computing*, pages 139–148, 2010.
5. S. Gyamfi, S. Krumdieck, and T. Urmee. Residential peak electricity demand response—highlights of some behavioural issues. *Renewable and Sustainable Energy Reviews*, 25:71–77, 2013.
6. D. Huang and R. Billinton. Effects of load sector demand side management applications in generating capacity adequacy assessment. *IEEE Transactions on Power Systems*, 27(1):335–343, 2012.
7. Ikatu. Khimo. www.khimo.com. May 2017.
8. S. Karnouskos. The cooperative Internet of Things enabled smart grid. In *14th IEEE Int. Symp. on Consumer Electronics*, pages 7–10, 2010.
9. D. Long, S. Ploix, M. Jacomino, and M. Hoang. Home energy management problem: towards an optimal and robust solution. In *Energy Management*. 2010.
10. E. Orsi and S. Nesmachnow. Smart home energy planning using IoT and the cloud. In *IEEE URUCON*, 2017.
11. P. Palensky and D. Dietrich. Demand side management: Demand response, intelligent energy systems, and smart loads. *IEEE Transactions on Industrial Informatics*, 7(3):381–388, 2011.
12. A. Soares, C. Antunes, C. Oliveira, and A. Gomes. A multi-objective genetic approach to domestic load scheduling in an energy management system. *Energy*, 77:144–152, 2014.
13. A. Soares, A. Gomes, and C. Antunes. Categorization of residential electricity consumption as a basis for the assessment of the impacts of demand response actions. *Renewable and Sustainable Energy Reviews*, 30:490–503, 2014.
14. A. Soares, A. Gomes, C. Antunes, and H. Cardoso. Domestic load scheduling using genetic algorithms. In *European Conference on the Applications of Evolutionary Computation*, pages 142–151, 2013.
15. STMicroelectronics. STPM01. Programmable single phase energy metering IC with tamper detection. May 2017.
16. Universidad de la República. Cloud Computing for Smart-City Energy Management (CC-SEM) STIC-AmSud project. May 2017.
17. K. Zhou and S. Yang. Understanding household energy consumption behavior: The contribution of energy big data analytics. *Renewable and Sustainable Energy Reviews*, 56:810–819, 2016.
18. Électricité de France. Linky. <http://www.erdf.fr/linky-communicating-meter>. May 2017.

Sistema de Control de una cámara de germinación hidropónica con IoT.

Marcelo Marinelli , Rubén R. Urquijo
Instituto de Biotecnología de Misiones "María Ebe Reca". Campus U.Na.M. Ruta
12 Km 7,5, Miguel Lanús, Posadas, Misiones, Argentina.
{marcelo, ruben}@gmail.com

Resumen. Se presenta el diseño de un sistema de control difuso y telemetría con tecnología IoT para una cámara de germinación de cultivos hidropónicos. El sistema fue desarrollado con software NODE-RED y hardware Raspberry Pi, toma mediciones de humedad y temperatura dentro de la cámara y, por intermedio de un controlador difuso tipo Sugeno, genera señales de control para un sistema de ventilación forzada. El programa envía datos de las variables intervinientes y de las reglas difusas que se activan por mail y Twitter permitiendo recibir el estado del sistema desde cualquier dispositivo con conexión a internet.

Palabras claves: Lógica Difusa, Hidropónia, Internet de las Cosas, Web Services, E-Mail.

1 Introducción

La cámara de germinación para cultivos hidropónicos consiste en un recipiente plástico con una solución de nutrientes sobre la cual flota una placa de poliestireno expandido con perforaciones rectangulares. La evolución de las tecnologías de Internet tienden a integrar servicios de la web, redes sociales, redes de sensores que pueden obtener información de variables ambientales o de cualquier otro tipo y además controlar actuadores, cámaras, etc.; la integración de estos dispositivos y servicios dan como resultado la IoT [2] [3]. Cuando se trabaja con un conjunto de sensores inalámbricos distribuidos espacialmente e interconectados por distintos concentradores, se denomina WSN (*wireless sensor network*) utilizado para monitoreo ambiental, sistemas interconectados de energía, etc., [4].

A la hora de desarrollar aplicaciones del tipo IoT, debemos tener en cuenta que se deben considerar la adquisición de datos proveniente de sensores, tomar decisiones para realizar acciones sobre actuadores, cámaras u otros dispositivos y en tiempo real interactuar con servicios en línea. Si tuviéramos que programar una aplicación IoT de este tipo, se necesitan conocer varios lenguajes y protocolos, así como desarrollar APIs específicas y vincularlas con todos los componentes del sistema [5]. Por esto, surgen varias aplicaciones que integran los componentes que permiten desarrollar aplicaciones IoT. Estas se componen de nodos con funciones específicas y se interconectan por conexiones que intercambian flujo de datos.

Para alojar cubos de espuma fenólica (Figura 1), en estos orificios se siembran las semillas para su germinación y desarrollo hasta su trasplante a otro sistema hidropónico que puede ser de raíz flotante o NFT (*Nutrient Film Technique*) [1]. La cámara posee iluminación artificial por medio de paneles led o luces de bajo consumo y un sistema de ventilación forzada cuyo flujo de aire se controla mediante un control PWM (*pulse-width modulation*), en su interior posee un sensor de humedad y temperatura DHT22 (Figura 2). Para mejorar el proceso de desarrollo de las plántulas, se diseñó e implementó un sistema basado en tecnología IoT (*Internet of Things* o Internet de las Cosas) en un sistema embebido, que consta de un control difuso que, en función de las variables humedad y temperatura, dentro de la cámara, proveen una salida de control de potencia del sistema de ventilación forzada [1].



Figura 1. Placa de poliestireno expandido con los cubos de espuma fenólica en los cuales se desarrollaron las plántulas.

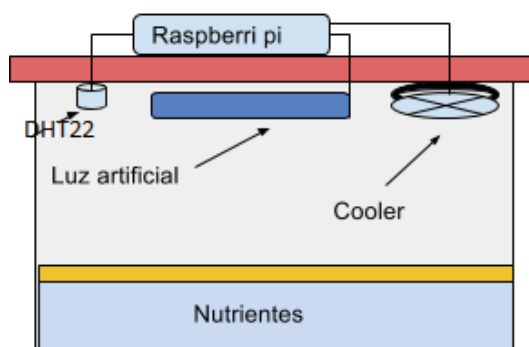


Figura 2. Diagrama de la cámara de germinación.

2 Hardware del sistema

En este trabajo se desarrolló un sistema de control de una cámara de germinación con un computador de placa simple o SBC (*Single board computer*) Raspberry Pi (<https://www.raspberrypi.org/>) que tiene la característica de ser un dispositivo de bajo costo y soporta sistemas operativos GNU/Linux ARM (Debian, Fedora, Arch Linux), RISC OS2.

Tiene un tamaño reducido (85 mm de longitud por 56 mm de ancho) y posee una salida HDMI, una ethernet y 4 USB con lo que, si se conecta un monitor, teclado y ratón, se convierte en un ordenador. Tiene una CPU ARM1176JZF-S (armv6k) a 700 MHz3, GPU Broadcom VideoCore IV3, memoria 512 MiB, Capacidad de almacenamiento Tarjeta SD o SDHC (Figura 3).

Debido a su reducido tamaño se puede montar sobre la cámara de germinación y con sus puertos de propósito general GPIO (*General Purpose Input/Output*) se pueden controlar sensores y actuadores para desarrollar un sistema de control de la misma.

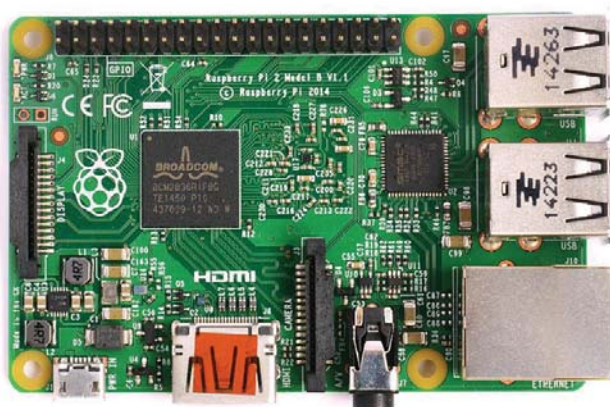


Figura 3. Raspberry Pi 2 modelo B.

2.1 Puertos GPIO

Los puertos GPIO son, como su nombre lo indica, puertos de entrada/salida de propósito general, es decir, un conjunto de conexiones que pueden utilizarse como entradas o salidas para diversos usos. Estos puertos representan la interfaz entre la Raspberry Pi y el medio exterior [6].

Posee un total de 17 puertos configurables como entrada o salida. Por defecto están todos configurados como entradas excepto los GPIO 14 y 15 que operan como salidas (Figura 4).

En este caso se utilizó el puerto GPIO 5 como entrada digital para obtener los datos provenientes del sensor de humedad y temperatura.



Figura 4. Salidas GPIO de Raspberry Pi.

2.2 Sensor de temperatura DHT22

Para obtener los valores de las variables humedad relativa y temperatura, se utilizaron sensores DHT22 (Figura 5 y Tabla 1) que miden estos parámetros. Estos sensores poseen un rango de operación de 0 a 100 % HR y de -40 a 80 °C y contando con una precisión de humedad de 2% RH y de temperatura 0,5%. Por otra parte, tiene la característica de trabajar con protocolo serial enviando los datos por una sola vía, por el pin de datos (Figura 5).

Tabla 1. Pines de salida de DHT22.

Pin	Función
1	5 v
2	Datos
3	NC
4	GND



Figura 5. Sensor DHT22.

De acuerdo con la Tabla 1, el pin 1 corresponde a la alimentación, el 4 a masa (GND) y el pin 2 tiene una resistencia de *pull up* de 10 K por donde son enviados los datos de humedad y temperatura en forma serial hacia el puerto GPIO correspondiente de la Raspberry Pi (Figura 6).

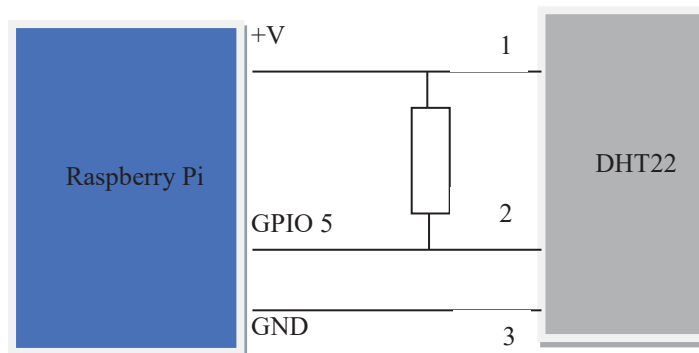


Figura 6. Conexiones del sensor DHT22 con el puerto GPIO.

2.3 Control del sistema de ventilación forzada

La salida PWM del puerto GPIO (GPIO 17) se conecta a un amplificador que permite controlar los niveles de potencia del sistema del *cooler*, esta salida entrega un nivel de potencia calculado por el controlador difuso en función de las variables de entrada y de la base de conocimientos del mismo.

3 NODE-RED

Es un editor basado en web para generar aplicaciones de IoT. De código abierto, creado por el equipo de *IBM Emerging Technology*. Posee una interfaz amigable que, mediante la función de arrastrar y soltar, permite colocar en el área de trabajo nodos que representan APIs de web, servicios en línea o dispositivos de hardware. Estos nodos se pueden unir con un cableado que representan el flujo de datos [7], también se pueden programar funciones en *JavaScript*; de esta manera es posible programar nodos con determinadas funcionalidades. El motor de tiempo real se basa en Node.js que es un entorno para ejecutar *JavaScript* realizado con el motor V8 de Chrome. Dado que Node.js trabaja con arquitectura basada en eventos no bloqueantes que se ejecutan del lado del servidor, esto hace que consuma poco costo de hardware lo que lo hace ideal para correr en placas Arduino (www.arduino.cc) o Raspberry Pi [8].

3.1 Controlador difuso

Se optó por un sistema de control difuso debido a la mayor interpretabilidad que posee para ser utilizado por expertos de otras disciplinas, como en este caso en que la base de conocimientos del sistema difuso se basó en información provista por expertos en Biología Vegetal [1].

Para la codificación del controlador difuso se seleccionó el tipo Sugeno [9], ya que al no poseer etapa de defusificación es posible programarlo con menor costo de procesamiento. Para la implementación se utilizó un nodo del tipo “función” y se codificó en JavaScript.

A la salida de este nodo se obtiene un flujo de datos que contiene la información de las reglas que se activan y el valor z que corresponde al porcentaje de potencia que se aplicará al nodo siguiente con salida PWM conectado al *cooler*.

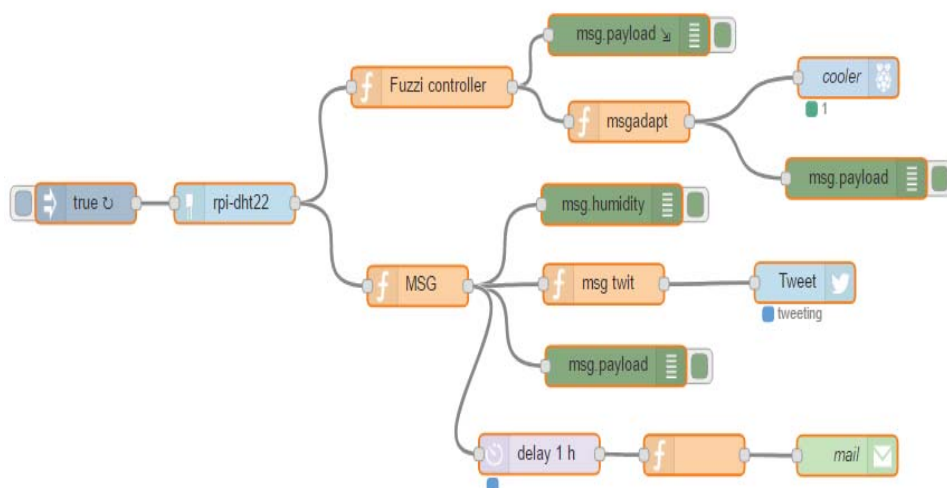


Figura 8. Diagrama de nodos del sistema de control de la cámara de germinación.

4 Telemetría social

La implementación de mensajería en redes sociales se hace a través de Twitter cada 10 minutos y por mail cada 60 minutos.

Estos mensajes son configurables y permiten tener información a distancia en un dispositivo móvil para estar al tanto de los valores de las variables intervinientes en la cámara de germinación y, en caso de algún valor anómalo, poder tomar una acción correctiva.

5 Resultados

El sistema de control diseñado combina tecnologías de IoT utilizando como soporte de hardware placas del tipo SBC (Raspberry Pi) y software Node-Red en el cual se implementó un sistema de control basado en lógica difusa.

La telemetría social permite informar los estados de las variables de la cámara por mail y por Twitter, de esta forma se tiene actualizado el estado de las variables intervinientes y poder tomar acción correctiva en caso de ser necesario.

6 Referencias

1. Marinelli, Marcelo J, Otegui, Mónica B, Zapata, Pedro D, & Acosta, Nelson H. (2015). Control Difuso de una Cámara de Germinación para Hidroponía. *Revista de Ciencia y Tecnología*, (24), 42-47. Recuperado en 25 de abril de 2017, de http://www.scielo.org.ar/scielo.php?script=sci_arttext&pid=S1851-75872015000200007&lng=es&tlng=es.
2. Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future generation computer systems*, 29(7), 1645-1660.
3. Kelly, S. D. T., Suryadevara, N. K., & Mukhopadhyay, S. C. (2013). Towards the implementation of IoT for environmental condition monitoring in homes. *IEEE Sensors Journal*, 13(10), 3846-3853.
4. Mainetti, L., Patrono, L., & Vilei, A. (2011, September). Evolution of wireless sensor networks towards the internet of things: A survey. In *Software, Telecommunications and Computer Networks (SoftCOM)*, 2011. 19th International Conference on (pp. 1-6). IEEE.
5. Blackstock, M., & Lea, R. (2014, October). Toward a distributed data flow platform for the web of things (distributed node-red). In *Proceedings of the 5th International Workshop on Web of Things* (pp. 34-39). ACM.
6. Jain, S., Vaibhav, A., & Goyal, L. (2014, February). Raspberry Pi based interactive home automation system through E-mail. In *Optimization, Reliability, and Information Technology (ICROIT)*, 2014 International Conference on (pp. 277-280). IEEE.
7. Kleinfeld, R., Steglich, S., Radziwonowicz, L., & Doukas, C. (2014, October). glue. things: a Mashup Platform for wiring the Internet of Things with the Internet of Services. In *Proceedings of the 5th International Workshop on Web of Things* (pp. 16-21). ACM.
8. Salihbegovic, A., Eterovic, T., Kaljic, E., & Ribic, S. (2015, May). Design of a domain specific language and IDE for Internet of things applications. In *Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 2015 38th International Convention on (pp. 996-1001). IEEE.
9. Takagi, T., & Sugeno, M. (1985). Fuzzy identification of systems and its applications to modeling and control. *IEEE transactions on systems, man, and cybernetics*, (1), 116-132.

VI Workshop Innovación en Educación en Informática (WIEI)

Coordinadores

Sanz, Cecilia – Depetris, Beatriz - De Vincenzi, Marcelo –
Cukierman, Uriel

Comité Científico

Baldassarri, Sandra (España)	Marco, Javier (España)
Cukierman, Diana (Canadá)	Motz, Regina (Uruguay)
De Giusti, Armando (Argentina)	Pesado, Patricia (Argentina)
Feierherd, Guillermo (Argentina)	Roig Vila, Rosabel (España)
Manresa-Yee, Cristina (España)	Simari, Guillermo (Argentina)

“Programar Jugando” con Scratch en escuelas rurales de la provincia de Catamarca.

Sosa Chasampi Cintia Stella Maris¹, Jais Carlos Elias¹, Gonzales Marcelo¹,
Tolozza Eduardo Adrián²

- ¹ Departamento de Informática y Comunicaciones, Universidad Nacional de Catamarca.
Facultad de Ciencias Exactas y Naturales. Belgrano 300, Catamarca, Argentina
- ² Departamento de Formación General y Pedagógica, Universidad Nacional de Catamarca.
Facultad de Ciencias Exactas y Naturales. Belgrano 300, Catamarca, Argentina
{cintia, carlos, mgonzalez}@exactas.unca.edu.ar, eduardoatolozza@gmail.com

Resumen. La Facultad de Ciencias Exactas y Naturales dependiente de la Universidad Nacional de Catamarca - UNCA (ARGENTINA), generadora de profesionales en las ciencias de la computación, propone una continua evaluación de resultados en el dictado de materias relacionadas con la programación, y reconoce la importancia que ha tomado esta área dentro del continuo desarrollo tecnológico en el que estamos actualmente, al generar proyectos que impulsen la mejora de la calidad educativa de los alumnos. Este trabajo expone la finalidad, objetivos perseguidos y actividades realizadas en el proyecto “Programar Jugando” en el cual se dictan talleres de programación adaptados a los alumnos del nivel inicial, con el fin de incentivar el desarrollo de habilidades en la resolución de problemas y uso de un lenguaje de programación acorde a su edad, de manera que puedan incorporar la programación como parte de su formación básica, y así fortalecer a futuro su actuación en el ámbito universitario.

Palabras Clave: Aprendizaje, Computación, Educación Primaria, Lenguaje de Programación, TICs.

1 Introducción.

La programación, una de las áreas más importantes de las Ciencias de la Computación, es el proceso de crear programas que realicen una tarea específica o resuelvan un problema, utilizando un lenguaje especial de computadora. Este tipo de desarrollos se vuelven complejos, ya que implican un cambio en la forma de pensar y el aprendizaje de un lenguaje específico de computadoras. Para ello es necesario el estudio desde el punto de vista lógico y formal de las computadoras que hacen posible su funcionamiento [1].

La Facultad de Ciencias Exactas (FaCEN), actualmente posee las carreras de Profesorado en Computación y Tecnicatura en Informática las cuales incluyen en su plan de estudio materias relacionadas con la programación. Pero un trabajo de análisis de rendimiento de las materias afines a programación de las carreras de la FaCEN

demostró que el alumno aspirante ingresa a la facultad con conocimientos casi nulos respecto las ciencias de la computación y sobre todo en conceptos de programación [2]. A esta problemática se suma el hecho de que las carreras de tecnicaturas prevén pocos años de cursado, lo que lleva a que el alumno no tenga el tiempo de maduración suficiente para llegar a comprender y aplicar los fundamentos de programación y así lograr un desarrollo más profundo y de mayor calidad de lo aprendido.

Esta problemática motivó la formación de un equipo de trabajo que realice tareas de análisis e investigación de herramientas disponibles actualmente, para favorecer el aprendizaje de la programación, y además a hacer extensión de las actividades realizadas dentro de la FaCEN, e intentar que el potencial alumno universitario incorpore nociones de programación en su formación previa a la educación superior, facilitando de esta manera la inserción futura a las carreras de computación. Además permitiendo al alumno que tiene vocación en áreas distintas a la computación, una comprensión más profunda de estas ciencias para su posterior formación.

En la era actual, la alfabetización digital debería formar parte integral e indisoluble de la educación básica, debiendo centrarse en el alumno y su entorno. En igual sentido, Buckingham [3] afirma que la alfabetización sobre medios, es tan importante como la alfabetización tradicional. De acuerdo a la OCDE [4] la alfabetización digital no es simplemente el uso de dispositivos digitales, sino que requiere una serie de competencias como el adecuado manejo de la información, la capacidad de evaluar, organizar, y aprender a través de las redes; entre otras capacidades semióticas, cívicas, y comunicativas [5].

Sosteniendo estos principios, se diseñó el proyecto “Programar Jugando”, enmarcado en un proyecto de voluntariado universitario convocado por la Secretaría de Política Universitarias del Ministerio de Educación de Argentina. El proyecto fue desarrollado en el marco de la convocatoria anual 2015, y llevado adelante por docentes y alumnos de la UNCA, en escuelas del nivel primario y secundario en San Fernando del Valle de Catamarca; y particularmente en escuelas rurales del interior de la provincia ¹.

La expectativa de logro fue ayudar a los alumnos a dar sus primeros pasos en los lenguajes de programación, particularmente con Scratch, que fue el lenguaje seleccionado para tal fin. Entendiendo los potenciales beneficios que puede aportar el aprendizaje de un lenguaje de computadoras en los niños. Programar en una computadora o dispositivo electrónico, implica generar estructuras de pensamiento, poniendo en juego la imaginación y creatividad imbuidas por la personalidad de cada individuo, y convertirlas en algo visible a través de la construcción, mediante el lenguaje de programación, para generar algo nuevo. Esta actividad, como afirman Kafai y Burke [6] posibilita desarrollar el pensamiento abstracto, lógico-matemático, y algorítmico; mejora la capacidad de previsión, la comprobación de resultados, la optimización de recursos, y la toma de decisiones. En definitiva, enseña a pensar, posibilitando además el aprendizaje colaborativo.

¹ Resolución Ministerial ME N° 83/15. Secretaría de Políticas Universitarias.

2. Marco Teórico

Como afirman Miller, Latham y Cahill [7] el sistema educativo ha perdido paulatinamente eficacia alejándose de las necesidades de los alumnos, sus estilos de aprendizaje, perspectivas y posibilidades de involucrarlos activamente a través de sus preferencias personales; provocando una desconexión del mundo real. Estos niños que en muchos casos emplean medios tecnológicos de maneras superfluas, desaprovechando las reales posibilidades de aprender en profundidad el funcionamiento de dichos dispositivos y lenguajes, son los destinatarios del proyecto.

Como sostienen Nelson y Braafladt [8], la tecnología innovadora no hace magia, pero puede ayudar a los niños a ser expertos en navegar por un mundo cada vez más conectado, a la vez que les ayuda a desarrollar razonamiento deductivo, matemáticas y otras habilidades vitales de lecto-escritura comprensiva. Como disciplina, la programación está orientada al desarrollo de una serie de habilidades de abstracción y operacionalidad. El primer tipo de habilidades incluye técnicas como la simplificación de problemas, la definición de soluciones generales aplicables a problemas similares, y la asignación de nombres significativos a las distintas partes de una solución. El segundo tipo de habilidades, las operacionales, suponen la definición de soluciones en términos de un conjunto de pasos que deben ejecutarse en un orden determinado para alcanzar un objetivo [9].

La innovación por otra parte no debe necesariamente ser cara, y generalmente tiene más que ver con el cambio de actitudes y el verdadero aprovechamiento de herramientas disponibles que suelen estar sin emplear o subutilizadas, creando nuevas oportunidades para los alumnos.

Una de las herramientas más simples y potentes desarrolladas para la programación se llama Scratch. Un lenguaje de programación visual, gratuito y fácil de usar, desarrollado por el Grupo Lifelong Kindergarten del MIT (Instituto Tecnológico de Massachusetts) Media que se puede utilizar para crear todo tipo de aplicaciones, desde animaciones hasta presentaciones y hasta juegos. Destinada principalmente a los niños entre los 8 y 16 años, aunque es usado por personas de todas las edades, y les permite explorar y experimentar los conceptos de programación mediante el uso de una sencilla interfaz gráfica, desarrollar proyectos y comunicar ideas [10].

Este tipo de aplicaciones que transforman los comandos del lenguaje en bloques encastrables es usado actualmente como herramienta didáctica para estudiantes y personas que no poseen mucha experiencia en el desarrollo de código en un lenguaje específico y esto facilita el aprendizaje del mismo. Existen diferentes aplicaciones similares en su filosofía de trabajo a Scratch. Por ejemplo “pilas bloques” desarrollada por Program.AR - Fundación Sadosky o MIT “app inventor” Una herramienta para la programación y la creación de aplicaciones para dispositivos con sistema operativo Android.

Tanto Scratch como las otras mencionadas poseen la particularidad, o la característica común de convertir un lenguaje de programación en una aplicación visual para resolver un problema, permitiendo programar utilizando bloques prediseñados con una acción o propiedad en particular. Arrastrando estos bloques de código hacia el área de trabajo para formar bloques de acciones más complejas, identificando grupos de bloques del mismo tipo con diferentes colores, eliminando la

dificultad que presenta la sintaxis de un lenguaje de programación escrito [10]. Pero la gran ventaja de Scratch es su gran difusión en el mundo lo que posibilita conseguir fácilmente mucho material y código fuente para trabajar y continuar aprendiendo.

Desde el año 2013, Scratch 2 está disponible en línea y como aplicación de escritorio para Windows, OS X y Linux. El código fuente de Scratch 1.x se liberó bajo licencias GPLv2 y "Scratch Source Code License"[11].



Fig. 1. Ejemplo de programación en bloques.

En el ámbito de la ingeniería informática, scratching quiere decir reutilizar código, esta característica dentro de la programación es muy beneficiosa y efectiva lo cual es una clave de Scratch ya que permite que los usuarios puedan descargar y trabajar sobre proyectos públicos subidos y desarrollados por otros usuarios y compartidos para toda la comunidad de programadores de Scratch.

Algunas ventajas que se destacan [11]:

- Permite desarrollar el pensamiento lógico algorítmico.
- Permite desarrollar métodos para solucionar problemas de manera metódica y ordenada.
- Permite desarrollar el hábito de hacer autodiagnos con respecto a su trabajo.
- Permite tener la posibilidad de obtener resultados complejos a partir de ideas simples.
- Permite trabajar cada cual a su ritmo en función de sus propias competencias.
- Permite aprender y asumir conceptos matemáticos: coordenadas, variables, algoritmos, aleatoriedad.
- Permite aprender los fundamentos de la programación.
- Permite usar distintos medios: sonido, imagen, texto, gráfico, etc.

- Posibilita el aprendizaje colaborativo a través del intercambio de conocimiento.

3. Antecedentes

Existen innumerables experiencias de aprendizaje en niños trabajando con la aplicación Scratch. Destacándose algunos como el expuesto por Nelson y Braafladt [8], llevados adelante en diversas bibliotecas públicas de Estados Unidos (Free Library of Philadelphia, Memphis Public Library, y Seattle Public Library) demostraron excelentes resultados empleando escasos recursos económicos en la preparación de los jóvenes para su futuro mundo universitario o laboral. Haciendo hincapié en el desarrollo de la educación no-formal en las bibliotecas públicas. Mientras que en países como Francia, Estonia, Reino Unido, Finlandia, Alemania, y Australia han institucionalizado en la curricula nacional, la enseñanza de la programación en los niños y el empleo transversal de Scratch.

En Argentina, a nivel nacional se presentaron algunas iniciativas que pretenden solventar esta problemática convirtiendo el aprendizaje de programación en un juego para niños. Una de ellas es la iniciativa del Program.AR creado en el año 2013 y es llevado adelante por la Fundación Sadosky del Ministerio de Ciencia, Tecnología e Innovación Productiva propiciando que aprendizaje significativo de Ciencias de la Computación llegue a las escuelas argentinas. Entendemos que una comprensión profunda acerca del funcionamiento de tecnología informática es una herramienta fundamental para la construcción de ciudadanía en el presente [12].

En la FaCEN surge entonces la iniciativa de incentivar el aprendizaje de este tipo de habilidades, pero desde la formación primaria, de modo que el alumno pueda incorporar de manera natural y pueda hacer uso de estas capacidades en todos los aspectos de su formación. Hasta el momento de presentación del proyecto no existía dentro de la planificación del nivel primario ni secundario, materias específicas relacionadas con la programación, y esta problemática se hace más profunda con el hecho de que existen en el interior de nuestra provincia establecimientos escolares en los cuales los alumnos no tienen acceso directo a tecnologías como Internet; y esto dificulta aún más incorporar dichos conocimientos a los sectores tradicionalmente excluidos. De manera que, si realmente aspiramos hacia la gestión de calidad y la excelencia en el ámbito educativo, debemos partir de las necesidades personales de los ciudadanos y de las necesidades sociales del territorio [13].

4 Descripción General del Proyecto

4.1 Objetivo General

Brindar mediante estrategias pedagógicas adecuadas a las edades de los alumnos, nociones básicas de programación a través de actividades que les permitan aprender un lenguaje de programación de manera divertida y simple.

4.2 Objetivos Específicos

- Analizar diversos entornos actualmente disponibles para el aprendizaje de conceptos de programación para niños, con el fin de determinar el más adecuado a utilizar.
- Diseñar estrategias de enseñanza centradas en el alumno como individuo
- Organizar junto a las autoridades del establecimiento educativo, talleres sobre conceptos básicos usados en programación, pero adaptados a la edad de los alumnos.
- Concientizar a los directivos y docentes de los establecimientos integrados al proyecto, respecto a la importancia del aprendizaje de un lenguaje de programación.
- Desarrollar material didáctico que permita el apoyo a los talleres.
- Desarrollar talleres para que los niños aprendan jugando a usar una herramienta simple para el desarrollo de sus primeros programas.
- Facilitar el aprendizaje de todo aquel que esté interesado en aprender a programar
- Incentivar el aprendizaje de nuevas tecnologías a los niños del nivel primario.
- Difundir la totalidad de las actividades desarrolladas en el proyecto a través de sitios webs, publicaciones en blogs, redes sociales en Internet.

4.3 Metas Propuestas.

- Formar un equipo de trabajo con docentes y alumnos, con la capacidad necesaria para conseguir la construcción de los objetivos del proyecto.
- Lograr que los alumnos puedan descubrir la capacidad de resolver problemas utilizando un entorno de programación y puedan disfrutar de la realización de diferentes programas.
- Incentivar a los alumnos para que puedan descubrir la capacidad de resolver problemas utilizando un entorno de programación y puedan disfrutar de la realización de diferentes programas.

5. Metodología

5.1 Análisis de Requerimientos

El grupo de trabajo integrado por alumnos y docentes de la FaCEN encargados de dictar las materias afines a programación y alumnos avanzados de las carreras vinculadas al proyecto se entrevistó con los directivos y docentes de cada escuela a fin de tomar conocimiento sobre las problemáticas propias de cada escuela. Para posteriormente diseñar las estrategias de intervención educativa más adecuadas, siempre centrándose en el perfil de los alumnos y su contexto.

A continuación se inició el desafío de seleccionar y re- diseñar los contenidos vinculados a la programación, y realizar análisis sobre dichos entornos, estudiar guías de ejercicios prácticos y manual de contenidos, con el fin de desarrollar luego material de referencia para profundizar posteriormente, los conceptos desarrollados.

Al encarar el gran desafío de actualizar los contenidos vinculados a la programación fue necesario invertir tiempo y esfuerzos por parte de los docentes, además se desarrolló experiencia e intercambio de opiniones y conceptos, fundamental para la creación de un grupo de trabajo capacitado para guiar al alumno en sus inicios en la programación, generando un nuevo enfoque en la enseñanza de la informática y que el alumno pueda hacer un uso más profundo de la tecnología.

5.2 Desarrollo de Material de Referencia

El equipo de trabajo desarrollo material de referencia para ser usado en el dictado de los talleres a los alumnos, y la ardua tarea de escribir conceptos básicos de programación de una forma sencilla, teniendo en cuenta la edad de los alumnos destinatarios con ejemplos para que sus contenidos puedan ser incorporados con facilidad, de manera que puedan ir aprendiendo estos conceptos de programación de manera gradual.

Los contenidos del material de referencia se presentaron en forma de secuencias didácticas, ejercicios y la realización de actividades que permiten la integración de los conceptos involucrados.

Las guías de ejercicios y manuales desarrollados incluyeron: Introducción al pensamiento lógico, lenguajes de programación, instrucción, secuencia de instrucciones, sentencias de selección, sentencias de repetición y resolución de problemas simples. Todos los conceptos dados, fueron presentados teniendo en cuenta la edad y conocimientos de los alumnos, de manera que puedan incorporar dichos conceptos sin mayores inconvenientes. Además los ejercicios fueron propuestos incorporando actividades y elementos propios del lenguaje scratch que permitió que el alumno se sienta motivado dada las características animadas que posee SCRATCH.

5.3 Beneficiarios

Para el desarrollo de esta parte del proyecto, se seleccionaron dos escuelas pertenecientes al interior de la provincia de Catamarca. Las mismas contaban con ciertas estructuras tecnológicas, pero se encuentran subutilizadas, de ahí la potencialidad para aprovechar y trabajar con dichas herramientas. Los establecimientos seleccionados fueron, la Escuela N° 10 Dr. Antonio del Pino perteneciente a la localidad de Copacabana, Escuela Primaria 233 "hipolito Vieytes" perteneciente a la localidad de El Salado, ambas del Departamento Tinogasta y Colegio Manuel Belgrano, Institucion ubicada en la zona urbana central de la capital de catamarca.

Los beneficiarios directos de aplicar lo planteado en el presente proyecto fueron los alumnos y docentes del nivel primario y secundario de las escuelas seleccionadas los cuales recibieron diversos talleres donde se les explicó de manera sencilla, conceptos introductorios sobre programación, resolución de problemas básicos usando un lenguaje natural y uso de la plataforma SCRATCH para la resolución de problemas con mayor complejidad.

Tabla 1. Instituciones seleccionadas para el desarrollo del proyecto “PROGRAMAR JUGANDO”. Tinogasta es una ciudad capital del departamento homónimo, en el oeste de la provincia argentina de Catamarca; a 271 km de la capital provincial San Fernando del Valle de Catamarca

Nombre de la Escuela	Ubicación
Escuela N° 10 “Dr. Antonio del Pino”	Copacabana, Tinogasta, Catamarca, Argentina
Escuela Primaria 233 “Hipólito Vieytes”	Ruta Nacional 60 - Salado, Tinogasta, Catamarca, Argentina
Instituto “Manuel Belgrano”	Junin 960. San Fernando del Valle de Catamarca, Catamarca, Argentina

El dictado de los talleres fue complementado con el material didáctico diseñado por el equipo de trabajo, el cual fue entregado a cada alumno y docentes del área tecnología de las escuelas destinatarias del proyecto.

6 Resultados

El equipo de trabajo dictó una serie de talleres de manera presencial en las instituciones seleccionadas, haciendo uso de las instalaciones propias de cada

institución y entregando a cada alumno todo el material de consulta y guías de ejercicios prácticos desarrollados.

Existen diferentes aspectos en base a los resultados obtenidos en el tiempo de trabajo que se consideran como puntos relevantes y de referencia que permiten evaluar el proyecto y nos aseguran el éxito del mismo:

- Los Docentes pertenecientes a las Instituciones Participantes, mostraron amplio interés en los talleres desarrollados, y participaron activamente de los mismos coordinando los alumnos en su rol de educador, aprendiendo de los mismos en su rol de profesional que se capacita continuamente.
- Los Alumnos destinatarios mostraron un gran entusiasmo en el dictado teórico de los talleres, en el que se les dió participación continua, enseñándoles los contenidos planteados, y brindándoles material de referencia.
- En el desarrollo práctico de los talleres, donde utilizaron la plataforma SCRATCH seleccionada como lenguaje de programación, los alumnos desarrollaron guías de ejercicios propuestos sin grandes complicaciones, mostrando diversión y entretenimiento, ya que el entorno les resultó fácil y divertido de utilizar.
- Los directivos de las Instituciones participantes se mostraron satisfechos con la actuación del equipo de trabajo del proyecto dentro del establecimiento, solicitando además una continua inclusión en futuros proyectos de similares características.
- El material de referencia y guías de ejercicios realizados por el equipo de trabajo constituido para el proyecto, forma parte del material de estudio de las organizaciones participantes. Pero también puede ser empleado por cualquier persona sin formación previa en el ámbito de las ciencias de la computación e interesadas en capacitarse en el área de programación, ya que el mismo se encuentra disponible en línea en la página principal de la FaCEN

6 Conclusiones y Futuras Líneas de Trabajo

Continuamente los profesionales de las ciencias de la computación formados en nuestras universidades generan grandes avances tecnológicos que impactan día a día en nuestra sociedad. Aun así creemos que es posible mejorar la calidad educativa y la cantidad de profesionales que se forman, si logramos realizar cambios que impulsen a que el alumno comience a desarrollar habilidades en este campo desde más temprana edad, y además implantando estos cambios en sectores tradicionalmente excluidos mediante estrategias que faciliten el acceso a la información y posibiliten una verdadera inclusión social. El uso de este lenguaje en los talleres desarrollados posibilitó el desarrollo de la creatividad y el razonamiento a través del desarrollo de animaciones, juegos o historias interactivas.

Los resultados expuestos en este trabajo pueden ser tomados como referencia para los proyectos vigentes actualmente, donde se pretende incorporar a la planificación de la enseñanza primaria materias inherentes a la programación que hoy en día no están contempladas, y de esta manera sirvan como base para el desarrollo de capacidades en

los niños y ayudar a la alfabetización de la sociedad en lo que hace a la programación, aspecto cada vez más relevante y necesario en los tiempos actuales.

Como trabajo futuro se identifica: (a) la necesidad de estudiar el desarrollo de nuevas herramientas de aprendizaje de la programación para lograr una continua motivación para los alumnos; y (b) lograr la implementación de una planificación permanente de los contenidos de programación propuestos en el sistema educativo actual.

Referencias

1. Mitch Resnick. “¿Por qué aprender programación en las escuelas?”. Revista Electrónica Educación 3.0, vol. 1, no. 23, (2015), <http://www.educaciontrespuntocero.com/noticias/por-que-aprender-programacion-en-las-escuelas-habla-mitch-resnick-en-su-charla-ted/29924.html>.
2. Cintia S. M. Sosa Chasampi. “Aprendiendo a Programar”. Semana de Puertas Abiertas de la Facultad de Ciencias Exactas y Naturales: los desafíos del nuevo siglo, vol 1, no 1, pp.151-153, (2015).
3. Buckingham, D. Educación en medios: alfabetización, aprendizaje y cultura contemporánea. Paidós. España. (2005).
4. OCDE Annual report. Organisation for economic co-operation and development. (2003)
5. Pérez Tornero, J. M. Digital literacy and media education: an Emerging Need. Academic Press (2004).
6. Kafai, Y., Burke, Q. Connected Code. Why Children Need to Learn Programming. MIT Press. (2014)
7. Miller, R., Latham, B. Cahill, B. Humanizing the Education Machine: How to Create Schools That Turn Disengaged Kids Into Inspired Learners. Wiley & son. Hoboken. USA. (2016)
8. Nelson, J., Braafladt, K. Technology and Literacy: 21st Century Library Programming for Children and Teens. American Library association. Chicago. USA. (2016)
9. Pablo Factorovich y Federico Sawady O'Connor. “Actividades para aprender a Program.AR”. 1a ed. - Ciudad Autónoma de Buenos Aires: Fundación Sadosky, E-Book. - Program.AR. (2015)
10. Marina Umaschi and Mitch Resnick “The Official ScratchJr Book: Help Your Kids Learn to Code”. Impreso en USA, (2016).
11. Plataforma Scratch. Acerca de: <https://scratch.mit.edu>, (2017).
12. Fundación Sadosky, “¿Por qué enseñar programación en la escuela? Disponible en: <http://program.ar/> (2016).
13. Ana Iglesias Rodríguez. “PLANIFICACIÓN y ORGANIZACIÓN en la educación inclusiva”, Aspectos clave de la Educación Inclusiva, Colección Investigación, Salamanca, (2009).

¿Invirtiendo la clase o invirtiendo en la clase? Enseñanza de tecnología para la producción multimedia mediante Aula invertida

Fernández, Mirta ^{1,2} - Godoy Guglielmone Maria V.², Mariño, Sonia I.² Barrios Walter G².

¹ Facultad de Artes, Diseño y Ciencias de la Cultura (FADyC)

² Facultad de Ciencias Exactas y Naturales y Agrimensura (FaCENA)
Universidad Nacional del Nordeste (UNNE)

{mirtagf@hotmail.com, mvvg2001@yahoo.com, simarinio@yahoo.com,
waltergbarrios@yahoo.com.ar}

Resumen. Haciendo alusión a los enfoques que han primado en la educación en general -y artística en particular-; en este trabajo, se interpela la introducción de métodos aplicables con el objetivo de generar interés entre jóvenes estudiantes y formarlos a través del Aula Invertida: un enfoque activo, centrado en el estudiante que surge de la premisa de extender el tiempo de una actividad con el objetivo de favorecer el pensamiento crítico y autonomía en el aprendizaje. Al respecto, se ponen en relieve algunas de las teorías de aprendizaje vinculadas; se plantea la propuesta; se detallan las decisiones didácticas, curriculares y tecnológicas. Se plasman resultados de su implementación en una asignatura de primer año de la Facultad de Artes de la Universidad Nacional del Nordeste para propiciar la enseñanza de una herramienta software en la realización de una narrativa multimedia interactiva. Finalmente se extraen conclusiones, sintetizadas en el impacto significativo del aprendizaje para la realización de las producciones en cuestión y derivaciones subyacentes.

Palabras Clave: Inversión en el aula, Aprendizaje ubicuo, Aplicación del modelo TPACK, Metodologías de aprendizaje activo, Evaluación del método Aula Invertida.

1. Introducción

Los avances recientes en tecnología e ideología han abierto nuevas direcciones para la investigación educativa buscando eliminar las barreras espaciotemporales [1]. Propiciada por el aumento de materiales y recursos distribuidos en la red, se está generando discusión y catalizando el cambio en el aula física.

Por otra parte, haciendo alusión a los enfoques que han primado en la educación en general -y artística en particular, tal como lo manifiestan Aguirre & Giráldez [2]; en este sentido interpela a la introducción de métodos aplicables con el objetivo de generar interés entre jóvenes estudiantes y formarlos a través del Aula Invertida: un enfoque activo, centrado en el estudiante que surge de la premisa de extender el tiempo de una actividad con el objetivo de favorecer el pensamiento crítico y la autonomía en el aprendizaje.

Referido a un enfoque de enseñanza en el que las actividades tradicionales de clase y de autoaprendizaje se invierten o se "vuelcan", el Aula Invertida o en inglés Flipped Classroom se presenta como un enfoque pedagógico, que emplea –en general– conferencia pregrabada o emprendiendo lecturas y actividades guiadas, seguido del tiempo de clase que se utiliza para la discusión interactiva, la resolución de problemas y otras actividades con el profesor [3], [4] y [5]. Como tal, el papel del docente cambia de ser el "sabio en el escenario" a un "guía" en el proceso. En [3], [6] y [7] se mencionan aspectos que favorecen la inclusión del Aula Invertida e inconvenientes que comúnmente se generan. Como ventajas se señalan:

- un aumento en la interacción entre estudiantes y profesores;
- un cambio en la responsabilidad de aprender sobre los estudiantes;
- la capacidad de los estudiantes para prepararse en un momento que les convenga, y tantas veces como satisfaga sus necesidades;
- una serie de recursos didácticos disponibles en cualquier momento y lugar;
- el trabajo colaborativo entre estudiantes;
- un aumento en la participación de los estudiantes y un cambio de la escucha pasiva al aprendizaje activo;

y, como posibles desventajas se mencionan:

- la inversión de tiempo y recursos para desarrollar cursos;
- la posible necesidad de inversión tecnológica;
- el tiempo para que los profesores y los estudiantes se adapten y adquieran las nuevas habilidades requeridas para este acercamiento más activo y autodirigido al aprendizaje.

Por ello, se asiente que la clave del éxito está en que los estudiantes asuman la responsabilidad de su aprendizaje y asistan preparados a la clase. Lo expuesto puede ser visto como una ventaja y una desventaja.

Según lo antes mencionado, en este trabajo se plasma una experiencia sustentada en las ventajas que proporciona implementar el Aula Invertida como estrategia de aprendizaje activo y en reconocer los diversos ritmos de aprendizajes en un grupo numeroso de alumnos de primer año de la asignatura Introducción a las Tecnologías Aplicada al Arte de la carrera Licenciatura en Artes Combinadas de la Facultad de Artes de la Universidad Nacional del Nordeste (UNNE). La propuesta se enmarca en el desarrollo de un trabajo práctico, en que el propósito principal es la elaboración de una narrativa multimedia mediante la integración de múltiples recursos (gráficos, imágenes en movimiento, sonido, botones, entre otros); las cual se desplegaron en etapas parciales: de análisis e interpretación, y de aplicación. Se realizó el seguimiento de las tareas mediante la tutorización en el aula Virtual y se retroalimentó la experiencia a través de un cuestionario que recogió la percepción de los participantes. En base a ello se proyectan la incorporación de algunas mejoras en el abordaje.

1.1. Bases teóricas y terminologías vinculadas al Aula Invertida.

Bishop & Verleger [1] en su estudio, explican como la metodología Aula Invertida o Inversa, representa una combinación de teorías de aprendizaje: actividades basadas en una ideología constructivista e instrucción derivadas de métodos bajo los principios conductistas. Es posible adherirse a ellos retomando la teoría de la Zona de Desarrollo Proximal (ZPD) de Vigotsky [8], que explica cómo el aprendizaje de los estudiantes depende de sus conocimientos previos en el área y cómo se articulan; explicitando, que requieren de la ayuda del educador para alcanzar todo su potencial, además de resaltar la importancia de interacción con sus pares, un andamiaje y una retroalimentación apropiada.

Por su parte, sus difusores y promotores Aaron y Bergmann [4], señalan que es simplemente girar el método tradicional en que el contenido educativo se presenta en el aula y las actividades de práctica se realizan en casa como se muestra en la **Fig. 1**. En este caso los alumnos reciben la "lección magistral" en casa y realizan la tarea en el aula, previamente pensada y planificada a tales fines.



Fig. 1. Esquema representativo de Aula Invertida.

En 2014, *The Flipped Learning Network (FLN)* [9] aportó una definición universalizada además de cuatro pilares muy difundidos, sintetizados y traducidos en la **Fig. 2**. Los cuales pretenden servir de guía para sus implementaciones.



Fig. 2. Los cuatro Pilares del Aula Invertida según FLN [9].

Aún con las numerosas experiencias existentes, una de las principales críticas al modelo es que invertir una clase puede, no necesariamente, conducir al Aprendizaje Invertido [9]. Esto está relacionado con la adecuada preparación y uso de las herramientas tecnológicas en las actividades "fuera del aula" [10], es decir, que se requiere emplear el tiempo de la clase de la manera adecuada y con la intencionalidad que promueve el modelo [4], [11].

1.2. Evaluación de la metodología

Siguiendo a Acuña [6] y Fernández & otros [7], las dimensiones que pueden considerarse las más aceptadas a la hora de valorar los efectos de invertir el aula, son:

- i. experiencial (del proceso de aprendizaje),
- ii. respecto al uso del video como mecanismo de instrucción,

Asimismo, es posible pronosticar que la combinación de nuevas tecnologías de enseñanza con actividades interactivas en el aula puede resultar en un mejor aprendizaje, pero desfavorable en la satisfacción de los estudiantes [12]. En este sentido Persky [13], comparte algunos de sus hallazgos, vinculados a la resistencia al cambio que viven los estudiantes: según su experiencia, aunque los estudiantes aprenden más, no les agrada el curso y afirma que mientras más utilizó Aprendizaje Invertido en sus clases, más bajó su evaluación de docente. Asintiendo que el descontento y reacción desfavorable de los estudiantes se puede deber a la herencia inherente del modelo tradicional de enseñanza en el que el profesor es el responsable de "enseñar" y "transmitir" los conocimientos. Así, la expectativa del estudiante tradicional es depositar en el profesor la responsabilidad de adquirir nuevos aprendizajes y conocimientos [14].

1.3. Resumiendo, ¿de que se trata Aula Invertida?

La literatura existente coincide que se crea un entorno de Aprendizaje Invertido, cuando los maestros elaboran o seleccionan lecciones disponibles, accediendo cuando y donde sea conveniente para el estudiante, en casa, en clase, durante la sala de estudio, en el autobús a un juego o incluso desde una cama de hospital [5]; aspectos que según Burbules [15] definen ubicuidad en enseñanza. Los maestros pueden entregar esta instrucción grabando y narrando presentaciones, creando videos de sí mismos enseñando, o seleccionando tutoriales de sitios confiables de Internet. Los estudiantes pueden ver los videos o capturas de pantallas tantas veces como lo necesiten, permitiendo mejorar su productividad en el aula.

Tal como se reseña en este trabajo, la clase pasa de ser de recepción pasiva a ser de participación más activa. El profesor, al liberarse del tiempo de clase que utiliza para la instrucción a los estudiantes, puede centrarse en la atención de las necesidades puntuales de aprendizaje y dar respuesta a la diversidad y a los distintos ritmos de aprendizajes.

1.4. Motivación en la selección del método

Una hipótesis que da origen a desplegar un cambio de paradigma en el aula es el tiempo que insume a los estudiantes de primer año familiarizarse con algunas tecnologías informáticas, en las que se ponen en juego habilidades y sensibilidades particulares. Una experiencia estrechamente vinculada a tal inferencia se presenta en Algonquin College [16], en una clase de producción de vídeo se utilizó Aula Invertida para enseñar el funcionamiento del software de edición, procedimiento engorroso de explicar en una clase convencional.

Lo antes mencionado entre otros factores vinculados, promueve la introducción de variaciones en el modo de abordar una propuesta didáctica en la asignatura Introducción a las Tecnologías Aplicada al Arte, en el año 2016 a fin de lograr un aprendizaje más significativo aprovechando las ventajas de "invertir" el aula. Como se ilustra la **Fig. 3.**, los roles y las responsabilidades de los participantes se vuelven más proactivos.

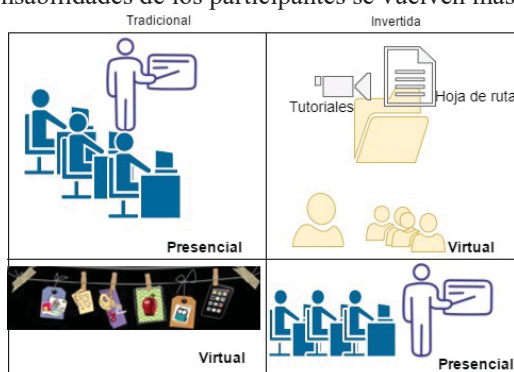


Fig. 3 Migración hacia Aula Invertida en la experiencia

La metáfora ilustra a la izquierda los componentes incorporados para la estrategia tradicional y a la derecha los recursos desplegados para brindar el andamiaje necesario a las tareas.

1.5. Pensar antes de actuar: utilización del modelo TPACK en el rediseño didáctico

En un proceso continuo de toma de decisiones en torno a los distintos elementos del currículo, una alternativa propicia es recurrir al modelo TPACK (por sus siglas Technological Pedagogical Content Knowledge) [17]. El modelo, permite enfatizar la dimensión creativa-constructiva de la preparación y desarrollo de procesos de enseñanzas, el rol del profesor como facilitador de entornos, la explicitación y discusión en torno a esos elementos para integrar las TIC de una forma eficaz [18], [19]. Tal como se ilustra en la **Fig. 4**, las orientaciones dan lugar a una serie de decisiones para tales fines.

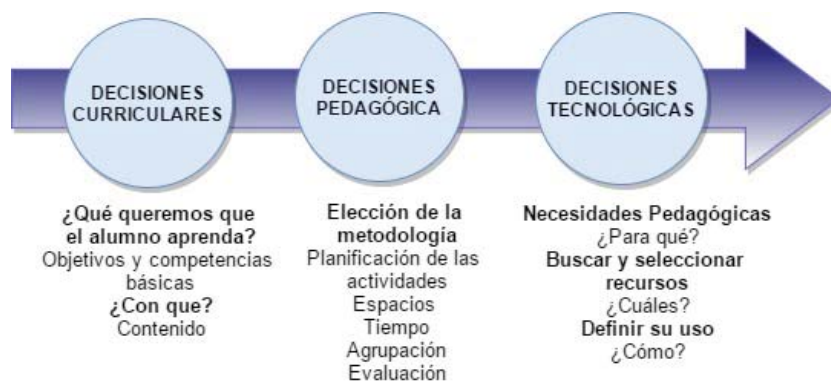


Fig. 4. Implicancias en la utilización de TIC según TPACK []

El marco TPACK y siguiendo a Harris & Hofer [20], aportaron al presente trabajo orientaciones para el rediseño de tareas, de acuerdo a las actividades con TIC o Learning Activity Types (Tipos de Actividades de Aprendizaje). De las mismas se tomaron las vinculadas a las Artes Visuales¹: Construir conocimiento/Conceptualizar y Aplicar, en el contexto curricular [21].

2. Metodología

En la elaboración didáctico y pedagógica, para la integración de tecnología en la enseñanza, se siguió los lineamientos que propone TPACK: se plantearon decisiones curriculares, pedagógicas y se realizó la selección de recursos tecnológicos. Posteriormente, se analizaron resultados y se validó la propuesta a fin de obtener retrospectiva por parte de los alumnos.

3. Resultados

La asignatura se presenta en el contexto del primer año. En la estructura curricular la temática a abordar originalmente se pensó para el desarrollo de ejercitaciones en el aula de informática, mediante la utilización de una herramienta de software para la realización de un trabajo aplicado.

En dicho contexto, se identificó la principal debilidad de la metodología tradicional es la disparidad de conocimientos previos o de preparación de los alumnos al llegar al aula, para aprender de la misma manera. El tiempo que insume a los estudiantes familiarizarse con algunas tecnologías es heterogéneo en el caso de los primeros años. Asimismo, se identificaron otros factores vinculantes:

- alumnos poco familiarizados con tecnología en los espacios académicos,
- alta tasa de inasistencia,

¹ Visual Arts Activity Types: <https://activitytypes.wmwikis.net/Visual+Arts>

- escasez de tiempo y espacio para el desarrollo de las clases personalizadas de informática.

Sumado a lo expuesto precedentemente, se reconoció que los alumnos se abrumaban con la infinidad de material disponible en Internet. Por ello, conformaron redes sociales en que se compartían videos y recursos para dar respuestas puntuales a cada grupo de trabajo fuera de la clase. Lo cual traía a colación la imposibilidad de conocer por parte de los docentes, cada material en particular, para dar la respuesta esperada.

Los factores señalados, contribuyeron a la decisión de repensar los mecanismos utilizados, por ello en el año 2016, se revisaron los objetivos curriculares y se reelaboró la estrategia didáctica, mediante lineamientos que se describen a continuación.

3.1. Decisiones curriculares: Definición de objetivos, competencias y contenido a abordar

Las decisiones curriculares consistieron en la definición de un trabajo práctico integrador, que permitió alcanzar un objetivo dual de las etapas:

- De análisis e interpretación: Identificar vocabulario, conceptos, significados y metáforas presentes en una narrativa del Autor José Luis Borges “El jardín de senderos que se bifurcan”.
- De aplicación: Desarrollar representaciones artísticas, en forma individual o colaborativa, usando diferentes medios para elaborar una narrativa interactiva multimedia (textos, imágenes fijas y en movimiento, botones, sonido, gráficos, etc.)

Estos objetivos se plantean mediante actividades que desafían a los estudiantes a transferir sus conocimientos usando una variedad de recursos y técnicas, y desarrollando las competencias requeridas en la asignatura que contribuyen al perfil profesional delineado en la carrera.

3.2. Decisiones Pedagógicas: Planificación de las actividades

La planificación sustentada en el Aula Invertida se basó en una serie de clases presenciales en que se establecieron pautas específicas mediante una hoja de ruta, disponible posteriormente en el aula virtual de la asignatura para la realización de las actividades programadas:

- *Actividades en clases.*

Se comenzó con una breve exposición de la dinámica de trabajo, objetivos, tareas implícitas, incluyendo fechas de entregas según las etapas parciales de:

- Mapa de navegación de la narrativa a plasmar.
- Narrativa interactiva en el software recomendado.

Seguidamente se analizó de manera preliminar el texto a abordar como una manera de anticipar el trabajo en fuera del aula; las demás clases estuvieron destinadas a la realización de las producciones grupales.

- *Actividad post-clases.*

Consistentes en un conjunto de tareas compuestas por al menos tres tutoriales, con el propósito de reconocer el entorno de trabajo, realizar prácticas individuales, desarrollar destreza en el manejo la herramienta y avanzar en la consecución de los objetivos particulares.

La evaluación del trabajo se orientó a la obtención parcial de las entregas, de acuerdo con los objetivos propuestos y los tiempos estipulados.

3.3. Decisiones Tecnológicas: selección de herramientas y recursos

Las decisiones tecnológicas se centraron en la elaboración de una hoja de ruta, la utilización del espacio virtual, la selección de tutoriales y en la implementación en la herramienta.

En la hoja de ruta se proyectaron detalladamente los objetivos y tareas.

Se recomendó el uso de la herramienta on-line **draw.io** para la confección del mapa de navegación, el cual permite que sea editada colaborativamente.

Se utilizó el aula virtual como apoyo a las actividades previstas, se dispusieron los materiales, enlaces y recursos. Para dar seguimiento, se utilizó principalmente la configuración de Grupos (de alumnos), por Agrupamiento (por tutores) y Tareas o Espacios de entregas como se muestra en **Fig. 5**; con la posibilidad de hacer presentaciones parciales y de esta manera lograr que el alumno conociera las devoluciones con rapidez.

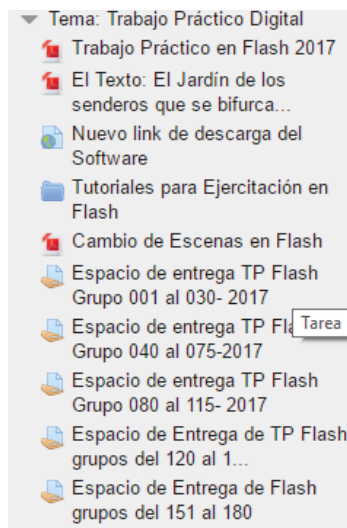


Fig. 5. Materiales, configuraciones de tareas y grupos en el aula virtual

Básicamente, los estudiantes debieron ejercitar tres tutoriales disponibles en la Web compuestos de una serie de instrucciones intermedias. La ventaja material prima entre otras:

- Se trata de sitio Web interactivo diseñado como un material con ilustraciones y la posibilidad de reproducir pequeños fragmentos de las instrucciones, como se muestra en la **Fig. 6**.
- Las instrucciones pueden obtenerse e imprimirse en formato texto (.pdf).
- Los primeros tres tutoriales brindan un andamiaje para la consecución de las tareas, con la posibilidad de extender con otros tutoriales y recursos del mismo sitio Web.
- Los textos e imágenes publicados del material en cuestión están sujetos a una licencia²² que permite copiarlos, distribuirlos y transmitirlos públicamente siempre que se cite el autor y la fuente.

Para la implementación en la herramienta, se continuó utilizando la misma de años anteriores (Adobe Flash CS 5).

Para la realización de encuesta de percepción, al final de la realización de los trabajos, se dispuso un formulario en Google Form, para recolectar las opiniones y posteriormente procesarlas.

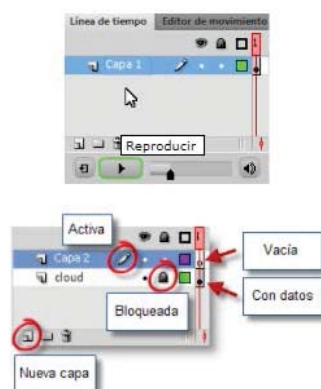


Fig. 6. Material Web interactivo seleccionado para su uso en Aula Invertida
(Fuente: <http://mosaic.uoc.edu/FlashCS5/cast/index.html>).

Las ventajas citadas brindaron -entre otras- la posibilidad al docente de unificar los materiales gestionados por los alumnos, y para éstos de accederlos en línea o descargarlos.

3.4. Evaluación de la propuesta de Aula Invertida

Respecto a la evaluación del proceso de aprendizaje, se propusieron dos dimensiones: del progreso y los logros de los alumnos.

En el progreso, los alumnos tuvieron dificultades de mayor de nivel de complejidad que excedían al material. Por ejemplo, en la utilización de determinado formato de sonido, en la incorporación de códigos o “script” para la interacción de botones y escenas, entre otros problemas de carácter “técnico”.

² *Reconocimiento-CompartirIgual 3.0 España (CC BY-SA 3.0 ES):*
<https://creativecommons.org/licenses/by-sa/3.0/es/>

En segundo lugar, vinculado a las capacidades e intereses de los alumnos y sobre la naturaleza del currículo, se observó diversidad de producciones con resultados satisfactorios en las etapas parciales. Algunos destacados como la grabación y edición de sonidos con herramientas ajenas a las propuestas, otros en la introducción de script que producía aleatoriedad en el recorrido de las escenas, algunos con enfoques lúdicos en la narrativa, entre otros.

En relación a los plazos, el aula virtual de la asignatura y los espacios destinados para tales fines, contribuyeron a la realización de entregas de manera organizada y en las fechas estipuladas.

3.5. Validación de la propuesta. Un caso de estudio en el año 2016.

Con el propósito de validar la propuesta, en el año 2016 se dispuso un formulario en línea y se seleccionó aleatoriamente una muestra de 60 alumnos. El sondeo se dirigió a reconocer las perspectivas y expectativas de los participantes en relación con la metodología utilizada, así como el uso de tutoriales; mediante cinco (5) interrogantes. En una escala del 1 al 4, en que uno (1) representó “en desacuerdo”, dos (2) “neutral”, tres (3) “de acuerdo”, en el otro extremo, el cuatro (4) “muy de acuerdo” se obtuvo:

- Respecto a “*me agradó la posibilidad de realizar un tutorial/ver un vídeo en lugar de asistir a una clase tradicional (exposición) de las herramientas del programa*”. Como se muestra en la Fig. 7, el 75% respondió entre la escala del 2 y el 4; en tanto que un 25% respondió que la opción 1 que representa “en desacuerdo”.

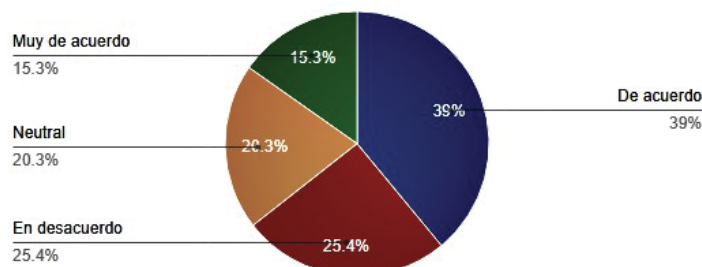
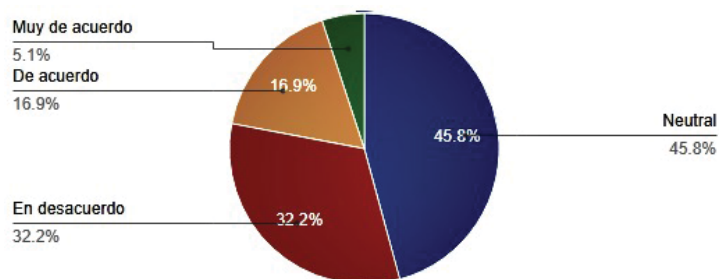
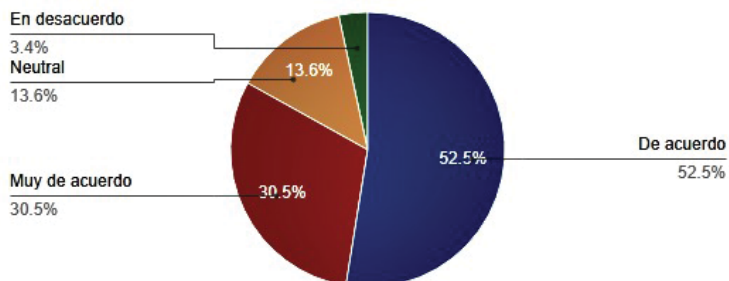


Fig. 7. Percepción del alumno respecto a la el uso de video y la metodología utilizada

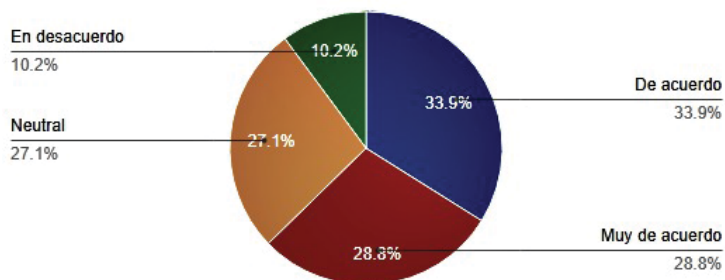
- Vinculado a la afirmación “*prefiero tener la clase tradicional (exposición) del docente en lugar de realizar trabajos activos y grupales en clases como los llevados a cabo*”; el 32,2% estuvo “en desacuerdo”, en tanto que el restante se concentró en las escalas restantes tal como se muestra en la Fig. 8.



- Fig. 8.** Gráfico en relación con la preferencia de método tradicional de enseñanza
- Referente a si “*el uso de videos me permitió aprender el material de estudio más eficazmente que hacer las lecturas en solitario*”, más del 80% correspondió a las opciones 3 y 4, tal como se ilustra en la **Fig. 9**. Lo cual significa que consideran que el material, les facilitó en gran medida sus elaboraciones.



- Fig. 9.** Resultado de encuesta en torno al uso de video.
- Al preguntar “*¿aprendí más cuando utilicé videos, lecturas cortas y actividades de aprendizaje activo en clase en comparación con el método tradicional (exposición del docente)?*”. Solo el 10,2 % estuvo en desacuerdo (**Fig. 10**).



- Fig. 10.** Nivel de aprendizaje que piensa (el alumno) que obtuvo con los tutoriales.
- Al indagar en torno al acompañamiento docente (**Fig. 11**), consultando puntualmente “*me sentía desconectado sin un profesor presente durante los videos o actividades virtuales*”. Un 40% contestó afirmativamente. Lo cual significa que el sistema tutorial debe ser revisado y optimizado.

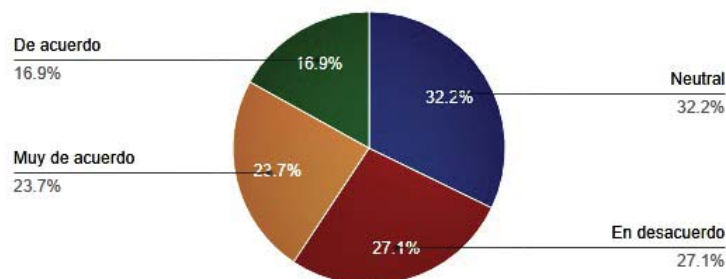


Fig. 11. Resultado del cuestionario con relación a la tutorización.

Asimismo, se solicitó indicar (en el formulario) el tutor que dirigió las tareas en cada caso, con la finalidad de obtener una retrospectiva de la tutorización e introducir mejoras en los mecanismos, en caso de ser necesario.

4. Conclusiones y trabajo futuros

De los resultados expuestos, es posible deducir que se aprovecharon las ventajas del Aula Invertida como estrategia de aprendizaje activa.

En este trabajo se expuso la redefinición de las actividades en la asignatura Introducción a las Tecnologías Aplicadas al Arte del primer año de la carrera de Licenciatura en Artes Combinadas, aplicando como estrategia de metodología el Aula Invertida y siguiendo el modelo TPACK. Así se orientó a la consecución de los objetivos de aprendizaje manifiesta en dos instancias: fuera del aula para acceder a información y contenidos didácticos y dentro del aula para la realización de las producciones.

La experiencia descrita se originó en el año 2012 como una actividad convencional presencial, por ello en 2016 se migró parte de la propuesta y se apoyó en el aula virtual de la asignatura para la tutorización y disposición de los materiales.

Además, se realizó el seguimiento del proceso de aprendizaje que, concluido se valoró a fin de obtener datos de retroalimentación. Los resultados arrojaron que una proporción significativa de alumnos prefieren la metodología tradicional, aunque reconocen que aprenden más, se propicia el trabajo autónomo y se contribuye a una adecuada gestión del tiempo.

Por otra parte, desde el punto de vista del docente, se destacan entre las ventajas: la facilidad para desarrollar el aprendizaje grupal personalizado, el empleo productivo del tiempo en el aula, la promoción del aprendizaje ubicuo, la posibilidad de mejorar el aprendizaje continuo y establecer un vínculo entre la educación formal y no formal mediante recursos disponibles en la Web. En definitiva, desempeña un papel de vital importancia en la capacidad de creación de los estudiantes, en la consecución de un aprendizaje más profundo y en la apropiación del contenido, motivo por el cual, el modelo Aula Invertida resulta de interés en la enseñanza de tecnología.

Se asume que la difusión de prácticas centradas en el aprendizaje activo en las aulas universitarias desde los primeros años es fundamental para lograr innovaciones educativas. Identificar aquellas que funcionan, pretende alentar a los diferentes actores de la Educación en la búsqueda de nuevos caminos para mejorar la enseñanza en los

diversos escenarios en que presenta la sociedad del conocimiento, y atendiendo a la diversidad del ritmo de estudio.

Particularmente, el diseño de la propuesta de Aula Invertida y su validación ha permitido corroborar la interpelación inicial que originó el trabajo: la inversión de la clase contribuye a la inversión en la clase, sustentado en que los docentes:

- Generaron nuevos formatos de interacción destinados a los estudiantes, definiendo el contenido y la tarea de seguimiento y evaluación correspondiente.
- Facilitaron experiencias de acceso autónomo al contenido, al trabajo crítico y constructivo en el tiempo presencial.
- Contribuyeron al desarrollo de prácticas de indagación, diferentes de las tradicionales.
- Siguieron los avances personales y grupales de los estudiantes en relación con la fidelidad de autoría en las producciones.

Por otra parte, la inversión en la clase se ilustra al dotar a los estudiantes de algunas competencias requeridas para la sociedad del siglo XXI, entre las que se destacan el trabajo con contenidos accesibles en distintos formatos, el trabajo en espacios virtuales, el trabajo cooperativo, entre otras.

En el año 2017, se realizó una implementación similar, en proceso de sistematización, así el análisis de ambas situaciones proporcionará datos con miras a mejorar el diseño propuesto. Además, como trabajos futuros se indagará en validar los aprendizajes de los estudiantes a través de cuestionarios intermedios o actividades en línea, con la finalidad de lograr un seguimiento más eficaz.

5. Bibliografía

1. Bishop, J. L., & Verleger, M. A. (2013). The Flipped classroom: A survey of the research. In ASEE National Conference Proceedings, Atlanta, GA (Vol. 30, No. 9, pp. 1-18).
2. Jiménez, L., Aguirre, I., & Pimentel, L. G. (2009). Educación artística, cultura y ciudadanía. Organização dos Estados Ibero-Americanos para a educação, a ciência ea cultura= organización de estados iberoamericanos para la educación, la ciencia y la cultura.
3. Halili, S. H., & Zainuddin, Z. (2015). Flipping the classroom: What we know and what we don't. *The Online Journal of Distance Education and e-Learning*, 3(1), 28-35.
4. Bergmann, J., & Sams, A. (2012). Flip your classroom: Reach every student in every class every day. International Society for Technology in Education.
5. Hamdan, N., McKnight, P., McKnight, K., & Arfstrom, K. M. (2013). The flipped learning model: A white paper based on the literature review titled a review of flipped learning. Flipped Learning Network/Pearson/George Mason University.
6. Acuña, B. P. (2015). Vectores de la pedagogía docente actual. ACCI (Asoc. Cultural y Científica Iberoameric.).
7. Fernández Delgado, A., Gutiérrez Rivas, P., & Tabasso, E. (2016). Humanizar la utilización de las TIC en educación. Dykinson.
8. Vigotsky, L. S. (1988). Interacción entre enseñanza y desarrollo. Selección de Lecturas de Psicología de las Edades I, 3.
9. Network, F. L. (2014). The four pillars of FLIP. Disponible en http://classes.mst.edu/edtech/TLT2014/BCH120/Abkemeier--FLIP_handout_FNL_Web.pdf

10. Domínguez, L. C., Vega, N. V., Espitia, E. L., Sanabria, Á. E., Corso, C., Serna, A. M., & Osorio, C. (2015). Impact of the flipped classroom strategy in the learning environment in surgery: A comparison with the lectures. *Biomédica*, 35(4), 513-521.
11. Rodríguez, W. P. (2016). Estudio de evidencias de aprendizaje significativo en un aula bajo el modelo flipped classroom. *EduTec. Revista Electrónica de Tecnología Educativa*, (55).
12. Missildine, K., Fountain, R., Summers, L., & Gosselin, K. (2013). Flipping the classroom to improve student performance and satisfaction. *Journal of Nursing Education*.
13. Simpson, K. (2014, abril 9). Flipped classroom - a model for the future? UNMC NEWSROOM. Recuperado de: <http://www.unmc.edu/news.cfm?match=12626>
14. Trends, E. (2014). Aprendizaje invertido. Monterey: Observatorio de innovación educativa del Tecnológico de Monterrey.
15. Burbules, N. C. (2014). Meanings of “ubiquitous learning”. *education policy analysis archives*, 22, 104.
16. Alshahrani, K., & Ally, M. (Eds.). (2016). *Transforming Education in the Gulf Region: Emerging Learning Technologies and Innovative Pedagogy for the 21st Century* (Vol. 171). Routledge.
17. Koehler, M., & Mishra, P. (2009). What is technological pedagogical content knowledge (TPACK)?. *Contemporary issues in technology and teacher education*, 9(1), 60-70.
18. Cabero Almenara, J. (2015). Reflexiones educativas sobre las tecnologías de la información y la comunicación (TIC). *Tecnología, Ciencia y Educación*, 1, 19-27.
19. Barroso Osuna, J. M., Cadena León, A. M., Castaño Garrido, C., Cukierman, U. R., Llorente Cejudo, M. D. C., Gallego Pérez, Ó., ... & Marín Díaz, V. (2014). La formación del profesorado en TIC: modelo TPACK.
20. Harris, J., Mishra, P., & Koehler, M. (2009). Teachers' technological pedagogical content knowledge and learning activity types: Curriculum-based technology integration reframed. *Journal of Research on Technology in Education*, 41(4), 393-416.
21. Fernández, M. G., Barrios, W. G., Godoy Guglielmone, M. V., & Gendin, G. (2013). Arte y TIC: Experiencias iniciales con herramientas de software en la formación de Licenciados en Artes Combinadas. In XVIII Congreso Argentino de Ciencias de la Computación.

Agradecimiento

Los autores integran el Proyecto acreditado por la Secretaría General de Ciencia y Técnica por Resol. 241/17 de la UNNE, PI 16F019: cuyo título es: “*TI en los Sistemas de información: Modelos, métodos y herramientas*”. Directora del mismo la Prof. Mariño, Sonia I. y Co-Directora, Prof. Godoy Guglielmone, Maria V. Se agradece el apoyo de la Secretaría.

ECMRE: Extended Concurrent Multi Robot Environment

J. Castro, L. De Giusti^{1,3}, G. Gorga¹, M. Sánchez¹, M. Naiouf¹, A. De Giusti^{1,2}

¹ Instituto de Investigación en Informática LIDI (III-LIDI) – Facultad de Informática –UNLP

² Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)

³ Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC)
camaju_25@hotmail.com, {ldgiusti,ggorga,msanchez,mnaiouf,
degiusti}@lidi.info.unlp.edu.ar

Abstract. El entorno ECMRE es una extensión de CMRE (Concurrent Multi Robot Environment) donde se incorporan características relacionadas a las arquitecturas paralelas actuales: heterogeneidad de procesadores, consumo energético, técnicas de cambios de velocidad del procesador en relación a la temperatura y/o consumo energético.

ECMRE permite incorporar los temas de concurrencia y paralelismo de forma sencilla y amena en cursos iniciales de las carreras de Informática mediante un entorno gráfico e interactivo.

Se ha realizado una prueba inicial en un curso con 42 alumnos para analizar la adaptación a este nuevo entorno y la utilización que pueden hacer del mismo.

Keywords: Concurrencia, Paralelismo, Procesadores heterogéneos, Algoritmos paralelos, Consumo energético.

1 Introducción

La Concurrencia ha sido un tema central en el desarrollo de la Informática y los mecanismos de expresión de procesos concurrentes que cooperan y compiten por recursos ha estado en el núcleo curricular de los estudios de Informática desde la década del 70, en particular a partir de los trabajos fundacionales de Hoare, Dijkstra y Hansen [1][2][3]. Por su parte el paralelismo, entendido como “concurrencia real” en la que múltiples procesadores pueden operar simultáneamente sobre varios threads o hilos de control en el mismo instante, resultó durante muchos años una posibilidad limitada por la tecnología de hardware disponible [4]. Las currículas informáticas clásicas [5][6][7] contenían conceptos de concurrencia en diferentes áreas (Lenguajes, Paradigmas, Sistemas Operativos) pero omitían casi totalmente el tratamiento del paralelismo, salvo al plantear los conceptos de sistemas distribuidos.

Los cambios tecnológicos han producido una evolución de los temas de mayor interés en informática, fundamentalmente por las nuevas aplicaciones que se desarrollan a partir de disponer de arquitecturas y redes de comunicación de mayor potencia y menor costo [8]. Por este motivo, las recomendaciones curriculares internacionales mencionan la necesidad de tratar los temas de concurrencia y paralelismo desde las etapas tempranas de la formación del alumno, dado que todas las arquitecturas y sistemas reales con los que trabajará son esencialmente paralelos

[9]. Sin embargo, la programación paralela (y los conceptos fundamentales de concurrencia) resulta más compleja para un alumno en las etapas iniciales de su formación, y es necesario contar con nuevas estrategias que permitan abordarlos.

Dados los estímulos que los alumnos reciben desde temprana edad, ya sea mediante juegos electrónicos, computadoras, celulares, tablets o cualquier otro dispositivo electrónico, la utilización de herramientas interactivas para la enseñanza de conceptos básicos a alumnos desde un curso CS1 [9][10][11] se ha vuelto una herramienta fundamental [12]. En este sentido, la posibilidad dar los primeros pasos en el mundo de la programación mediante un entorno gráfico e interactivo permite reducir la brecha que tradicionalmente existió entre la abstracción y la posibilidad de ver gráficamente la aplicación de los conceptos estudiados en un entorno que conceptualmente es similar a los utilizados en la vida cotidiana [8][13].

El entorno gráfico CMRE, donde se cuenta con un conjunto de robots que se mueven en una ciudad, ha permitido incorporar la enseñanza de los conceptos básicos de concurrencia y paralelismo en un curso inicial de Informática. En un artículo previo [14] se ha presentado la idea de incorporar al entorno características avanzadas de las arquitecturas paralelas actuales (tales como heterogeneidad, consumo de energía, temperatura generada). Continuando con ese trabajo, se ha realizado dicha extensión generando el entorno ECMRE (Extended Concurrent Multi Robot Environment), el cual se presenta en este trabajo.

El artículo está estructurado de la siguiente manera: en la Sección 2 se detallan las características avanzadas de las arquitecturas paralelas modernas que se han incluido en el entorno; en la Sección 3 se presenta el entorno CMRE actual. En la Sección 4 se explican las extensiones realizadas para generar el ECMRE; mientras que en la Sección 5 se muestra una prueba del nuevo entorno en un curso de primer año. En la Sección 6 se detallan las conclusiones.

2 Características avanzadas de Arquitecturas Paralelas

Las arquitecturas paralelas actuales poseen características avanzadas que son importantes de incluir dentro de los conocimientos iniciales de concurrencia y paralelismo. En particular la heterogeneidad de la arquitectura y el consumo energético.

2.1 Heterogeneidad de las Arquitecturas Paralelas

Históricamente se ha buscado incrementar el poder computacional de las máquinas. Sin embargo, se ha llegado a una situación en la que resulta difícil acelerar la velocidad de los procesadores incrementando la frecuencia de reloj de los mismos. Son dos los problemas que los arquitectos de hardware deben enfrentar: la generación de calor y el consumo de energía. La solución que presentan los diseñadores a estos problemas ha sido integrar, dos o más, núcleos computacionales dentro de un mismo chip, lo cual se conoce como procesador multicore o multinúcleo. Los procesadores multicore mejoran el rendimiento de una aplicación al distribuir el trabajo entre los núcleos disponibles [15][16].

Actualmente la investigación se está enfocando principalmente en Arquitecturas Multicores Heterogéneas (poseen *cores* con diferentes características en cuanto a rendimiento y consumo de energía, pudiendo utilizar o no distintos set de instrucciones), dado que tener *cores* de diferentes tipos permite optimizar el rendimiento, y al realizar una correcta distribución de las tareas entre los núcleos, se logra una mayor eficiencia en la relación rendimiento/energía.

En este tipo de arquitectura, la heterogeneidad se da en diferentes aspectos, los más importantes son: la potencia de cómputo (velocidad de cómputo) de los *cores*; el tiempo de acceso a memoria y la velocidad de comunicación entre *cores*. Estos tres aspectos definen el tiempo de ejecución de las instrucciones en cada *core*, por lo que una misma sentencia ejecutada en dos núcleos diferentes puede insumir tiempos distintos. Por otro lado, al existir un cierto grado de independencia entre las características que producen la heterogeneidad, no todas las instrucciones se ven influenciadas en la misma proporción. Es decir, una operación de punto flotante ejecutada en el *core* A, puede tardar la cuarta parte del tiempo que en el *core* B, mientras que en el caso de una operación de escritura puede tardar la mitad al ejecutarse.

2.2 Consumo Energético de las Aplicaciones Paralelas

El consumo energético es un punto clave en los procesadores actuales. En general la performance de un algoritmo paralelo no se mide sólo por el tiempo de ejecución del mismo, sino también por la energía consumida. Así aparecen índices que relacionan Flops/Watt o Flops/Joule según se relacione el cómputo con potencia instantánea o energía total [17][18].

Resulta importante en la formación de los alumnos de Informática hacer hincapié en las métricas de consumo como un indicador de calidad de los algoritmos. Asimismo, comprender los mecanismos automáticos que desarrollan los procesadores en función de la temperatura que alcanzan (que es una función directa de la energía consumida en un intervalo de tiempo) [17].

Existen técnicas de ajuste de rendimiento que utilizan los procesadores actuales tomando el consumo energético, la temperatura, y otros valores como indicadores para toma de decisiones. *Overclocking* y *underclocking* son dos de las técnicas más utilizadas para aumentar o disminuir la frecuencia del *clock* del procesador con el objetivo de incrementar la performance o normalizar los valores de consumo y temperatura cuando el procesador se encuentra sobrecargado.

3 El entorno CMRE actual

Las características principales del entorno CMRE pueden resumirse de la siguiente manera [15] [19]:

- Existen múltiples procesadores (robots) que realizan tareas y que pueden cooperar y/o competir. Los mismos representan los “*cores*” de una arquitectura multiprocesador real. Estos robots virtuales pueden tener un reloj propio y diferentes tiempos para la ejecución de sus tareas específicas.

- El modelo de ambiente (“ciudad”) en la que desarrollan sus tareas admite áreas privadas, parcialmente compartidas y totalmente compartidas. En un área privada sólo puede moverse un único robot, en un área parcialmente compartida se especifica el conjunto de robots que pueden moverse en ella y en un área totalmente compartida todos los robots definidos en el programa pueden moverse dentro de ella.
- Si se instancia a un sólo robot en un área que abarque toda la ciudad, se repite el esquema del Visual Da Vinci [20][21].
- Cuando dos o más robots están en un área compartida (parcial o totalmente), compiten por el acceso a las esquinas del recorrido y a los recursos que allí existan. Para esto deben sincronizar.
- Cuando dos o más robots (en un área común o no) desean intercambiar información (datos o control) deben hacerlo por mensajes explícitos.
- La sincronización se da por un mecanismo equivalente a un semáforo binario.
- La exclusión mutua puede generarse con la declaración de las áreas alcanzadas por cada robot. Acceder a otras áreas de la ciudad, así como salir de ellas no está permitido.
- Todo el modelo de ejecución es sincrónico y permite la existencia de un reloj virtual de ciclos, que a su vez permite asignar tiempos específicos a las operaciones, simulando la existencia de una arquitectura heterogénea.
- El entorno permite ejecutar el programa de manera tradicional, o paso a paso por instrucciones, dando al usuario un control detallado sobre la ejecución del programa, de manera de poder controlar situaciones típicas de concurrencia tales como conflictos (colisiones) o *deadlocks*.
- En la ejecución paso a paso, el efecto de las operaciones se puede reflejar en los robots físicos, comunicados vía *wifi*. Los robots físicos poseen un sistema operativo Linux que permite ejecutar un servidor http implementado en NodeJS [22]. De esta manera el entorno se comunica con los robots (cada robot físico se corresponderá con uno virtual en el ambiente). La comunicación entre ellos es punto a punto, y bidireccional, es decir, el entorno envía las instrucciones al robot físico y luego este último envía su respuesta al entorno indicando la finalización de la instrucción indicada.

4 Extensiones de CMRE

El entorno ECMRE (Extended Concurrent Multi Robot Environment) incorpora a su versión anterior (CMRE) los conceptos de arquitecturas multicore heterogéneas, el consumo energético, la temperatura de los procesadores y las técnicas de Overclocking y underclocking.

El alumno puede trabajar en ECMRE representando diferentes tipos de arquitecturas multicore y observar en forma gráfica e interactiva información relacionada con los tiempos de trabajo de cada robot y las variaciones de los valores de consumo y temperatura durante la ejecución de su algoritmo. A través del análisis de esta información, el alumno podrá modificar su algoritmo (por ejemplo,

balanceando la carga de trabajo) para obtener soluciones que resulten eficientes desde el punto de vista del consumo de energía y de la temperatura alcanzada.

4.1 Representación de la Heterogeneidad de las Arquitecturas Paralelas

En esta sección se describen las principales adaptaciones efectuadas al entorno CMRE para abordar los aspectos de velocidad y consumo energético incorporados al nuevo entorno ECMRE.

4.1.1 Rendimiento de los procesadores.

En ECMRE existe un área de *Detalles* (Fig.1(a)) donde se configuran características generales de la aplicación, y en particular de cada robot. En ella se incorpora en la tabla ROBOTS, una columna que permite definir la velocidad de cada uno. De esta forma se facilita el trabajo con los robots que podrán mostrar una variación en su rendimiento dado que ejecutarán las instrucciones del algoritmo a distinta velocidad. Existen 3 velocidades elegibles y la relación entre ellas es la siguiente:

- *Max* es la máxima velocidad disponible.
- *Med* es la mitad de la velocidad establecida como velocidad máxima.
- *Min* es la mitad de la velocidad establecida como velocidad media.

En ECMRE se define T como unidad de tiempo (equivale a 100 milisegundos) para medir el rendimiento de los robots/procesadores. Se ha decidido trabajar con un subconjunto de 10 instrucciones primitivas (*bloquearEsquina*, *depositarFlor*, *depositarPapel*, *derecha*, *enviarMensaje*, *liberarEsquina*, *mover*, *recibirMensaje*, *tomarFlor* y *tomarPapel*) clasificadas según su grado de complejidad y sobre ellas se asignó un tiempo de ejecución, consecuente con la velocidad del robot.

Además, en ECMRE existe un área de *Información de Ejecución* donde se visualiza la información actualizada de cada robot durante la ejecución del algoritmo (Fig.1(b)), uno de los datos corresponde al tiempo de ejecución en T unidades. Esta información es de gran utilidad para evaluar la performance del algoritmo.

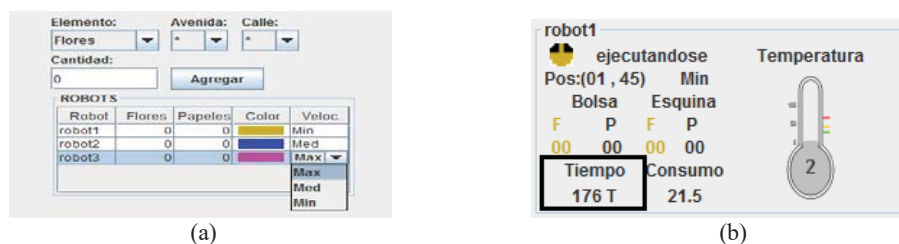


Fig. 1. Áreas de ECMRE: (a) *Detalles*, donde se configura cada robot, (b) *Información de Ejecución*, donde se observa el tiempo de ejecución del robot.

El manejo de velocidad por robot y la asignación de tiempos a las instrucciones primitivas en ECMRE, permite trabajar con robots/procesadores con distinto rendimiento, simulando una de las características de una arquitectura multicore heterogénea.

4.1.2 Consumo energético

El entorno ECMRE incorpora una nueva sección llamada Robot/Procesador donde se puede parametrizar el consumo energético (en Joules) para un conjunto de instrucciones primitivas (Fig.2(a)). A partir de esta información cada robot almacena su consumo y el mismo es actualizado durante la ejecución del algoritmo, lo cual se visualiza en el área de Información de Ejecución (Fig.2(b)).



Fig. 2. ECMRE (a) Consumo energético por instrucción en sección Robot/Procesador, (b) Información de Ejecución, donde se observa el consumo energético del robot

La velocidad del procesador es un factor que incide sobre el consumo (a mayor velocidad, mayor es el consumo). En ECMRE, el consumo generado por un robot r para ejecutar una instrucción i está dado por el consumo de i especificado previamente multiplicado por un coeficiente que representa a la velocidad de r en ese momento (0.25 para *Min*, 0.5 para *Med* y 1 para *Max*).

4.2 Representación de la Temperatura.

En la sección Robot/Procesador mencionada anteriormente también se puede parametrizar, para cada robot, la temperatura producida al ejecutar cada instrucción del conjunto de primitivas (Fig.3(a)), expresado en grados centígrados.

Se ha implementado la lógica necesaria para que cada robot registre su temperatura, que la misma sea actualizada y a la vez pueda visualizarse durante la ejecución del algoritmo. Para ello, se incorporó un termómetro en el área de Información de Ejecución del robot (Fig.3(b)) que refleja la temperatura con un valor numérico. El termómetro cambia de color (gris, verde, naranja y rojo) a medida que disminuye o aumenta la temperatura, representando con el color gris a la temperatura mínima y con rojo a la máxima.



Fig. 3. ECMRE (a) Temperatura por instrucción en sección Robot/Procesador, (b) Información de Ejecución, donde se observa la temperatura del robot

Esta característica también está influenciada por la velocidad del robot (a mayor velocidad, mayor es el incremento de la temperatura). En ECMRE, la temperatura de un robot r después de ejecutar una instrucción i está dado por: la temperatura previa

más la que generada la instrucción i especificada previamente, y este total se multiplica por un coeficiente que representa a la velocidad de r en ese momento (0.92 para *Min*, 0.95 para *Med* y 0.98 para *Max*).

4.3 Representación de las Técnicas de Overclocking y Underclocking.

Por otra parte, ECMRE contempla la posibilidad de que los robots apliquen *overclocking* y *underclocking* durante la ejecución del algoritmo. En la sección *Robot / Procesador* de ECMRE, se configuran los siguientes parámetros que activan dicha funcionalidad (Fig.4):

- *TCase Max*: es la temperatura máxima del procesador. Si se sobrepasa este valor:
 - Se intentará aplicar *underclock* para disminuir la temperatura del procesador.
 - Si no se admite *underclock* o la velocidad es la mínima, el robot se detiene hasta normalizar su temperatura.
- *Utilizar Overclock*: habilita la utilización de *overclocking* en el robot.
- *Utilizar Underclock*: habilita la utilización de *underclocking* en el robot.
- *Intervalo de Verificación (IV)*: indica cada cuanto tiempo verificar el incremento del consumo energético y aplicar *overclocking/underclocking* según corresponda (en unidades de tiempo T).
- *Variación de Consumo Esperada (VCE)*: indica la variación de consumo energético (en joules) esperada para el tiempo definido en *IV*. Valor utilizado para aplicar *overclocking/underclocking*.

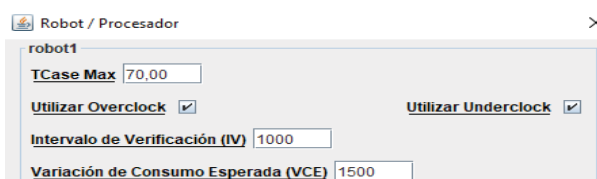


Fig. 4. ECMRE - Parámetros de ajuste de rendimiento en sección Robot / Procesador

El algoritmo de ajuste de rendimiento en cada robot trabaja de la siguiente manera: Por cada instrucción ejecutada evalúa:

- Si la temperatura del robot es mayor a la temperatura máxima especificada (*TCase Max*):
 - Si el robot soporta *underclocking* y la velocidad actual no es la mínima, se aplica *underclock* para disminuir la temperatura. Notar que si la velocidad es *Min* no es posible seguir disminuyendo su valor.
 - En caso contrario, el robot se detiene hasta enfriar su temperatura a 25 grados y luego continúa su procesamiento.
- Si el incremento de consumo registrado durante el intervalo de verificación (*IV*), supera la variación de consumo esperada (*VCE*):
 - Si el robot soporta *underclocking* y la velocidad actual no es la mínima, se aplica *underclock* para disminuir el consumo energético y la temperatura. Notar que si la velocidad es *Min* no es posible seguir disminuyendo su valor.
- Si el incremento de consumo registrado durante el intervalo de verificación (*IV*), es menor a la variación de consumo esperada (*VCE*):

- Si el robot soporta overlocking y la velocidad actual no es la máxima, se aplica overlock para incrementar la performance. Notar que si la velocidad es Max no es posible seguir aumentando la velocidad.

Las operaciones de *overclock* y *underclock* no se aplican mientras el tiempo de variación de consumo sea inferior al parámetro *IV*. Cada vez que se aplica *overclock*, *underclock* o el robot se detiene para enfriarse, los valores de variación de consumo y tiempo transcurrido son reiniciados a 0.

4.4 Logs del Procesador.

Las nuevas funcionalidades incorporadas a ECMRE pueden requerir realizar un análisis post ejecución del algoritmo para evaluar si el resultado obtenido cumple con lo esperado. En el caso que, el tiempo de ejecución o el consumo energético sea mayor al esperado, es posible que se proponga modificar la solución implementada o bien reasignar las tareas entre los procesadores involucrados para obtener un mejor resultado. Con el objetivo de facilitar este análisis se ha incorporado a ECMRE una nueva sección llamada *Log del Procesador* que muestra de forma amigable y resumida (mediante la utilización de tablas y gráficos) la ejecución de un algoritmo.

ECMRE registra información de los procesadores durante la ejecución del algoritmo. Dentro de la sección *Robot/Procesador* se puede parametrizar la *frecuencia de log* durante la ejecución de cada robot (cantidad de instrucciones).

Los eventos que se registran pueden ser: procesador detenido por sobrecarga, overlock y underclock. Por cada entrada de log registrada se almacena la siguiente información del robot: evento ocurrido (LOG, DETENIDO, OVERCLOCK y UNDERCLOCK), velocidad antes y después de la ocurrencia del evento, temperatura y consumo energético.

La sección *Log del Procesador* cuenta con 5 subsecciones diferentes denominadas: Temperatura, Gráfico Temperatura, Consumo, Gráfico Consumo y Gráfico Consumo x Instrucción y en cada una de ellas el usuario puede seleccionar el robot objeto de análisis.

5 Sesión de pruebas con alumnos

El entorno ECMRE fue presentado en la materia Taller de Programación de la Facultad de Informática de la UNLP. La materia corresponde al 1er año de las carreras de la Facultad y se compone de 3 módulos, el tercero corresponde a la introducción de los conceptos básicos de la Programación Concurrente, donde actualmente utilizan el entorno CMRE.

La sesión se inició con un breve repaso de los conceptos de concurrencia y paralelismo abordados en el entorno como son las arquitecturas multicore (homogéneas y heterogéneas), consumo energético, temperatura y técnicas de ajuste de rendimiento en procesadores, y balance de carga. Para cada ítem se revisó el concepto y su importancia en relación a las arquitecturas de computadoras actuales y se presentaron los nuevos elementos (temperatura, consumo y velocidad) que han sido incorporados al entorno CMRE dando origen al entorno ECMRE.

Una vez finalizada la revisión teórica, se planteó una actividad práctica que fue desarrollada con la supervisión del docente y el conjunto de los alumnos. Los alumnos no interactuaron directamente con el entorno debido a que aún estaba en la etapa de desarrollo. La actividad práctica consistió en la resolución de un problema modelo, utilizando el entorno ECMRE y donde se mostró la variación en tiempo de ejecución, consumo energético y temperatura de acuerdo a las diferentes configuraciones de los robots, utilizando tablas, gráficos y logs explicados en la sección anterior.

Al finalizar la actividad práctica, se entregó a cada alumno una breve encuesta anónima, con el objetivo de tener un primer feedback de los alumnos, quienes serán usuarios finales de la herramienta. La misma se compone de 5 preguntas cuyas respuestas cubren una escala de Likert que varía de 1 (en completo desacuerdo) a 5 (completamente de acuerdo). La encuesta se realizó a los 42 alumnos que asistieron a la clase de Taller el día de la experiencia y se obtuvieron los resultados de la Tabla 1.

Tabla 1. Resultados encuesta alumnos de Taller de Programación.

PREGUNTA	RESULTADOS OBTENIDOS				
	Comp. de acuerdo	De acuerdo	Ni de acuerdo, ni en desacuerdo	En desacuerdo	En completo desacuerdo
La aplicación ECMRE colabora en el aprendizaje de los conceptos presentados.	36%	57%	7%	0%	0%
Es útil contar con una herramienta de aplicación práctica que permita mediante ejemplos concretos visualizar los contenidos teóricos aprendidos en clase.	55%	38%	7%	0%	0%
Los contenidos en ECMRE se encuentran organizados y su utilización es intuitiva.	24%	52%	24%	0%	0%
La iconografía y gráficos utilizados en la aplicación tienen el tamaño adecuado y coinciden con la función asociada.	48%	38%	14%	0%	0%
Es beneficioso contar con esta aplicación como complemento a las clases teóricas.	52%	36%	12%	0%	0%

6 Conclusiones

Los temas de heterogeneidad, consumo energético y temperatura en arquitecturas paralelas son de gran importancia y se ha presentado una extensión de CMRE para incluirlos (ECMRE).

ECMRE aparece como una herramienta muy útil para la introducción de estos conceptos en los cursos iniciales de las carreras de Informática. Para lograr esto, se han realizado dos etapas, por un lado, la modificación del entorno CMRE para permitir manejar estas características en los robots y, por otro, se incorporaron herramientas gráficas para poder visualizar y posteriormente analizar esta información por parte de los alumnos de forma sencilla y amena.

De esta manera, el nivel de complejidad de los posibles escenarios aumenta, planteando un desafío mucho más ambicioso, que responde a la realidad tecnológica de los procesadores actuales.

Referencias

1. Hoare C. "Communicating Sequential Processes". Prentice Hall, 1985.
2. Dijkstra E. W. "Finding the Correctness Proof of a Concurrent Program". In Program Construction, International Summer School, Friedrich L. Bauer and Manfred Broy (Eds.). Springer-Verlag, 24-34, 1978.
3. Hansen P. B. "The Architecture of Concurrent Processes". Prentice Hall, 1977.
4. Dasgupta S. "Computer Architecture. A Modern Synthesis. Volume 2: Advanced Topics". John Wiley & Sons, 1989.
5. ACM Curriculum Committee on Computer Science. "Curriculum '68: Recommendations for the undergraduate program in computer science". Communications of the ACM, 11(3):151-197. 1968.
6. ACM Curriculum Committee on Computer Science. "Curriculum '78: Recommendations for the undergraduate program in computer science". Communications of the ACM, 22(3):147-166. 1979.
7. ACM Two-Year College Education Committee. "Guidelines for associate-degree and certificate programs to support computing in a networked environment". New York: The Association for Computing Machinery. 1999.
8. Hoonlor A., Szymanski B. K., Zaki M. J., Thompson J. "An Evolution of Computer Science Research". Communications of the ACM. 2013.
9. ACM/IEEE-CS Joint Task Force on Computing Curricula. "Computer Science Curricula 2013". Report from the Task Force. 2013.
10. ACM/IEEE-CS Joint Task Force on Computing Curricula. "Computer Engineering 2004: Curriculum Guidelines for Undergraduate Degree Programs in Computer Engineering". Report in the Computing Curricula Series. 2004.
11. ACM/IEEE-CS Joint Interim Review Task Force. "Computer Science Curriculum 2008: An Interim Revision of CS 2001". Report from the Interim Review Task Force. 2008.
12. De Giusti, L., Leibovich, F., Sanchez, M., Chichizola, F., Naiouf, M., De Giusti, A. "Desafíos y herramientas para la enseñanza temprana de Concurrencia y Paralelismo". Congreso Argentino de Ciencias de la Computación (CACIC), 2014.
13. AMD. "Evolución de la tecnología de múltiple núcleo". <http://multicore.amd.com/es-ES/AMD-Multi-Core/resources/Technology-Evolution>. 2009.
14. De Giusti, L., Leibovich, F., Chichizola, F., Naiouf, M., De Giusti, A. "Incorporando conceptos en la enseñanza de Concurrencia y Paralelismo utilizando el entorno CMRE". Congreso Argentino de Ciencias de la Computación (CACIC), 2015.
15. Gepner P., Kowalik M.F. "Multi-Core Processors: New Way to Achieve High System Performance". In: Proceeding of International Symposium on Parallel Computing in Electrical Engineering 2006 (PAR ELEC 2006). Pags. 9-13. 2006.
16. Mc Cool M, "Programming models for scalable multicore programming", 2007, <http://www.hpcwire.com/features/17902939.html>.
17. Ballardini J., Rucci E., De Giusti A., Naiouf M., Suppi R., Rexachs D., Luque E. "Power Characterisation of Shared-Memory HPC Systems". Computer Science & Technology Series—XVIII Argentine Congress of Computer Science Selected Papers. Págs. 53-65. 2013.
18. Brown D. J., "Toward Energy-Efficient Computing", Magazine Communications of the ACM Volume 53 Issue 3, March 2010
19. De Giusti, A., De Giusti L., Leibovich, F., Sanchez, M., Rodriguez Eguren, S. "Entorno interactivo multirrobot para el aprendizaje de conceptos de Concurrencia y Paralelismo". Congreso Tecnología en Educación, Educación en Tecnología. 2014.
20. Champredonde R., De Giusti A. "Herramienta visual para la enseñanza de programación". Congreso Argentino de Ciencias de la Computación (CACIC), 1996.
21. Champredonde R., De Giusti A. "Design and Implementation of the Visual DaVinci Language". Congreso Argentino de Ciencias de la Computación (CACIC), 1997.
22. <https://nodejs.org/api/http.html>

Educational Experiences in Informatics: Innovations in Internet Searches in Academic Environments

Pablo Argañarás^{1,2}, Paola Britos^{1,2}, Martín Vilugrón^{1,2}

¹ Sede Andina, Universidad Nacional de Río Negro, Mitre 630 1° B,
8400 S. C. de Bariloche, Río Negro, Argentina

² Grupo de Investigación de Explotación de Información, Laboratorio de Informática Aplicada, Sede Atlántica, Universidad Nacional de Río Negro, 8500 Viedma, Río Negro, Argentina
{parganaras, pbritos, mrvilugron}@unrn.edu.ar

Abstract. This paper presents innovative experiences in the way of teaching to search information of academic style in Internet to first year students, of the UNRN (Andina headquarters) where Informatics is a transversal topic to their careers. Since 2013 in the Workshop on Informatics a group of teachers inquiries about innovate the way how students make Internet information searches on the academic section, trying that they acquire the ability and the criteria to potentiate their searches, to value their results and to give meaning to the iterative cycle of conformation of correct search keywords, trying this to be useful in the development of their performance as students and later as professionals. We will discuss the methodological recommendations that are offered, the types of tasks and guidelines with which it is tried to innovate the search of information in Internet in academic environments, and the obtained results since 2016.

Keywords: search information of academic style in Internet, higher education, learning, collaborative environment, b-learning.

1 Introduction

Internet information searches in traditional academic settings such as the university are exercises in which students have to collect information about a particular topic or a particular subject without an immediate and conscious construction of knowledge in the act.

In the Informatics and ICT Workshop, which has as its main characteristic its great practical load, it is worked so that students acquire skills and abilities in the management and use of an isolated or networked computer system for educational purposes.

As a result of the socio-cultural reality of the region, our students are mostly the first generation of university students of their families to whom we try to insert as active participants in the information and knowledge society, respecting the project of each of them, emerging from different contexts and responding to different perceptions and expectations [1], and taking in mind the type of group to which they

belong (older adults, adults, young people and students with completed higher education) [2]. Then, it is an interesting challenge for teachers to inquire more about how much of the knowledge provided in the courses is assimilated by our students in college.

If we wish to improve the quality of our students' research results in the academic field, we must take into account that the evolution of information technologies poses new challenges to education, because in the future the acquisition and organization of information it will become in the vital activity that dominate a good part of the population [3]. Therefore continue teaching to look for information as we were doing before the development of web 2.0 and web 3.0 would become a bad practice.

Also, it is interesting to use some innovative strategy in didactic for the teaching of subjects oriented to the TICs, trying to "humanize" the teaching of the contents of Informatics, which would allow approach the students in the understanding of the phenomena, in a more simple, complete and realistic form [4].

The Internet represents a significant and indispensable tool for the agile search for academic and scientific information that teachers and students must take advantage of, which imposes a rethinking, reflection and questioning of contexts for learning in the field of higher education [5].

There are studies that analyze the relationship between cognitive style and academic achievement in which, for example, issues that bring ICTs, field independence and learning achievement are addressed. And in research involving university students who searched information in databases with a hypermedia structure, the results showed that the independent students of the fields used the tools autonomously, while the dependent students of the field needed social assistance [6]. That is, if there is a background in which ICTs are used in relation to the search for information and academic achievement, from the classroom we can make our contribution so that the learning of internet searches is productive and allows our students to enrich their learning process so that they value the possibility of learning to learn.

1.1 Motivation

The importance of this research is based on the search for quantifications and recommendations on the subject in the UNRN. Even though there are similar studies with specific courses, this research is justified in the university teaching by the interdisciplinarity of the students of the same course, where Informatics and ICTs are a transversal subject, and the interest of the chair to implement experiences where incorporate innovations in the way of looking for information of academic style realized by the contemporary university students, whom we will call university students 3.0.

The strengths of this experience may be the commitment of the coordinator and of the teachers of the Informatics and ICTs workshop, as well as the good reception of the study among the students of the Workshop. And the weakness detected lies in the quality of the internet connection, with a deficient service in both the wireless connection and the wired connection; this despite the good will and predisposition of the technical team that maintains the resources of the Informatics Laboratory, which is beyond the conditions of study that teachers can propose and control.

1.2 Context

This research is framed within the set of tasks that the members of the Information Exploitation Research Group carry out as research teachers, who try to improve the educational level in Information Technology and ICTs. The presented analysis involves students of the Schools of Humanities and Social Studies; Production Technology and Environment; and of Economy, Administration and Tourism; and to the teachers of the Workshop on Informatics and ICTs of the Andina Headquarters of Bariloche and El Bolsón.

The axes of the present investigation are:

- Internet searches in academic environments
- The use of collaborative work environments in university courses
- Assessment methods in computer science and university ICT workshops when students use Web-based work environments.

Among the software tools used in the workshop on Informatics and ICT are:

- A **Virtual Platform in Moodle**, where teachers organize and store all course material including practical activities, homeworks delivery time, evaluation dates, and qualifications. It is a useful tool for students that cannot attend the class for some special reason (health, labor, or climatic factors) and for one or very few days.
- **Google's collaborative environment** with its applications, especially the Google Docs with their word processor, spreadsheet, and digital presentations generator.
- **Local free software applications** installed in the computer lab of UNRN over the Linux Ubuntu operating system, like LibreOffice and Firefox.

1.3 Objectives

Some of the questions that guide this research are: *How do the students who enter the university seek information? How do ICT help university students in their academic production? Are varied the search engines that students use? How much does help to the process of learning IT and ICT the search for information in academic environments? Do help internet searches to university students for their continuous training?*

The objectives of the present investigation are:

- Minimize student dropout from the Informatics and ICT Workshop
- That students be learn and use different Internet search tools and familiarize themselves with web-based work tools to improve academic production
- That the assessment of knowledge acquired in the workshop be reflect the ability of students to use learned techniques.

We will focus on the teaching of Informatics and ICTs in the university with students of different careers, and in its relation with the school performance applying the b-learning modality for the accomplishment of group works for integrate knowledge, and mainly in the search of information and including contextualized

searches for students of each career in an exam of Informatics and ICTs, to which the notions of Copyright, Copyleft and Creative Commons are associated, and lastly, the answers are analyzed to reveal a consummate learning associated with knowledge theoretical practical of Informatics and ICTs, or not.

2 Description of Experience

This experience consists of:

- Teach to formulate a search strategy, that includes:
 - Selecting a correct search source
 - Selecting search keys relevant to the topic to be searched
 - Defining a strategy based on operators and/or filters by gathering at least three search keys
 - Evaluate the results of the search
 - Redesign the strategy and return to the selection of the search keys
 - Finally store the results of the searches and information about how to quote those results.
- Teaching to make internet searches through the different types of search engines that students can find and use (i.e. search engines for common search, metasearch, visual search, specific search, semantic search, and resource directories), going from text keys to image keys.
- Propose some examples in class where the students should to solve short problems of internet searches combined with writing exercises using Google's collaborative word processor and applying the technics learned with the known search engines that best fit each problem.
- To design an assessment that contains internet searches using keywords related to the career of each student. This means that the teacher should think about and should stay reported on topics that appear in scientific articles, technical manuals, technical reports, or other academic or professional recognition publications, preferably with reference. Such keywords should be known or understood by first-year college students.
- The assessment must combine some questions of traditional exams where the answer depends of the study and reasoning of the student, with some innovative questions with answers that are not deterministic and can vary from one student to another.

3 Preliminary Results

Since the incorporation of the use of collaborative environments, of a virtual platform, and of the implementation of this new form of internet search in the Workshop of Informatics and ICTs, in the year 2016 were generated around 6 integrator works by course in the headquarters Andina Bariloche, and the dropout rate of 4 students per year was maintained on the highest dropout course.

In two commissions in which this new way of searching for information was applied, one student couldn't promoted the workshop with a grade equal to or greater than seven (7) and none of them disapproved the course.

The question about traditional searches in the evaluation, ask to mark True options in 30% of the evaluation, and to write definitions and examples in a 10%. The search questions that we will call innovators are those that by using two search engines available, and three contextualized keywords to each of the careers of the students of the course (given by the teacher), complete two rows of a table with the addresses of sites or pages that contain material with all of those keywords and which are licensed by Copyright, Copyleft and Creative Commons respectively.

We will comment on the experience carried out in two commissions of School of Production in headquarter Andina Bariloche, with 23 students in total, each one of which was separated into two groups to carry out the evaluation (due to the computers availability and times preference) for two hours.

To this first question considered innovative: *Using a pair of available search engines complete the table with the first two "useful" sites marked with Copyright license, two with Copyleft and two with Creative Commons*, they were associated three assigned search keywords according the careers of the students, as it is shown in Table 1.

Table 1. Internet search keywords assigned by career.

Career	Keywords		
Nursery Technician (TEVI)	Nurseries	Patagonia	bulbs
Chemistry Teacher (PQUI)	chemistry	teaching	molecules
Physics Teacher (PFIS)	Physics	optics	lenses

For the searches of TEVI the correct answers that are expected are two with Copyright and with Creative Commons, and none with Copyleft; and for PQUI and PFIS searches the correct expected answers are two with each type of license.

The number of students who answered without error the results with the Copyright license was almost the double of the results with the Creative Commons license and the results with the Copyleft license (Fig. 1). A similar relation it is verified between the answers with one error for the Copyright license and the Creative Commons license versus the Copyleft license (Fig. 2). And finally, those students who answered incorrectly the two questions of Copyleft were 14 and of Creative Commons were 6. None of the students have had two wrong answers for results with Copyright license.

The second question considered innovative it is: *Provide a list of at least two Application Software that you believe could be useful to you in your professional activity after your graduation, with type license CC-BY-NC and indicate it the career you are studying and ¿what for do you believe you would use them?*

The students who answered correctly to this question were 4, answered fifty percent correct 10 students, less than fifty percent 5 and totally incorrect 4. The category of less than fifty percent corresponds to the answers that show a correct explanation about the type of application software selected and the use the student think for it but couldn't identify an application by name (i.e. word processor or

spreadsheet). The distribution of the answers of TEVI, PQUI and PFIS for this second innovative question it is shown in the Fig. 3.

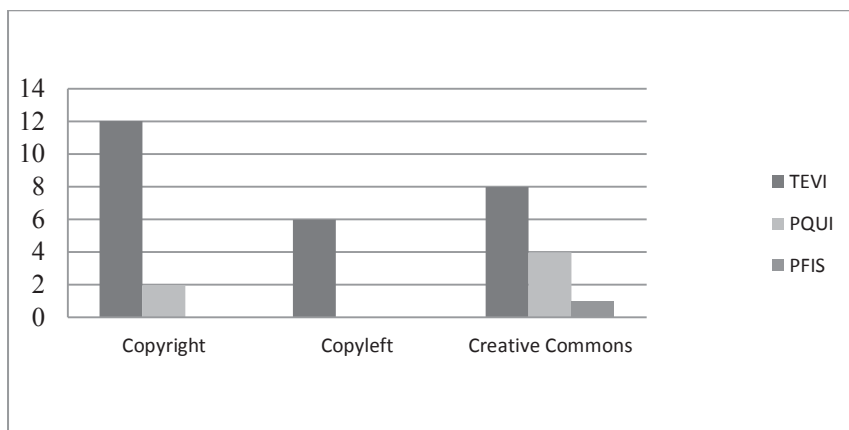


Fig. 1. Answers to the First Innovative Question without errors

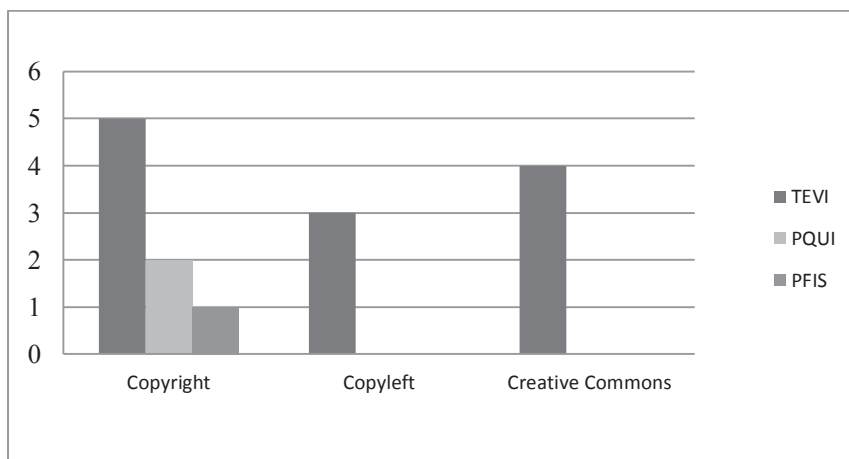


Fig. 2. Answers to the First Innovative Question with one error

The third question considered innovative it is: *Choose a search engine (record your choice in this evaluation) and use it for information about "the benefits of informatics knowledge in college". Perform three (3) iterations and list the first 5 results (your links and names) you obtained on each iteration. Explain what changes you make between one iteration and the next. Select the best result obtained.*

This last question was added to a subgroup of the students since the first semester of 2017 looking for new ways to investigate how complete it is the knowledge of our students about the searching process when they must to perform some internet search of academic type refining the searching key, and this will be measured and analyzed at the end of this year or early next year.

It is important to keep in mind that to these innovative questions there is associated another traditional question in the same assessment (*Define Copyright and Copyleft on the web, and explain their differences*), so that the student can demonstrate his knowledge about the topics Copyright, Copyleft and Creative Commons. Then with the internet search the student would finish demonstrating the understanding of both topics and could gain knowledge about available sites and resources with information related to his career.

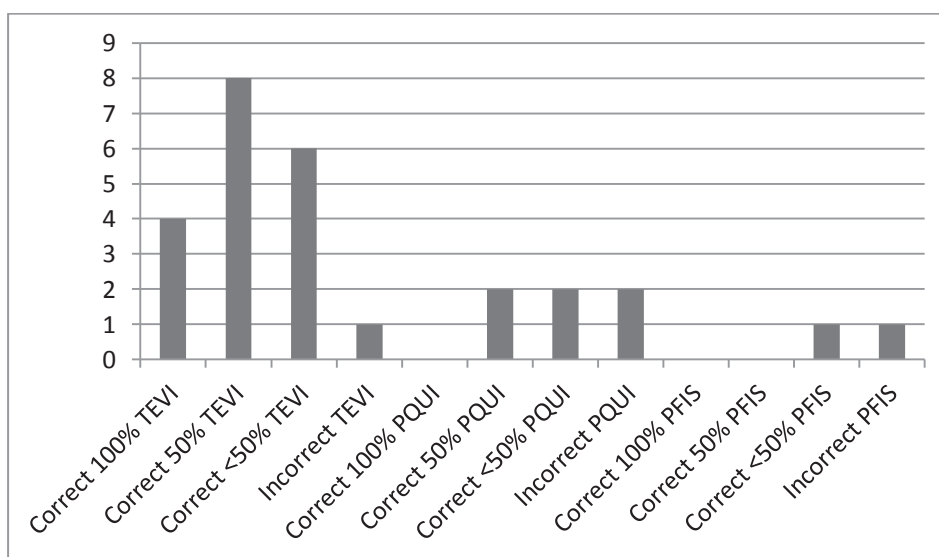


Fig. 3. Answers to the Second Innovative Question

4 Conclusions

It is observed that the searches performed by students with ordinary search engines for academic topics do not generate too many disadvantages when the result is licensed by Copyright; without going too deeply into the quality of the content. On the contrary, finding results with a Creative Commons license it present them with some degree of difficulty, and they find it difficult to find results with Copyleft license.

The result of the first innovative question may be caused because the new generation of university students is very impatient and they do not read consciously the information of the sites or pages that choose as correct answers. In many internet searches they have found keywords as part of advertisements or links to another pages or sites, which are not part of the main text, and probably that was the reason of the mistake in their selection. Another reason for this result may be the few experience of the students in the field of knowledge to which belong the questions.

The result of the second innovative question exposes many doubts of the students related with application software concept by one hand and with the presumption of the license of the applications that they know and usually use by the other. Some of

the proposed applications with name are shown in Fig. 4 and the generic applications type most proposed are shown in Fig. 5

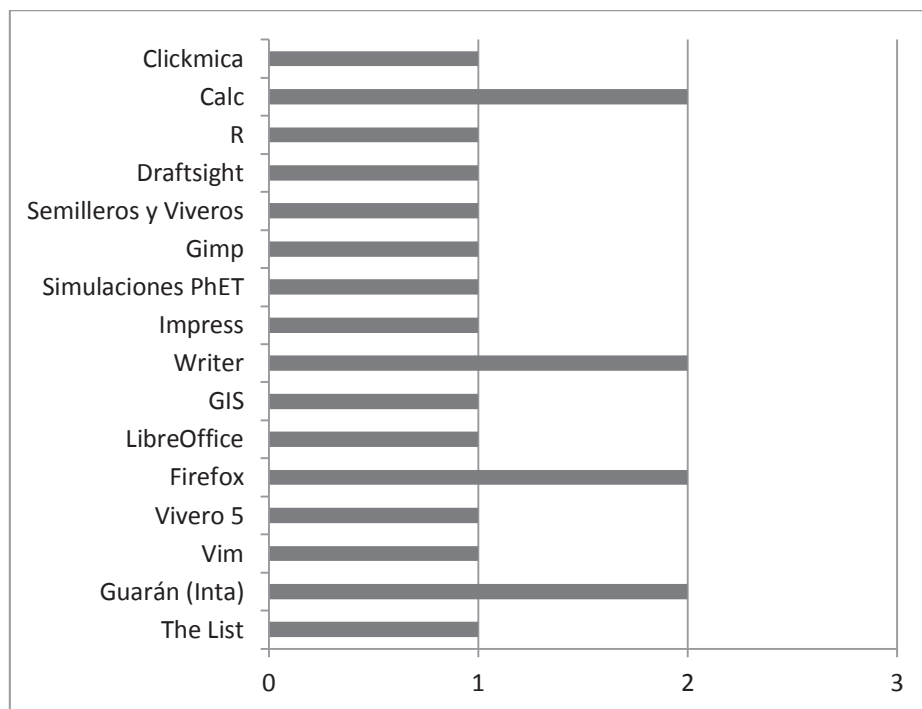


Fig. 4. Applications software with name

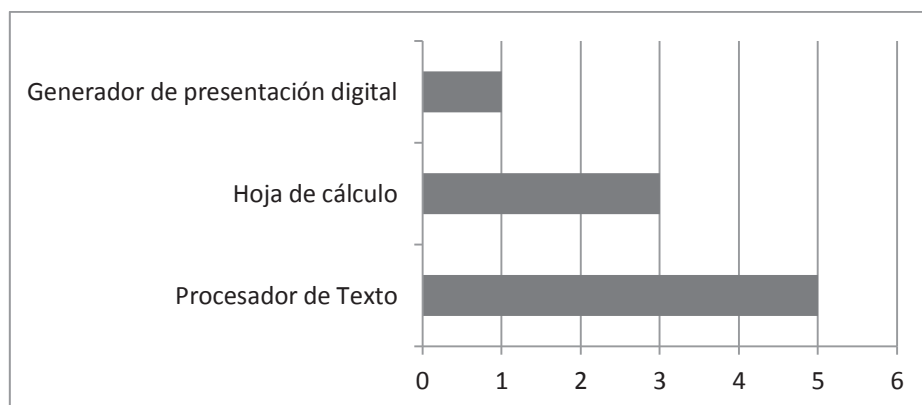


Fig. 5. Generic application type

The third innovative question tries to impulse to the students to make a certain number of repetitions of the cycle of refinement of the search key using connectors, synonyms and all the necessary elements to improve the results of the internet

searching process. The study and analysis of the responses to this question will be part of future works in this direction.

With this we can conclude that although the Workshop promote the use of free software and work with collaborative and cooperative resources, the materials of interest that our students can easily find on the web are not always marked with licenses that allow their reusability and redistribution, such as those licensed with Creative Commons and Copyleft.

5 References

1. Fandiño Parra, Y.J.: La educación universitaria en el siglo XXI: de la sociedad de la información a la sociedad del conocimiento. *Revista Iberoamericana de Educación*, vol. 55, no 3, ISSN 1681-5653 (2011)
2. Argañaraz, P.E., Britos, P.V., Verdun, N.: Experiencias educativas en el uso de TIC en la educación superior en la sede andina de la Universidad Nacional de Río Negro. XVI Workshop de Investigadores en Ciencias de la Computación, pp. 1069--1073, Ushuaia, Tierra del Fuego, Argentina (2014)
3. Salinas, J.: Nuevos ambientes de aprendizaje para una sociedad de la información. *Revista pensamiento educativo*, 20, 81--104 (1997)
4. Mon, Alicia, Horacio Del Giorgio, and Bettina Donadello. "Estrategias didácticas Innovadoras en la enseñanza de TICs para ingeniería en informática." - XXI Congreso Argentino de Ciencias de la Computación - pp. 1083--1091 (Junín, 2015). 2015.
5. Pérez, M.A., Pérez, R., León, S., Maninat, M., Triana, J.: Moscaweb contenidos-instrumento para la evaluación de contenidos académicos provenientes de internet. *Revista de la Facultad de Ingeniería U.C.V.*, vol. 25, no 3, pp. 87 -- 98, Caracas (2016)
6. López-Vargas, O., Hederich-Martínez, C., Camargo-Urbe, A.: Estilo cognitivo y logro académico. *Revista Educación y Educadores*, vol 14, pp. 67--82, Colombia (2011)

El Pensamiento Computacional: experiencia de su aplicación en el aprendizaje de la resolución de problemas

María V. Rosas, Mariela E. Zúñiga, Jacqueline M. Fernández, Roberto A. Guerrero

Laboratorio de Computación Gráfica, Departamento de Informática, FCFMyN
Universidad Nacional de San Luis
Ejército de los Andes 950 (San Luis, Argentina)
{mvrosas, mezuniga, jmfer, rag}@unsl.edu.ar

Resumen. El Pensamiento Computacional (PC) es el núcleo de las disciplinas actuales en Ciencia, Tecnología, Ingenierías y Matemáticas, representando una alternativa que mejore índices de deserción y desgranamiento en carreras universitarias. Este trabajo relata una experiencia piloto realizada en el ámbito de la Facultad de Ciencias Físico Matemáticas y Naturales, destinada a alumnos inscriptos al ingreso 2017. El curso tuvo como propósito desarrollar conceptos básicos del PC, centrándose en: Descomposición, Abstracción, Reconocimiento de patrones y Algoritmo. Se registró además el impacto que éste provocó mediante una encuesta al comienzo y otra al finalizar. Del análisis de los datos se promueve un espacio de reflexión que favorezca transformaciones que garanticen el ingreso y la permanencia en la universidad, comenzando por la formalización como curso de ingreso obligatorio para todos los alumnos que en el primer año de su carrera tengan una materia relacionada con la resolución de problemas computacionales y/o la programación.

Palabras clave: Pensamiento Computacional, resolución de problemas, curso de ingreso. Ingreso y permanencia de los alumnos al nivel superior.

1 Introducción

En todo momento el contexto económico, político, tecnológico y cultural presenta nuevos desafíos al sistema educativo provocando modificaciones y reestructuraciones en sus propuestas pedagógicas. Estos cambios se hacen más evidentes en los niveles superiores de formación. La tendencia actual determina que los nuevos profesionales estén capacitados para afrontar una gran variedad de situaciones, incluso estar preparados para afrontar problemas que tal vez aún no existen como tal. Por lo tanto, cabe a las instituciones educativas adaptarse a los requerimientos del mundo actual, formando a las nuevas generaciones en habilidades de pensamiento como complemento de los contenidos específicos que fueron el eje de la enseñanza hasta el momento [1].

La universidad como parte del sistema educativo no puede estar ajena a los desafíos actuales, debiendo capacitar a sus futuros egresados en una formación más generalista, centrada en prácticas de resolución de problemas e incentivando la

creatividad para desempeñarse satisfactoriamente en su campo profesional. El mundo laboral va sufriendo transformaciones, surgiendo nuevas profesiones desconocidas o inimaginables hasta hace unos años atrás. La grieta entre lo que se espera de los nuevos profesionales y cómo ellos egresan de la universidad puede ir disminuyendo a partir de centrar los esfuerzos en formar a los alumnos en rutinas y prácticas de pensamiento y no en conocimientos específicos, y que esto sea una constante desde el comienzo de su formación. En este sentido, podría considerarse que llevar a cabo estas acciones desde el primer año de cualquier carrera universitaria contribuiría a la adaptación y permanencia del alumnado favoreciendo su desempeño académico durante todo el proceso de formación de grado [2] [3].

En particular, en carreras vinculadas a las Ciencias e Ingeniería, diferentes autores han planteado la problemática recurrente del desgranamiento en los primeros años y la falta de motivación, presentando diferentes propuestas para hacer frente a estas realidades [4][5]. Algunas de estas propuestas implementaron períodos de aprestamiento utilizando herramientas lúdicas, cursos de ingreso con herramientas específicas y prácticas de programación, trabajos interinstitucionales de articulación con el nivel medio, entre otros. Estas iniciativas lograron resultados prometedores mediante prácticas acotadas a resoluciones concretas y con objetivos orientados en su mayoría al aprendizaje de la programación [6] [7] [4] [8].

Por otro, el desarrollo del Pensamiento Computacional (PC) ha evidenciado un fuerte avance teórico en los últimos años. Varios autores manifiestan la importancia de promover su desarrollo desde edades muy tempranas, considerando que esto favorece significativamente las habilidades de los estudiantes para enfrentar y resolver los problemas de diversa índole: académicos, personales, sociales, entre otros. [9] [10] [4] [11] [12]. Según Jeannette Wing, el PC se define como el conjunto de los procesos de pensamiento implicados en la definición de problemas y en la representación de sus soluciones, de manera que dichas soluciones pueden ser efectivamente ejecutadas por un agente de procesamiento de información (humano, computadora o combinación de ambos) . Por lo tanto, el PC se basa en resolver problemas haciendo uso de conceptos básicos de la programación [13] [14].

Durante el proceso de resolver un problema, el pensamiento computacional se pone en juego cuando [6] [11] [14]:

- Se descompone el mismo en pequeños subproblemas; denominado proceso de **Descomposición**. El mismo consiste en dividir un problema complejo (situación o tarea) en subproblemas más pequeños y manejables cuyas soluciones combinadas proveen la solución al problema general.
- Se centra la atención en las características más importantes; denominado proceso de **Abstracción**. Se pretende captar la esencia del problema filtrando las características no fundamentales y conservando los rasgos más relevantes, para luego crear una representación o modelo simplificado del mismo.
- Se utiliza el conocimiento de problemas similares resueltos con anterioridad; denominado proceso de **Reconocimiento de Patrones**. Consiste en buscar similitudes entre distintos problemas y dentro de un mismo problema. Se trata de encontrar similitudes o patrones de un problema (o subproblema) complejo con otro similar ya analizado-y resuelto efectivamente. Mientras

más patrones se reconozcan, más fácil y rápida será la tarea general de resolver problemas.

- Se elabora un plan de acción a ejecutar: **Algoritmo**. Consiste en un conjunto de instrucciones claras y precisas, que se identifican y se planifican en un determinado orden para la resolución a un problema.

La experiencia presentada en este trabajo consistió en promover, desde un sentido más general, prácticas para el desarrollo de las habilidades de pensamiento. Las mismas se centraron en actividades de resolución de problemas que desarrollan cuatros pilares del Pensamiento Computacional, utilizando como soporte diferentes herramientas de programación e incorporando el modelo lúdico como elemento de motivación. La iniciativa surge en el marco del proyecto de investigación “Estrategias para la Mejora de la Enseñanza de la Programación a Alumnos Ingresantes de las carreras de Ciencias e Ingeniería” y en concordancia con los objetivos planteados en su definición [1] [15].

El presente trabajo se organiza de la siguiente manera: en la sección 2 se explica cómo se diseñó el curso en función de los potenciales alumnos, las características del mismo y los instrumentos de recolección de datos; en la sección 3 se hace el análisis de los resultados obtenidos a partir de las encuestas realizadas, finalmente en la sección 4 se presentan algunas conclusiones y trabajos futuros.

2 Metodología del curso

Para la implementación de la experiencia que permitiera indagar qué habilidades relacionadas al PC ponen en juego los alumnos que ingresan a la UNSL en lo que respecta al aprendizaje de la programación en carreras de Ciencias e Ingeniería, se diseñó el curso “Inicio a la Programación”. El mismo estuvo orientado a alumnos que ingresaban en el ciclo 2017 a la Facultad de Ciencias Físico Matemáticas y Naturales de la UNSL, con el objetivo de introducirlos en la práctica de resolución de problemas usando fundamentos del PC. [17]

Debido a que en este período los alumnos deben cursar en forma obligatoria dos cursos de nivelación y evaluaciones diagnósticas, se optó por la modalidad B-learning para la organización del curso. Se estructuró en tres semanas completas de actividad, cerrando cada una con un encuentro presencial.

El aspecto no presencial fue cubierto mediante un aula virtual en la plataforma Moodle del Campus virtual de la UNSL, lo que facilitó la comunicación, organización y distribución de los recursos necesarios para la realización del curso [16].

El material de estudio fue desarrollado por los docentes responsables del dictado y puesto a disposición de los estudiantes semanalmente. En ellos se plantearon los objetivos que se pretendían alcanzar, el marco teórico y las actividades a realizar para afianzar dichos contenidos. Así mismo, se incorporaron diferentes recursos digitales (audio, video, enlace web) que complementaban el material de estudio.

Para poner en práctica los contenidos abordados y favorecer el desarrollo de las habilidades planteadas en los objetivos propuestos, las actividades fueron programadas con una modalidad integradora y complejidad incremental.

La comunicación entre los participantes del aula virtual se incentivó a través de la habilitación de foros para debatir e intercambiar ideas, conocimientos y dudas, y el uso de mensajería privada especialmente pensada para la comunicación entre el equipo docente y cada estudiante.

Los contenidos y las actividades fueron organizados de la siguiente manera:

Semana 1

- Resolución de problemas computacionales y no computacionales
- Pilares del Pensamiento Computacional: descomposición, reconocimiento de patrones, abstracción y algoritmo

Semana 2

- Resolución de problemas. Estrategias.
- Estructuras de Control: Secuenciales, Repeticiones (simples y condicionales).

Semana 3

- Estructuras de Control: Condicionales.
- Variables. Parametrización.

Con el propósito de obtener la información necesaria para realizar a posteriori un balance del curso y evaluar el impacto y la relevancia de la experiencia, se planificaron 2 cuestionarios. Un cuestionario previo al inicio de las actividades del curso y otro posterior a la finalización de las mismas. El primer cuestionario pretendía indagar los conocimientos previos de los alumnos para poder focalizar la orientación de las actividades durante el dictado, mientras que el cuestionario a posteriori pretendió observar la utilización apropiada de la terminología abordada en el curso, así como también la comprensión de los nuevos conceptos.

Al finalizar el tercer y último encuentro se habilitó un espacio de diálogo donde los participantes pudieron valorar verbalmente la experiencia. De este intercambio se registraron algunos criterios a tener en cuenta sobre los contenidos trabajados, su relevancia, la modalidad propuesta y la integración de los contenidos de este curso con las materias curriculares del primer año en cada carrera.

3 Evaluación de Resultados

La experiencia se planificó tomando como población al conjunto de 429 alumnos pre-inscriptos, para el año 2017, en las carreras de la FCFMyN. Un alumno reviste la condición de pre-inscripto a partir del momento que ha iniciado el proceso de inscripción y no necesariamente lo ha finalizado ó ha confirmado su matriculación.

A los fines de configurar la muestra se procedió con el análisis y depuración de los datos que los potenciales alumnos declararon, eliminando registros repetidos o con errores. De la muestra obtenida, fueron seleccionados aquellas personas que habían expresado su intención de ser alumnos de carreras que en el primer año de su plan de estudio tuvieran alguna materia vinculada a la programación. Resultó así un muestreo de 100 personas a quienes se los convocó a través del correo electrónico declarado durante la pre-inscripción. El 50% de los alumnos respondieron a la invitación, expresando su conformidad de participar, en forma voluntaria, de esta experiencia. (Ver tabla 1).

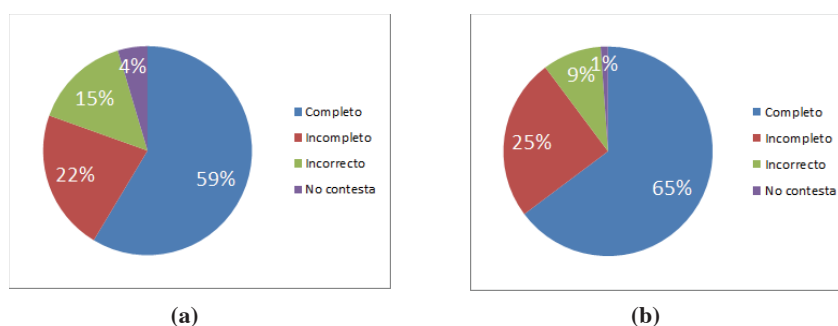
Tabla 1. Tamaño de la muestra en las distintas instancias

Pre inscriptos FCFMyN	Convocados	Se interesaron en la propuesta	Asistieron al menos una vez	Terminaron el curso
429	100	50	40	22

La recopilación de datos se realizaron en dos momentos determinados y a través de cuestionarios con la intención de determinar las condiciones y características del grupo al momento del inicio del curso (Pre) y al terminar (Pos). Ambos cuestionarios mantuvieron características similares y fueron estructurados en 8 ejercicios prácticos comprendiendo todos los pilares del PC abordados. Cada pilar contó con dos ejercicios prácticos y la respuesta a cada uno fue calificada según los siguientes criterios: *Completo*, *Incompleto*, *Incorrecto* y *No contesta*.

El análisis se organiza por niveles desde una evaluación general del impacto de la experiencia a un estudio más específico sobre cada ejercicio planteado.

La Figura 1 representa los resultados globales de los porcentajes obtenidos por tipo de respuesta en los cuestionarios Pre y Pos. De la comparación se observa que los resultados alcanzados en el cuestionario Pos mejoraron con respecto a los obtenidos en el Pre. Se redujo aproximadamente en un 50% el porcentaje de ejercicios sin resolver o resueltos de manera incorrecta, aumentando el porcentaje de ejercicios Completos o Incompletos. El índice de respuesta *Completo* aumentó un 6% mientras que la cantidad de respuestas de tipo *Incompleto* aumentó en un 3%. Los alumnos intentaron resolver el 90% de los ejercicios lo que significa que comprendieron las consignas de los mismos. Se infiere, entonces, que los alumnos, en un porcentaje muy significativo, supieron resolver el problema o entendieron la consigna pero no llegaron a la solución.

**Fig. 1.** Comparación entre Cuestionarios Pre y Pos, totales globales por tipo de respuesta.

En la Figura 2, se muestran los resultados del cuestionario Pre (a) y el cuestionario Pos (b) agrupados por tipo de respuesta distinguiendo los diferentes pilares. Muestra la distribución de los porcentajes vistos en la Figura 1, especificados por cada uno de los pilares trabajados. En el caso del gráfico (a), en las respuestas de tipo Completo los ejercicios asociados al pilar de *Abstracción* claramente fueron los mejores resueltos mientras que en la *Descomposición*, el *Reconocimiento de Patrones* y en la definición del Algoritmo se evidencia que, si bien encontraron mayor dificultad al resolverlos, el desempeño fue similar y relativamente bueno. Al analizar las

respuestas de tipo *Completo* del gráfico (b) se visualiza una distribución en los distintos pilares más homogénea que en el caso del gráfico (a), mejorando el desempeño en todos ellos. La *Abstracción* es el segundo pilar en que sobresalen luego de la *Descomposición* que mejoró significativamente. Se infiere entonces, que los alumnos fortalecieron en su estrategia el análisis del problema, que les permitió identificar los subproblemas que componen al problema planteado y asegurar un buen desempeño en las fases restantes.

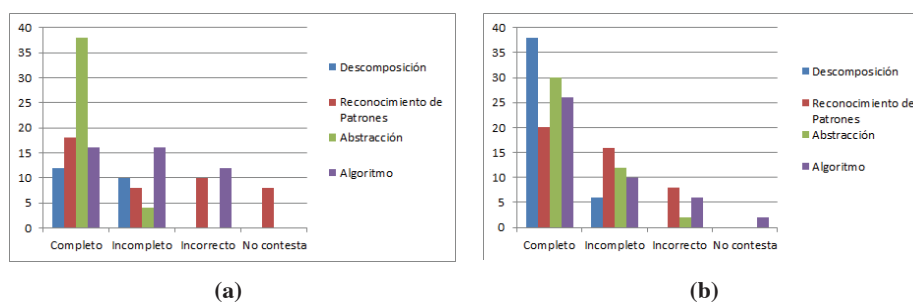


Fig. 2. Comparación Cuestionarios Pre y Pos, totales por cada pilar y tipo de respuesta.

En la Figura 3, se muestran los gráficos con un mayor detalle permitiendo analizar cada pilar en función del tipo de respuesta obtenidas en ambos cuestionarios. El gráfico (a) muestra los valores del pilar de *Descomposición*, donde se visualizan los mejores resultados alcanzados, ya que el incremento del tipo de respuestas *Completo* es significativo en el Cuestionario Pos. En las dos instancias, todos los alumnos respondieron, un porcentaje muy alto lo hizo de manera incorrecta en el Pre mientras que en el Pos ninguno lo resolvió mal o no lo hizo. Por otro lado, en el gráfico (b), correspondiente a *Reconocimiento de patrones*, la mejora en la respuesta del tipo *Completo* no fue tan relevante como en el gráfico anterior. Sin embargo, se observa un significativo incremento en las respuestas de tipo *Incompleto* del Cuestionario Pos. Esto podría mantener la relación con la disminución en las respuestas de tipo *Incorrecto* y la ausencia de respuestas en *No contesta*. El pilar de *Abstracción*, como se visualiza en el gráfico (c), arrojó los resultados menos satisfactorios, lo que lleva a plantear una evaluación más detallada del abordaje de este pilar durante el curso, con la intención de establecer los aspectos a mejorar. Finalmente, en el gráfico (d) que corresponde al pilar *Algoritmo* se destaca el incremento de tipo de respuestas *Completo* y se acentúa la disminución de los tipos de respuestas *Incompleto* e *Incorrecto*. Se infiere entonces, que se logró una notable mejora en la definición de algoritmos al finalizar el curso.

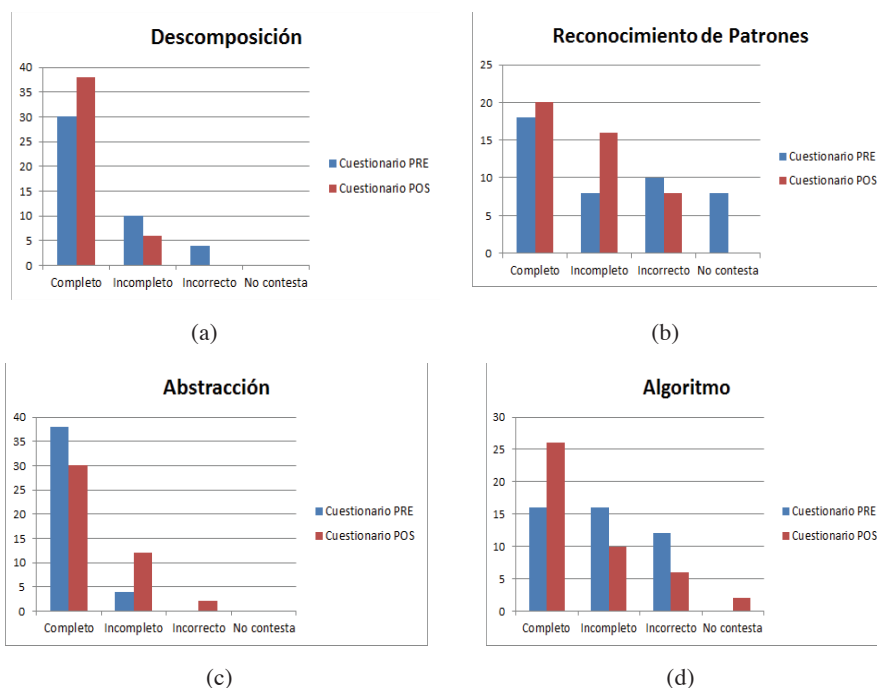


Fig. 3. Comparación entre Cuestionario Pre y Pos cada ejercicio por tipo de respuesta.

En la Figura 4, se pueden interpretar con un mayor detalle los resultados por ejercicio, mostrando la cantidad de tipos de respuesta en los cuestionarios Pre (a) y Pos (b). Una primera observación, permite identificar que las respuestas de tipo *Completo* se destacan en el cuestionario Pos, ya que se separan del resto de los tipos de respuesta en la mayoría de los ejercicios. Así mismo, se destaca que la respuesta de tipo *No contesta* en el cuestionario Pos se registró en un único ejercicio. Una comparación más profunda con respecto a cada ejercicio permite visualizar aquellos que mostraron una baja en el cuestionario Pos, en la cantidad de respuestas de tipo *Completo*. De este modo se plantea la necesidad de indagar las posibles causas. Entre las acciones que se podrían llevar a cabo se encuentran el revisar la práctica desarrollada durante el curso, seleccionando cuidadosamente los problemas y ejercicios, y el control de la complejidad y legibilidad de los enunciados en ambos cuestionarios.

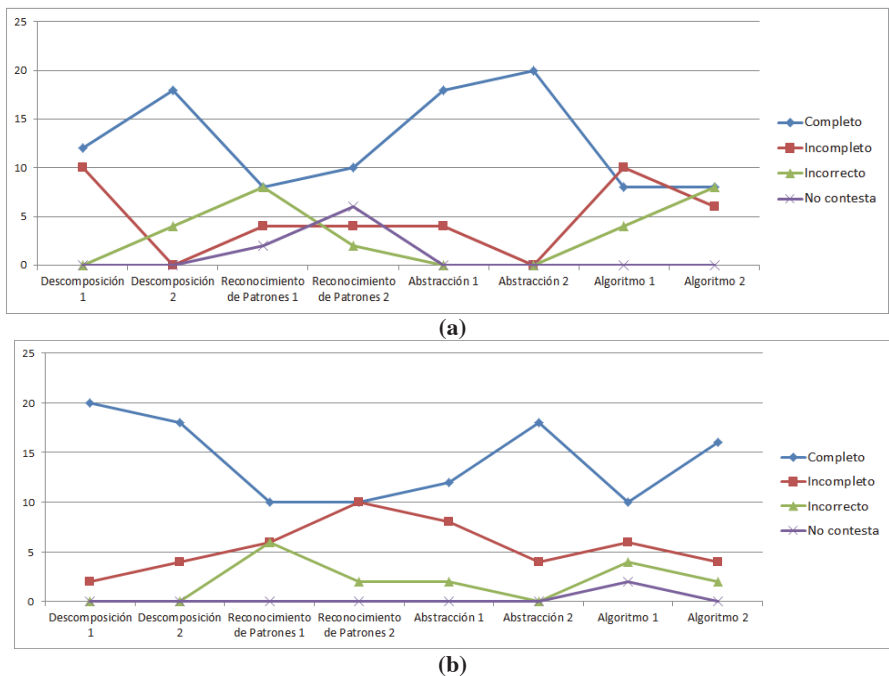


Fig. 4. Comparación entre Cuestionario Pre y Pos cada ejercicio por tipo de respuesta.

Como cierre del cuestionario Pos los alumnos tuvieron la posibilidad de autoevaluar su participación en las diferentes actividades propuestas en el curso y a través del campus. La Figura 5 refleja los porcentajes de actividades realizadas según los manifestado. Los alumnos, en su mayoría, justificaron estos valores aduciendo que si bien les resultaba muy entretenido el curso y las actividades propuestas eran motivadoras, dieron prioridad a los trabajos requeridos en los cursos obligatorios del ingreso. Más del 60% de los estudiantes cumplieron con varios o todos los requerimientos realizados. Se infiere entonces que, a pesar del nivel de participación relativamente bajo, el curso les resultó de interés, donde se propusieron problemas atractivos y la complejidad de las herramientas utilizadas estuvieron acordes a la situación. Los resultados obtenidos muestran un impacto general positivo.

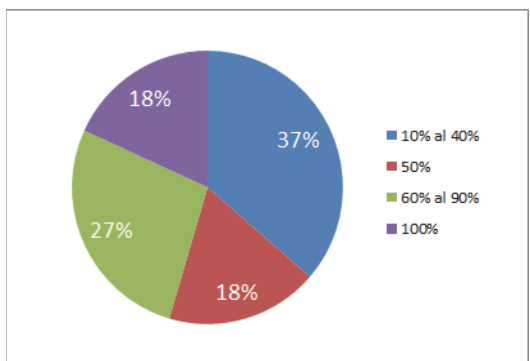


Fig. 5. Porcentaje de realización de las actividades propuestas en el aula virtual.

4 Conclusiones y Trabajo Futuro

Del análisis de la realidad áulica realizado en el período 2013-2016 en las asignaturas vinculadas a la resolución de problemas y/o programación de primer año de carreras de la FCFMyN, se detectó el bajo rendimiento académico de los estudiantes, el alto porcentaje de deserción en el primer año y el significativo desgranamiento en los años superiores, fundamentalmente en carreras afines a la informática. Así mismo se visualizó que el desgranamiento se acentuaba en los años superiores de cada carrera. En función de los datos registrados y la indagación sobre las posibles causas se configuraron diferentes acciones a llevar a cabo con la intención de atenuar esta situación. En este sentido el presente trabajo intenta contribuir en uno de los posibles factores que influyen en el bajo rendimiento académico, la diferencia entre los conocimientos previos de los futuros alumnos universitarios y los nuevos conceptos y prácticas impartidas en las materias vinculadas a la resolución de problemas y/o programación del primer año académico.

Del dictado del curso “*Inicio a la Programación*”, de las observaciones realizadas y del análisis de los datos recogidos es posible categorizar a la experiencia como muy positiva, en tanto se logró:

- Enriquecer los conocimientos previos de los alumnos participantes, sobre prácticas de resolución de problemas a partir del PC.
- Generar un espacio para propiciar la nivelación, en relación a los conocimientos previos, dada la variedad de orientaciones académicas que presentaron los alumnos participantes.
- Favorecer los tiempos de aprestamiento de los estudiantes ingresantes a la vida universitaria estimulando el sentido de pertenencia a la institución.
- Integrar prácticas en modalidad B-learning, favoreciendo el acceso a los contenidos y la interacción docente-alumno y alumno-alumno.
- El reconocimiento desde el Dpto. de Informática, espacio donde se valorizó la propuesta desarrollada.

Para el futuro, se espera:

- Dictar el curso en carácter obligatorio, en conjunto con los cursos de ingreso como Matemática y Comprensión lectora. En el ámbito de la Facultad de Ciencias Físico Matemáticas y Naturales de la UNSL.
- Lograr la adecuación de las prácticas propuestas en función de los resultados obtenidos y los aspectos débiles observados.
- Reforzar la propuesta con mecanismos que promuevan la participación y la realización de la mayor cantidad de actividades planteadas en función de incrementar positivamente los resultados obtenidos.

Referencias

- [1] Rosas M., Zúñiga M., Fernández J., Guerrero R.: El Pensamiento Computacional en el Ámbito Universitario. XIX Workshop de Investigadores en Ciencias de la Computación (WICC 2017), Instituto Tecnológico de Buenos Aires, Buenos Aires, Argentina. Aceptado. Aún no publicado. (2017)

- [2] Bit & Byte. Referentes de diferentes universidades argentinas opinan sobre la formación de profesionales informáticos. Año 2, no. 4. ISSN: 2468-9564. p. 23-28 (2016)
- [3] Simari G. Los fundamentos computacionales como parte de las ciencias básicas en las terminales de la disciplina Informática. VIII Congreso de Tecnología en Educación y Educación en Tecnología. UNSE. Santiago del Estero. (2013)
- [4] Szpiniak D., Rojo G.: Enseñanza de la programación. TE&ET: Revista Iberoamericana. (2006).
- [5] Dapozo G., Greiner C., Pedrozo Petrazzin G., Chiapello, J.: Investigación para fortalecer actividades de promoción y retención de alumnos en carreras de Informática. Anales del XX Congreso Argentino de Ciencias de la Computación. 1a ed. San Justo. Buenos Aires. Universidad Nacional de La Matanza, 2014. ISBN 978-987-3806-05-6. (2014).
- [6] Compañ-Rosique, P., Satorre-Cuerda, R.: Enseñando a programar: un camino directo para desarrollar el pensamiento computacional. RED-Revista de Educación a Distancia, 46(11). 15-Sept.-2015 DOI: 10.6018/red/46/11 http://www.um.es/ead/red/46/faraon_et_al.pdf (2015)
- [7] Dapozo G., Greiner C., Petris R.: Herramientas lúdicas como apoyo a la enseñanza de la programación. XXII Congreso Argentino de Ciencias de la Computación (CACIC 2016). Octubre 2016. Universidad Nacional de San Luis. (2016)
- [8] Dapozo G., Petris C., Greiner C., Espíndola M., Company A., López M.: Capacitación en programación para incorporar el pensamiento computacional en las escuelas. X Congreso de Tecnología en Educación y Educación en Tecnología (TE&ET 2015). ISBN 978-987-3977-30-5. Universidad de Morón. Buenos Aires. (2015)
- [9] Bravo-Lillo, C.: Pensamiento Computacional: una idea a la que le llegó el momento. Bits de Ciencia- Revista del Departamento de Ciencias de la Computación de la Universidad de Chile, 48-51. (2015)
- [10] Brookshear, J.: Introducción a las Ciencias de la Computación. Wilmington Delaware (U.S.A.): Addison-Wesley Iberoamerican, S.A. (1995)
- [11] Zapata-Ros.: Pensamiento computacional: Una nueva alfabetización digital. RED-Revista de Educación a Distancia, N° 46. (2015)
- [12] Wing J.: Computational Thinking: What and Why? Disponible en <http://www.cs.cmu.edu/~CompThink/resources/TheLinkWing.pdf> (2010)
- [13] Wing J.: Computational Thinking and Thinking about Computing. Philosophical transactions. Series A, Mathematical, physical, and engineering sciences, pp. 3717-3725. (2008)
- [14] Wing J.: Computational thinking. Commun. ACM 49, 3 (March 2006), 33-35. DOI: <https://doi.org/10.1145/1118178.1118215> (2006)
- [15] Zúñiga M., Rosas M., Fernández J., Guerrero R.: El Desarrollo del Pensamiento Computacional para la Resolución de Problemas en la Enseñanza Inicial de la Programación. XVI Workshop de Investigadores en Ciencias de la Computación (WICC 2014), Universidad Nacional de Tierra del Fuego, Ushuaia, Tierra del Fuego, Argentina. ISBN 978-950-34-1084-4. pp. 340-343 (2014)
- [16] <https://moodle.org>.
- [17] Linn, M., Aho, A., Blake, B., et al.: Report of a Workshop on the pedagogical aspects of COMPUTATIONAL THINKING. The National Academies Press. ISBN 978-0-309-21474-2. (2011)

Explotación de Información para el análisis de los procesos de enseñanza. Caso de estudio: el rol del Ingeniero y las percepciones de los alumnos.

Luciano Straccia¹, Ariel Deroche¹, Pablo Pytel¹, Ma Florencia Pollo-Cattaneo¹

¹Grupo de Estudio en Metodologías de Ingeniería de Software (GEMIS). Universidad Tecnológica Nacional. Facultad Regional Buenos Aires. Argentina.
lstraccia@frba.utn.edu.ar, arielderoche@gmail.com, ppytel@gmail.com, flo.pollo@gmail.com

Resumen. La información es un activo fundamental para las organizaciones que buscan mejorar sus procesos. Es relevante que la toma de decisiones sobre los procesos de enseñanza en las Universidades también se realice a partir de información objetiva y confiable. Las técnicas de Explotación de Información, asociadas a la Inteligencia de Negocios, pueden otorgar ventajas significativas para la mejora de los procesos educativos. En este trabajo se presenta el análisis de encuestas realizadas a alumnos que se inician en la carrera de Ingeniería en Sistemas de Información de UTN-FRBA con el objetivo de obtener información acerca de sus percepciones sobre el rol del profesional; para ello se utilizan técnicas de Explotación de Información que son presentadas de manera detallada para que los docentes y gestores académicos puedan aplicarlas independientemente de sus conocimientos en el área.

Palabras Claves: formación de ingenieros; rol del ingeniero; Explotación de Información; inteligencia de negocios; C4.5.

1 Introducción

Dentro del ámbito de la Facultad Regional Buenos Aires de la Universidad Tecnológica Nacional (UTN-FRBA) se ha conformado el Grupo GEMIS (Grupo de Estudios de Metodología para Ingeniería de Software y Sistemas de Información), un equipo de personas con interés en la investigación vinculada a Ingeniería en Sistemas de Información y Tecnología Aplicada a la Educación, que lleva adelante el Proyecto de Investigación y Desarrollo (PID) denominado “Intervenciones tecnológicas en dispositivos didácticos con herramientas de tecnología informática” cuyo objetivo es describir y analizar el uso de la tecnología informática en las intervenciones didácticas de los profesores de las asignaturas de la carrera de Ingeniería en Sistemas de Información (ISI) de la UTN-FRBA y desarrollar nuevos artefactos tecnológicos que favorezcan la mejora en las intervenciones didácticas.

En este trabajo se busca indagar, mediante el aprovechamiento de herramientas informáticas y técnicas avanzadas de análisis de información, en las percepciones de los alumnos respecto a la carrera y profesión que han escogido, realizando este

análisis al principio y finalización del primer cuatrimestre de cursada, buscando especialmente analizar el impacto que los procesos de enseñanza y aprendizaje desplegados a lo largo del cuatrimestre, pudieran tener sobre estas percepciones y qué variables contextuales pueden explicarlas.

Este trabajo presenta el contexto en el cual se desarrolla (sección 2), los objetivos previstos (sección 3), los materiales y métodos (sección 4) y los resultados obtenidos por la ejecución de los procesos de Explotación de Información (sección 5). En la sección 6 se presentan las conclusiones y futuras líneas de trabajo.

2 Contexto

2.1. Universidad

La carrera Ingeniería en Sistemas de Información en la UTN posee un tronco integrador que “está constituido por un conjunto de materias cuya finalidad es crear (...) un espacio de estudio multidisciplinario de síntesis, que permita al alumno conocer las características del trabajo ingenieril” [1]. En el plan de estudios la materia integradora del primer nivel es Sistemas y Organizaciones.

El tronco integrador de la carrera es la respuesta a través de los diseños curriculares al perfil de graduado esperado y a las problemáticas asociadas a "la apatencia del alumno que viene a la Universidad a buscar, desde el inicio de su carrera, temas que lo vinculen a la profesión que él ha elegido y que, al no encontrarlos, sobreviene su desmotivación", siendo esta uno de los factores de deserción. [2] Los niveles de deserción inicial son altos en la actualidad, llegando en la UTN-FRBA aproximadamente al 30% sólo en el primer año de la cursada [3]. Las causas que llevan a un alumno a dejar de cursar son diversas y no es objetivo de este trabajo indagar en ellas, sin embargo resulta particularmente interesante poner en discusión qué aportes pueden realizar las asignaturas (especialmente las asociadas al tronco integrador) para el acompañamiento de estas situaciones.

Se presentan a continuación algunos desafíos actuales del ámbito universitario:

- la mera definición de un plan de estudios, asignaturas asociadas y los programas de ellas no son más que declaraciones que deben ser llevadas a la práctica cotidiana por los propios docentes de las materias. La relación entre los ámbitos de gestión educativa y la actividad docente es necesaria para dar cumplimiento a los objetivos previstos, siendo ésta no siempre lo necesariamente eficiente.
- la aplicación de tecnología informática en los procesos de análisis de información y toma de decisiones en los ámbitos de gestión, es una práctica necesaria. Y la toma de decisiones por parte de estos equipos acerca de qué tecnologías utilizar, cómo y el aprovechamiento de esos resultados resulta muy relevante. Por ello, a lo largo de este trabajo se presentarán las diversas herramientas de tecnología informática utilizadas, su justificación y su mecanismo de aplicación.

- el uso de encuestas es una práctica habitual en las Universidades y no siempre los docentes, titulares de cátedra u otras autoridades cuentan con las habilidades y conocimientos requeridos para la real explotación de la información que permita optimizar la toma de decisiones.

2.2. El perfil del profesional

Según el perfil profesional de la carrera de Ingeniería en Sistemas de Información en UTN-FRBA “el ingeniero en Sistemas de Información es un profesional de sólida formación analítica que le permite la interpretación y resolución de problemas mediante el empleo de metodologías de sistemas y tecnologías de procesamiento de información” [4]. Krick [5] afirma que “un ingeniero es un solucionador de problemas”, poniendo énfasis en la necesidad de identificar y analizar problemas y, no sólo, en la mera construcción de artefactos (especialmente software).

Como se observa en [6], libro que forma parte de la bibliografía obligatoria de la asignatura Sistemas y Organizaciones, “en reiteradas oportunidades, se observa que la mayoría de los alumnos ingresantes a las carreras de Ingeniería en Sistemas de Información, no tienen respuesta para tales cuestiones y arriesgan dubitativamente proposiciones que aluden más al uso de una herramienta (la computadora) y a una técnica (la programación)”. La asignatura Sistemas y Organizaciones, como asignatura del tronco integrador de la carrera y ubicada en el primer nivel del plan de estudios, se plantea entre sus objetivos clarificar a los estudiantes que inician la carrera sobre el perfil profesional de un ingeniero en sistemas de información.

A partir de lo expresado en los párrafos anteriores se hace necesario trabajar sobre las percepciones de los alumnos sobre la vinculación de la programación y la resolución de problemas con el rol del ingeniero en sistemas de información.

3 Objetivos

El presente trabajo propone la aplicación de técnicas de Explotación de Información sobre encuestas realizadas a los alumnos del primer nivel de ISI con el objetivo de obtener datos significativos que favorezca la toma de decisiones sobre las actividades que desarrollan los cuerpos docentes. Además, dado que el trabajo se realiza en el marco de proyectos de investigación y desarrollo de equipos del propio campo disciplinar de Sistemas de Información y Tecnología de Información, se busca aplicar técnicas propias de la Ingeniería de Software. Finalmente, se espera que las herramientas y técnicas aquí descriptas puedan ser utilizadas por cualquier equipo docente o de gestión para indagar en las percepciones de sus alumnos y los resultados de sus propios procesos de enseñanza.

4 Materiales y métodos

Para la recolección de datos se han llevado a cabo encuestas en todos los cursos de la asignatura Sistemas y Organizaciones en los años 2016 y 2017 [7], arrojando un total de 690 alumnos encuestados en el año 2016 y 723 alumnos encuestados en el año 2017 en las encuestas iniciales y 570 alumnos y 684, respectivamente, en las encuestas llevadas a cabo al final del cuatrimestre.

La encuesta fue organizada en cinco secciones: datos generales del alumno, sus percepciones sobre la vinculación de la carrera con ciertas áreas, sus percepciones sobre las actividades asociadas a la carrera, su conocimiento sobre algunos aspectos del plan de estudios (especialmente respecto de asignaturas electivas, dada la relevancia que las asignaturas electivas poseen en la formación del ingeniero [8]) y, una pregunta abierta final donde se le solicita “explicar brevemente con sus palabras en qué consiste la labor del Ingeniero en Sistemas de Información”. En aquellos casos que se trató de preguntas cerradas con escala numérica, la escala utilizada fue 1 a 5.

Las preguntas utilizadas para el presente trabajo son aquellas que permitan dar respuestas especialmente a los objetivos planteados, siendo las restantes preguntas fuentes de información relevantes para otros trabajos de equipos de investigación o equipos docentes. Se considerarán en este trabajo la información general del alumno (de la primera sección de la encuesta), la calificación asociada al ítem de Programación (en adelante ítem a, de la segunda sección) y la calificación asociada a los ítems Desarrollo de Productos-Solución (ítem n) y Detección, análisis y resolución de problemas (ítem q), ambos de la tercera sección. Y algunas de las preguntas restantes sólo son utilizadas para analizar la percepción general de los estudiantes en cada área.

En este trabajo se ha aplicado un proceso de Explotación de Información a través de un método basado en sistemas inteligentes denominado TDIDT (*Top Down Induction Decision Trees*). Según Watson [9] la inteligencia de negocios “es una categoría amplia de aplicaciones, tecnologías y procesos para la recolección, almacenamiento, el acceso y análisis de datos para ayudar a los usuarios a tomar mejores decisiones de negocio” [10]. La Explotación de Información es “la subdisciplina informática que aporta a la Inteligencia de Negocio las herramientas para la transformación de información en conocimiento. Se ha definido como la búsqueda de patrones interesantes y de regularidades importantes en grandes masas de información” [11].

Para la aplicación de estos algoritmos se debe identificar la variable clase y las variables explicativas. La variable clase será aquella cuyo comportamiento se desea evaluar (en este trabajo, por ejemplo, la variación de la percepción de los alumnos entre el inicio y el final del cuatrimestre) y, las variables explicativas son aquellas variables que se intenta evaluar su impacto sobre la variable clase (en este trabajo, por ejemplo, la situación laboral de los alumnos, entre otras).

Existen diferentes algoritmos TDIDT que permiten “el desarrollo de descripciones simbólicas de los datos para diferenciar entre distintas clases” [11]. Para este proyecto, entre los algoritmos TDIDT, se ha seleccionado el C4.5 [12], que genera un árbol de decisión a partir de los datos y “considera todas las pruebas posibles que pueden dividir el conjunto de datos y selecciona la prueba que resulta en la mayor ganancia de información. Para cada atributo discreto, se considera una

prueba con n resultados, siendo n el número de valores posibles que puede tomar el atributo. Para cada atributo continuo, se realiza una prueba binaria sobre cada uno de los valores que toma el atributo en los datos” [13].

Para realizar la Explotación de Información a través del algoritmo C4.5 se ha utilizado el software Tanagra [14]. Para su uso se requiere un archivo de formato filas-columnas que contengan en las columnas las variables a analizar y, en cada fila los valores asociados. Este archivo es ingresado dentro del software como *dataset* (conjunto de datos). Entre las variables definidas como columnas en el archivo que contiene los datos, y a través de la interfaz de usuario del software, algunas de ellas serán consideradas como variables explicativas (en Tanagra, se denominan input) y otras como variables clases (*class*).

5. Resultados

En esta sección se presenta el análisis de la Explotación de Información. En [15] pueden hallarse los reportes obtenidos a través del software Tanagra.

5.1. Percepciones generales del alumnado al inicio de la carrera

En esta sección se indica cuáles son las percepciones de los alumnos respecto de la carrera al inicio de la cursada, cómo estas varían cada año y si los resultados son similares a los esperados por la cátedra. Se compara la información de los resultados obtenidos en cada uno de los años y se realiza una comparación interanual. En la tabla 1 pueden hallarse los resultados para cada ítem de la encuesta.

Tabla 1. Percepciones por pregunta.

Ítem	2016	2017	Δ
Programación	3,84	3,82	-0,02
Computación	3,84	3,79	-0,05
Informática	4,04	4,02	-0,02
Administración	3,69	3,63	-0,06
Gestión	3,58	3,67	+0,11
Diseño de software	3,64	3,60	-0,04
Investigación	3,53	3,55	-0,02
Planeamiento	3,93	3,86	-0,07
Gestión de Recursos	3,46	3,50	+0,04
Desarrollo de Productos-Solución	4,03	4,01	-0,02
Diseño y Programación de aplicaciones informáticas	3,99	4,01	+0,02
Administración de circuitos informacionales	3,84	3,78	-0,07
Detección, análisis y resolución de problemas	4,59	4,47	-0,12
Auditorías y pericias sobre procesamiento de datos	3,73	3,69	-0,04
Investigación y simulación de modelos	3,63	3,67	+0,04
Desarrollo de lenguajes y técnicas de programación	3,61	3,70	+0,09
Planificación y control de la seguridad Informática	3,74	3,88	+0,14

Los ítems que los alumnos más relacionan con la carrera al iniciar el ciclo lectivo son la informática, el desarrollo de productos-solución y la detección, análisis y resolución de problemas. En general, los resultados obtenidos son positivos y siguen los objetivos de la asignatura. En las próximas secciones en este trabajo se realizan análisis más detallados sobre algunos de los ítems.

Además se observa que las diferencias entre el año 2016 y el año 2017 son menores, por lo cual puede generarse la hipótesis que los alumnos que ingresan lo harán con similares percepciones. La mayor variación se produce en planificación y control de la seguridad informática. Cabe preguntarse qué expectativas tienen respecto a esta temática los alumnos, siendo un tema de actualidad. Debido a ello, la cátedra ha tomado la decisión de incluir en el ciclo lectivo 2018 algunas temáticas vinculadas a la seguridad de la información (resguardos, planes de recuperación ante desastres), para introducir a los alumnos en las diferentes aristas de esta temática vinculadas a la profesión, con el objetivo de continuar incorporando temas relacionados con el rol del ingeniero y buscando generar conceptos más acabados de ciertos términos.

5.2. Percepciones respecto a la programación

En esta sección se analiza la percepción de los alumnos respecto a la programación, al inicio de la cursada y a la finalización del primer cuatrimestre. Para este análisis, se utilizaron las encuestas correspondientes al año 2017, por tratarse de las más recientes.

En la tabla 1 puede hallarse que el valor promedio otorgado por los alumnos en las encuestas de inicio de cursada al ítem Programación es 3,82. Sin embargo, entre los alumnos que completaron la encuesta de cierre el promedio original era 3,65. Se considera a este grupo de alumnos (los que completaron ambas encuestas), para realizar todo el análisis que sigue a continuación. La distribución de respuestas, según el nivel de importancia, puede encontrarse en la tabla 2.

Tabla 2. Valoración de la programación como actividad del Ingeniero en Sistemas al inicio.

Valoración	1	2	3	4	5
% de alumnos	3%	11%	27%	25%	24%

Puede observarse que entre la valoración 4 y 5 se halla el 49% de los alumnos. ¿Existe algún patrón asociado a otras variables que permita identificar por qué algunos alumnos otorgan valoración alta (puntuación 4 o 5) y otros valoración menor (1 a 3)? A fines de poder dar respuesta a esta pregunta, se ha procesado esta información en el software Tanagra, considerando esta valoración (alta o baja), como variable clase y, como variables explicativas, se han utilizado el turno de cursada (en adelante, [T]), la relación del título secundario con la informática ([TS]), la razón por la cual eligieron la carrera ([RE]) y la situación laboral ([SL]). A partir de estas variables clases y explicativas, se han conformado diversos escenarios (tabla 3).

Tabla 3. Escenarios con la valoración como variable clase.

Escenario	1	2	3	4
Variables explicativas	[T]	[TS]	[RE]	[SL]
Reglas generadas	SI	NO	NO	NO

Para el caso de los escenarios 2 al 4 no se produjeron resultados significativos. Para el primer escenario, se puede observar en la figura 1, las reglas generadas donde se concluye que existe una relación entre el turno de cursada y la valoración al inicio: en el turno mañana la valoración es alta, mientras que en el turno noche, baja.

Decision tree

- [T] Turno in [Mañana] then [VAI] Valoración Inicio A = Alta (61,42 % of 591 examples)
- [T] Turno in [Noche] then [VAI] Valoración Inicio A = Baja (55,56 % of 54 examples)
- [T] Turno in [Tarde] then [VAI] Valoración Inicio A = Baja (61,54 % of 39 examples)

Figura 1. Reglas generadas por la ejecución del escenario 1 de la tabla 3.

A partir de estos resultados se buscó si otra variable podría influir dentro de cada turno. Para ello, se generaron escenarios con atributos explicativos combinados, según el detalle de la tabla 4.

Tabla 4. Escenarios con la valoración como clase y atributos explicativos combinados

Escenario	1	2	3
Variables explicativas	[T] + [TS]	[T] + [RE]	[T] + [SL]
Reglas generadas	SI	SI	SI

En el escenario 1 se identificó que en el turno tarde los alumnos con título secundario asociado a informática, dieron valoración alta al ítem, mientras que en el turno noche los alumnos con el título secundario asociado a informática realizó la valoración opuesta (figura 2).

Decision tree

- [T] Turno in [Mañana] then [VAI] Valoración Inicio A = Alta (61,42 % of 591 examples)
- [T] Turno in [Noche]
 - [TS] ¿Su título secundario tiene relación con informática o sistemas? in [No] then [VAI] Valoración Inicio A = Alta (53,85 % of 39 examples)
 - [TS] ¿Su título secundario tiene relación con informática o sistemas? in [Si] then [VAI] Valoración Inicio A = Baja (80,00 % of 15 examples)
- [T] Turno in [Tarde]
 - [TS] ¿Su título secundario tiene relación con informática o sistemas? in [No] then [VAI] Valoración Inicio A = Baja (85,71 % of 21 examples)
 - [TS] ¿Su título secundario tiene relación con informática o sistemas? in [Si] then [VAI] Valoración Inicio A = Alta (66,67 % of 18 examples)

Figura 2. Reglas generadas por la ejecución del escenario 1 de la tabla 4.

En el escenario 2 se identificaron reglas para los alumnos del turno noche (figura 3), sin embargo la cantidad de valores asociados a cada una de estas reglas es reducido respecto al total general de alumnos.

- Decision tree
- [T] Turno in [Mañana] then [VAI] Valoración Inicio A = Alta (61,42 % of 591 examples)
 - [T] Turno in [Noche]
 - [RE] ¿Por qué razón decidió inscribirse en la carrera Ingeniería en Sistemas de Información? in [Interés en la currícula de la carrera] then [VAI] Valoración Inicio A = Baja (58,33 % of 36 examples)
 - [RE] ¿Por qué razón decidió inscribirse en la carrera Ingeniería en Sistemas de Información? in [Futuro laboral] then [VAI] Valoración Inicio A = Alta (75,00 % of 12 examples)
 - [RE] ¿Por qué razón decidió inscribirse en la carrera Ingeniería en Sistemas de Información? in [Continuar con la orientación del secundario] then [VAI] Valoración Inicio A = Baja (100,00 % of 3 examples)
 - [RE] ¿Por qué razón decidió inscribirse en la carrera Ingeniería en Sistemas de Información? in [Antecedente familiar (padres o hermanos)] then [VAI] Valoración Inicio A = Alta (0,00 % of 0 examples)
 - [RE] ¿Por qué razón decidió inscribirse en la carrera Ingeniería en Sistemas de Información? in [Afinidad con el trabajo actual] then [VAI] Valoración Inicio A = Baja (100,00 % of 3 examples)
 - [RE] ¿Por qué razón decidió inscribirse en la carrera Ingeniería en Sistemas de Información? in [Continuar con amigos] then [VAI] Valoración Inicio A = Alta (0,00 % of 0 examples)
 - [T] Turno in [Tarde] then [VAI] Valoración Inicio A = Baja (61,54 % of 39 examples)

Figura 3. Reglas generadas por la ejecución del escenario 2 de la tabla 4.

Entre las reglas generadas en el escenario 3, el resultado significativo es la diferenciación, sólo para alumnos de turno noche, entre el grupo de alumnos que trabaja (que valoran alto) y, aquellos que no lo hacen (que valoran bajo) (figura 4).

- Decision tree
- [T] Turno in [Mañana] then [VAI] Valoración Inicio A = Alta (61,42 % of 591 examples)
 - [T] Turno in [Noche]
 - [SL] ¿Trabaja actualmente? in [No] then [VAI] Valoración Inicio A = Baja (66,67 % of 27 examples)
 - [SL] ¿Trabaja actualmente? in [Si. En un trabajo no relacionado con la carrera.] then [VAI] Valoración Inicio A = Alta (60,00 % of 15 examples)
 - [SL] ¿Trabaja actualmente? in [Si. En un trabajo relacionado con la carrera.] then [VAI] Valoración Inicio A = Alta (50,00 % of 12 examples)
 - [T] Turno in [Tarde] then [VAI] Valoración Inicio A = Baja (61,54 % of 39 examples)

Figura 4. Reglas generadas por la ejecución del escenario 3 de la tabla 4.

Al analizar las respuestas de fin de cuatrimestre, el promedio de valoración del término programación como asociado a la carrera es 3,26, disminuyendo significativamente respecto de los valores iniciales. La distribución de la valoración por porcentaje de alumnos se puede hallar en la tabla 5.

Tabla 5. Valoración de la programación como actividad del Ing. en Sistemas.

Valoración	1	2	3	4	5
% de alumnos	7%	18%	30%	31%	14%

Puede observarse que, la puntuación alta disminuyó desde el 49% al 45% de los alumnos. A continuación se buscan hallar explicaciones para estas variaciones.

A partir de las valoraciones al inicio y al final se han conformado cuatro grupos de alumnos según el cambio producido en la valoración: “de alta a alta”, “alta a baja”, “baja a alta” y “baja a baja”. El 64% de los alumnos mantienen la valoración, mientras que, un 25% la disminuye y, el restante 11% la incrementa. Considerando como clase a esta variación de respuestas se han ejecutado diferentes escenarios conformados a partir de variables explicativas diferentes (tabla 6).

Tabla 6. Escenarios con la variación del ítem A como variable clase.

Escenario	1	2	3	4	5
Variables explicativas	[T]	[P]	[TS]	[RE]	[SL]
Reglas generadas	SI	SI	NO	NO	NO

Los resultados permiten definir que ninguna variable impacta significativamente en las variaciones de las percepciones, siendo que las reglas generadas sólo permiten identificar grupos en los cuales se mantienen las valoraciones.

5.3. Percepciones respecto al análisis de problemas y desarrollo de soluciones

En esta sección se repetirá el proceso seguido en la sección anterior, pero para otros ítems de la encuesta. Se omitirá el detalle de los escenarios planteados, aunque se definirán las variables utilizadas y los resultados obtenidos.

El valor promedio otorgado por los alumnos al ítem Desarrollo de Productos-Solución (ítem n) al inicio de la cursada es 4,01. Entre los alumnos que finalizaron la cursada el promedio es 4,03. El valor promedio otorgado por los alumnos al ítem Detección, análisis y resolución de problemas (q) al inicio de la cursada es 4,47. Entre los alumnos que finalizaron el cuatrimestre el promedio es 4,50. A diferencia del ítem de programación donde se producían variaciones relevantes entre el inicio y el cierre de la cursada, en este ítem las diferencias no son significativas. Se considera al grupo de alumnos que completaron ambas encuestas para realizar el análisis que sigue a continuación. Para estos ítems las respuestas se categorizan en tres conjuntos: valoración baja (puntuación 1 a 3), media (4) y alta (5).

La distribución de respuestas según el nivel de importancia para el ítem n ha sido el siguiente: el 25% otorgó valoración Baja; el 34%, Media y el 41%, Alta. En el ítem q las valoraciones han sido: el 10%, Baja; el 26%, Media y el 64%, Alta. En ambos casos, al realizar la Explotación de Información, vinculando estas categorías con las variables explicativas [T], [TS], [RE], [SL] no se produjeron hallazgos significativos.

En segundo lugar, se busca analizar la variación que cada alumno tuvo en sus respuestas, considerando como variable clase la Variación (por ejemplo, valoración baja al inicio y media al cierre). En el ítem n el 49% de los alumnos mantuvo su valoración, mientras que en el ítem q, el 57%. En el primer caso, el 29% incrementó su valoración; mientras que en q, lo hizo el 23%. Los restantes alumnos disminuyeron la valoración. A partir de estos datos se hace necesario buscar cuáles variables podrían explicar estas diferencias.

Para el ítem n se halló que la principal variable que afecta a los resultados es la situación laboral de los alumnos. Los alumnos que no trabajan incrementaron su valoración del ítem. Los alumnos que trabajan, en temas relacionados con la carrera, mantuvieron una valoración alta. Contrariamente a estos grupos, aquellos que trabajan en temas no vinculados a la carrera, disminuyeron su valoración del ítem. Al analizar las respuestas de alumnos de este último grupo se verifica que existe un grupo en particular que, a pesar de trabajar en temas no vinculados a la carrera, mantuvo alta la valoración en lugar de disminuirla; se trata de aquellos que eligen inscribirse en la carrera por antecedentes familiares. Para el ítem q no se hallaron resultados significativos que permitan comprender, a partir de las variables explicativas, las diferencias entre las respuestas.

6 Conclusiones

El proceso de Explotación de Información permite otorgar a los equipos docentes y de gestión universitaria resultados objetivos sobre sus temas de interés. En este trabajo se

analizó la percepción de los alumnos sobre algunas temáticas vinculadas sobre el perfil profesional y el perfil de la carrera.

Entre los hallazgos significativos pueden mencionarse la valoración alta de los alumnos sobre la resolución de problemas como parte de las acciones del ingeniero y la menor valoración de la programación. Se halló la relevancia que toman las temáticas de seguridad informática y que ameritan su tratamiento en el marco de la asignatura en que se desarrolló este trabajo para poner en discusión académica qué implica la participación del ingeniero en esa área.

Se halló también que algunos alumnos, a pesar de los trabajos realizados por la cátedra, modifican sus percepciones en camino contrario a los objetivos planteados. Esto, sumado a la falta de hallazgos del impacto de algunas variables (estudios secundarios, situación laboral, etc.) en la explicación de los cambios en la percepción de los alumnos, requiere un relevamiento detallado acerca de las prácticas que cada docente realiza sobre los temas analizados en este trabajo, con el objetivo de identificar nuevas variables, comportamientos, etc. que pudieran incidir sobre la percepción de los alumnos. Finalmente, dado que la asignatura tiene modalidad anual, se prevé llevar a cabo en el segundo cuatrimestre nuevas actividades que permitan otorgar mayor información para completar el análisis iniciado en este trabajo.

Referencias

1. UTN (2007). Ordenanza 1150. Aprobación del diseño curricular de la carrera de Ingeniería en Sistemas de Información. Universidad Tecnológica Nacional.
2. UTN-FRBB (2001). Revista UTEC Noticias. Número 8. Agosto 2001. Universidad Tecnológica Nacional. Facultad Regional Bahía Blanca.
3. La Voz de San Justo (2017). La UTN advirtió sobre el nivel de los ingresantes de las escuelas públicas a sus carreras de grado. Diario La Voz de San Justo. 19 de mayo de 2017.
4. DISI UTN-FRBA (s.f.). Perfil Profesional. Departamento de Ingeniería en Sistemas de Información. UTN FRBA.
5. Krick, E. (1999). Introducción a la Ingeniería y al diseño en la ingeniería. Limusa.
6. Pollo-Cattáneo, M.F. (2012). Resolviendo problemas en los sistemas de información: enfoque para informáticos. Centro de Estudiantes de Ingeniería Tecnológica, CEIT.
7. Grupo GEMIS. Templates de las encuestas realizadas. <https://goo.gl/rzBnPA>
8. Isola, A.; Guerrero, M.; Portillo, M.; Ferreiros Cabrera, S.; Bursztyn, A.; Granado Peralta, S. (2013). Las asignaturas electivas como eje en la formación del ingeniero.
9. Watson, H. (2009). Tutorial: Business Intelligence – Past, Present, and Future. Communication of the Association for Information Systems, Vol. 25.
10. Vega, J. J. C., Aguilar, L. J., & Marín, L. M. G. (2016). La inteligencia de negocios como una herramienta en la gestión académica. Revista Científica, 1(24), 120.
11. Britos, P.; García-Martínez, R. (2009). Propuesta de Procesos de Explotación de Información. Proceedings XV Congreso Argentino de Ciencias de la Computación.
12. Quinlan, J. R. (2014). C4. 5: programs for machine learning. Elsevier.
13. Servente, M. (2002). Algoritmos TDIDT aplicados a la Minería de Datos Inteligente. Tesis de Grado en Ingeniería Informática. Universidad de Buenos Aires. Facultad de Ingeniería.
14. Rakotomalala, R. (2005). Tanagra: data mining software for academic and research purposes. in Actes de EGC'2005, RNTI-E-3, vol. 2, pp. 697-702.
15. Grupo GEMIS (2017). Resultados de exportación de proceso de Explotación de Información en Tanagra. <https://goo.gl/V6crPW>

Incidencia de los estilos de aprendizaje de los estudiantes de informática en la valoración de un Objeto de Aprendizaje sobre Compuertas Lógicas

Natali Salazar Mesia^{1,2}, Cecilia Sanz^{2,3}, César Estrebou².

¹ Becaria Tipo A – Universidad Nacional de La Plata

²Instituto de Investigación en Informática LIDI. Facultad de Informática – Universidad Nacional de La Plata

³Investigador Asociado de la Comisión de Investigaciones Científicas de la Prov. de Buenos Aires

{nsalazar, csanz, cesarest}@lidi.info.unlp.edu.ar

Abstract. En este trabajo se presenta un estudio que busca relacionar los estilos de aprendizaje que presentan los estudiantes de primer año de una Carrera en Informática, con las actividades que se realizan en un Objeto de Aprendizaje, orientado a la comprensión del tema Compuertas Lógicas. Este material educativo ha sido utilizado en el marco de diferentes experiencias en una asignatura de primer año, mostrando los estudiantes un alto nivel de satisfacción con las actividades que éste presenta. Los resultados obtenidos vislumbran una incidencia entre los estilos de preferencia de los estudiantes, que son el teórico y el pragmático, y la satisfacción por las actividades realizadas que proponen vincular lo teórico con lo práctico. Se agregan otros resultados, que también ha permitido tomar decisiones sobre los tipos de actividades a realizar con estos grupos de estudiantes.

Keywords: Objetos de Aprendizaje, Estilos de Aprendizaje, Compuertas Lógicas

1 Introducción

Existen diferentes formas de aprender, algunas personas logran un mayor entendimiento con contenido audiovisual, otras prefieren la lectura de un texto descriptivo, y otras necesitan que el contenido sea interactivo para poder alcanzar el mismo nivel de comprensión.

Al mismo tiempo, hoy en día se visualiza un marcado crecimiento en la producción de materiales educativos digitales, con el objetivo de contribuir y mejorar los procesos de enseñar y aprender. Los Objetos de Aprendizaje (OA) son un ejemplo de estos materiales educativos digitales. Su utilización en el escenario educativo ha ido evolucionando, aunque se trata aún de un concepto dilemático [1], [2]. Los OA son materiales autónomos que permiten que los estudiantes puedan utilizarlos de forma independiente. El concepto adoptado para este trabajo es: “Un OA es un tipo de material educativo digital que: Desde el punto de vista pedagógico: se caracteriza

por orientarse a un objetivo específico de aprendizaje, y presentar mínimamente: una serie de contenidos con el fin de abordar la temática relacionada con el objetivo, al menos una actividad que permita al alumno poner en práctica o problematizar el contenido presentado, y al menos una autoevaluación que posibilite conocer al alumno, si ha podido comprender esos contenidos vinculados al objetivo. Desde el punto de vista tecnológico deberá: contener un conjunto de metadatos estandarizado que permitan su almacenamiento, búsqueda, recuperación y reutilización en diferentes entornos tecnológicos, estar integrado, preferentemente, utilizando un modelo de empaquetamiento que respete estándares y de esta manera, permita su diálogo con diferentes entornos tecnológicos” [3].

A su vez, existe una gran variedad de metodologías para el diseño y producción de OA. En este trabajo se pone el foco en un OA sobre Compuertas Lógicas, creado con el fin de introducir a los estudiantes en la comprensión del funcionamiento de estas compuertas. Para el desarrollo del OA que se presenta en este trabajo se utilizó la metodología CROA [4]. El OA se encuentra disponible en el SEDICI¹, repositorio institucional de la Universidad Nacional de La Plata y se lo viene utilizando en materias de primer año de carreras en Informática.

Por otro lado, es posible perfilar a los estudiantes de acuerdo a su forma de aprender a través de los estilos de aprendizaje. Un estilo de aprendizaje se define como las características, fortalezas y preferencias en relación a la forma en que las personas reciben y procesan la información. Cuando se habla de estilo de aprendizaje, se refiere al hecho de que cada persona tiene su propio modo o conjunto de estrategias cuando aprenden. Un modelo de estilo de aprendizaje clasifica a los estudiantes de acuerdo a dónde se ajustan en una serie de escalas relacionadas con la forma en que reciben y procesan la información [5].

De acuerdo a [6], existen variadas teorías acerca de los estilos de aprendizaje y hay diferentes cuestionarios que se pueden aplicar.

Kolb desarrolló un modelo de aprendizaje mediante experiencias para ser aplicado en la vida adulta de las personas. Expresa que el aprendiz necesita cuatro clases diferentes de capacidades: experiencia concreta, observación reflexiva, conceptualización abstracta, y experimentación activa. Realizó un instrumento al que denominó Inventario de Estilos de Aprendizaje (LSI) que es un cuestionario compuesto por doce series de palabras que es preciso ordenar por preferencia [7].

También, Felder y Silverman elaboraron un cuestionario llamado Índice de Estilo de Aprendizaje (ILS) para conocer las preferencias de aprendizaje en cuatro dimensiones: activa/reflexiva, sensitivo/intuitivo, visual/verbal, y secuencial/global. El cuestionario consta de 44 ítems que tienen un enunciado y dos opciones a elegir.

En este trabajo se utiliza el Cuestionario Honey-Alonso sobre Estilos de Aprendizaje (CHAEA), que es una adaptación al contexto académico español del *Learning Style Questionnaire* (LSQ) de Honey y Mumford, que a su vez se basa en el *Learning Style Inventory* (LSI) de Kolb [7], [8].

En el presente trabajo se busca correlacionar la valoración obtenida sobre las actividades del OA de Compuertas Lógicas con los estilos de aprendizaje de los estudiantes, a fin de obtener información que permitirá reconocer de qué manera se puede enriquecer el OA para que sea mejor aprovechado por estos estudiantes.

¹ URL del OA de Compuertas Lógicas: http://163.10.22.82/OAS/compuertas_logicas/

Este artículo se organiza de la siguiente manera, de aquí en adelante: en la sección 2 se presentan antecedentes sobre Objetos de Aprendizaje y Estilos de Aprendizaje, en la sección 3 se describe el Objeto de Aprendizaje Compuertas Lógicas, en la sección 4 se realiza la descripción de la experiencia, en la sección 5 se analizan los resultados obtenidos de la valoración del OA y el cuestionario CHAEA, y finalmente en la sección 6 se presentan conclusiones y trabajos futuros.

2 Antecedentes

La enseñanza de programación ha ido evolucionando con la incorporación de diferentes materiales educativos digitales que contribuyen a los procesos de enseñar y aprender.

En [9] se presenta un conjunto de Objetos de Aprendizaje para la enseñanza de programación estructurada utilizando el lenguaje de programación C. Los autores, diseñaron y desarrollaron los OA para que cada uno resuelva un problema pasando por las etapas de resolución de problemas: análisis, diseño, desarrollo y prueba con la posibilidad de ejecución. Organizaron todos los OA en un sitio web donde planifican su utilización posterior a cada tema dado en las clases teóricas. En [10] se describe el diseño de Objetos de Aprendizaje con el fin de ayudar a estudiantes de primer año a incorporar los conceptos básicos de programación y disminuir así el índice de deserción de estos cursos. Realizaron dos experiencias con OA sobre los temas: Punteros y Árboles Binarios que dieron resultados positivos en cuanto al aprendizaje y la motivación de los estudiantes. Estos antecedentes marcan un buen punto de partida para la utilización de OA en la enseñanza de carreras en Informática.

Por otra parte, la mayoría de los sistemas educativos no considera las características particulares de los estudiantes y utilizan el mismo contenido y las mismas estrategias pedagógicas para enseñar. La utilización del cuestionario CHAEA para identificar el estilo de aprendizaje de los estudiantes de las carreras informáticas puede ayudar a considerar esta cuestión.

El Cuestionario CHAEA consta de 80 afirmaciones, y está dividido en cuatro secciones de 20 ítems correspondientes a los cuatro estilos de aprendizaje (activo, reflexivo, teórico y pragmático). Es una prueba con puntuación dicotómica, de acuerdo (signo +) o en desacuerdo (signo -). La puntuación absoluta que el estudiante obtenga en cada sección indica el grado de preferencia [11]. Se utiliza un Baremo general basado en la muestra de una experiencia para identificar la preferencia de los estilos de aprendizaje en la siguiente escala: Muy Baja, Baja, Moderada, Alta y Muy Alta. La tabla 1 muestra este Baremo.

Tabla 1. Estandarización de los estilos de aprendizaje.

Estilo de Aprendizaje	Preferencia				
	Muy Baja (10%)	Baja (20%)	Moderada (40%)	Alta (20%)	Muy Alta (10%)
Activo	0-6	7-8	9-12	13-14	15-20
Reflexivo	0-10	11-13	14-17	18-19	20
Teórico	0-8	7-9	10-13	14-15	16-20
Pragmático	0-6	9-10	11-13	14-15	16-20

En la tabla 2 se puede observar una caracterización de cada estilo, donde se describen las situaciones en las que el alumno aprende mejor y las que les cuesta realizar en su proceso de aprendizaje.

Tabla 2. Se describen los diferentes estilos de aprendizaje del cuestionario CHAEA (tomado de [11]).

Estilos	Características Generales	Aprenden Mejor...	Les cuesta cuando...
Activo	<ul style="list-style-type: none"> Se involucran totalmente sin prejuicios en experiencias nuevas Disfrutan el presente y se dejan llevar por los acontecimientos Tienden a actuar primero y pensar después sin medir consecuencias Tienen muchas actividades Se rodean de gente y son el centro de las actividades. 	<ul style="list-style-type: none"> Se lanzan a actividades que le presentan desafíos Realizan actividades cortas con resultados inmediatos Hay emoción, drama y crisis 	<ul style="list-style-type: none"> Deben adoptar un papel pasivo Deben asimilar, analizar e interpretar datos Deben trabajar solos.
Reflexivo	<ul style="list-style-type: none"> Adoptan postura de observador que analiza sus experiencias desde muchas perspectivas distintas. Recogen datos y los analizan detalladamente antes de llegar a una conclusión. Son precavidos y analizan todas las implicaciones de cualquier acción. En las reuniones observan y escuchan, pasan desapercibidos. 	<ul style="list-style-type: none"> Adoptan la postura del observador. Ofrecen observaciones y analizan la situación. Piensen antes de actuar. 	<ul style="list-style-type: none"> Son forzados a convertirse en el centro de la atención. Se apresuran entre una actividad y otra. Deben actuar sin poder planificar previamente.
Teórico	<ul style="list-style-type: none"> Adaptan e integran las observaciones que realizan en teorías complejas y bien fundamentadas lógicamente. Piensen de forma secuencial y paso a paso. Analizan y sintetizan la información y premia la lógica y la racionalidad. Se incomodan con los juicios subjetivos y las técnicas de pensamiento lateral. 	<ul style="list-style-type: none"> A partir de modelos, teorías, sistemas de ideas y conceptos que presenten un desafío. Cuando pueden preguntar e indagar 	<ul style="list-style-type: none"> Con actividades ambiguas e inciertas. Cuando enfatizan las emociones y sentimientos. Cuando deben actuar sin un fundamento teórico.
Pragmático	<ul style="list-style-type: none"> Prueban ideas, teorías y técnicas nuevas y comprobar si funcionan en la práctica. Buscan ideas y las ponen en práctica No tienen largas discusiones. Son gente práctica, apogada a la realidad, a la que le gusta tomar decisiones y resolver problemas. Los problemas son un desafío y siempre están buscando una manera mejor de hacer las cosas. 	<ul style="list-style-type: none"> Con actividades que relacionen la teoría y la práctica. Cuando ven a los demás hacer algo. Cuando tienen la posibilidad de poner en práctica inmediatamente lo que han aprendido. 	<ul style="list-style-type: none"> Cuando lo que aprenden no se relaciona en nada con sus necesidades inmediatas. Con aquellas actividades que no tienen una finalidad aparente. Cuando lo que hacen no está relacionado con la "realidad"

En [12], por ejemplo, se describe el uso del cuestionario CHAEA para identificar los estilos de aprendizaje de 794 estudiantes de diferentes cohortes que ingresan a la universidad. Se analizan diferentes resultados de acuerdo al género, y a la elección de alguna de las seis carreras de licenciatura. Los resultados de este trabajo, con tres cohortes diferentes, ofrecen una valiosa información psicopedagógica de la forma en que el estudiante universitario prefiere enfrentar el aprendizaje.

A partir de estos antecedentes, se retoma la idea de utilizar OA para la enseñanza, en este caso de Compuertas lógicas, y también analizar la incidencia de los estilos de aprendizaje en la utilización de estos materiales educativos.

3 OA Compuertas Lógicas

El OA Compuertas lógicas se diseñó y desarrolló en el marco de una Convocatoria de Creación de Objetos de Aprendizaje de la Facultad de Informática de la Universidad Nacional de La Plata.

Este OA cumple con la definición adoptada en este trabajo y posee los componentes referidos en dicha definición. Su estructura se observa en la Fig. 1.



Fig. 1. Secuencia de contenidos del OA Compuertas Lógicas.

Posee contenido teórico sobre compuertas lógicas elementales y compuertas lógicas derivadas así como también actividades relacionadas.

Se caracteriza por la interactividad de los contenidos que presenta. Dentro de cada explicación teórica, cita ejemplos tanto de las compuertas lógicas y las tablas de verdad correspondientes con las que se puede interactuar para analizar su funcionamiento intercambiando el valor de las entradas por ceros o unos. Al mismo tiempo, se puede observar que esa interacción también se refleja, en forma automática, en la tabla de verdad. La Fig. 2 muestra un ejemplo de esta interactividad.

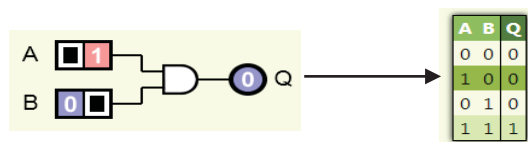


Fig. 2. Ejemplo de la compuerta Lógica AND donde la interacción de las entradas se marca en la fila correspondiente de la tabla.

Tanto las compuertas elementales como las compuertas derivadas poseen actividades de asociación de símbolos y asociación de operadores. En estas actividades, se deben asociar los símbolos con la compuerta lógica o se deben asociar los símbolos con el operador lógico correspondiente. Son preguntas de opción múltiple que tienen un *feedback* instantáneo para que puedan corregir su respuesta. Como actividad integradora poseen un editor de circuitos, donde tienen la posibilidad de armar circuitos combinatoriales con todas las compuertas vistas. A medida que arman el circuito, se arma de forma automática la tabla de verdad y la expresión lógica correspondiente a ese circuito. En la Fig. 3 se muestra un ejemplo de un circuito armado en el editor.

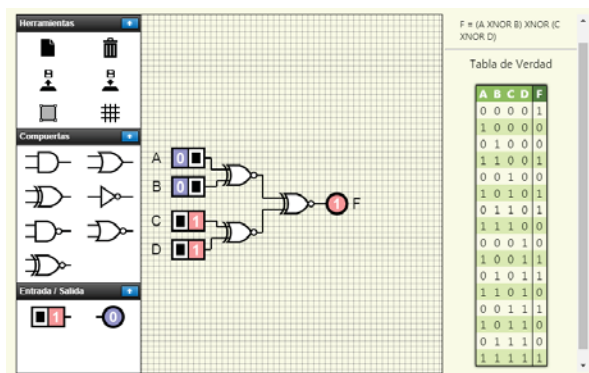


Fig. 3. Ejemplo del editor de circuitos con el generador de tablas automática.

Para finalizar con el OA disponen de una evaluación integradora con preguntas de opción múltiple, que tiene una calificación numérica sobre todos los conceptos tratados en el OA. Finalmente, el OA presenta un cierre con un resumen de los temas abordados y sugerencias para continuar con otros relacionados. El OA “Compuertas Lógicas” se encuentra disponible tanto para verlo *online* como para descargarlo en formato SCORM² y ser reutilizado en algún entorno tecnológico compatible.

4 Experiencia de utilización del OA Compuertas Lógicas en un curso inicial de la carrera informática

La experiencia se llevó a cabo con 22 alumnos de Programación 1. Se trata de una asignatura de la carrera Ingeniería en Computación de la Facultad de Informática de la UNLP. La experiencia se organizó teniendo en cuenta el trabajo previo con otros materiales educativos digitales realizados por autores de este trabajo [13].

Antes de comenzar se realizó una evaluación *pretest* con preguntas referidas a tablas de verdad, compuertas lógicas y símbolos que representan las operaciones lógicas, esto permitió tener un diagnóstico previo de la comprensión de los estudiantes acerca del tema y las principales dificultades.

Se diseñó una guía para entregar a los estudiantes con el fin de que realicen todas las actividades pautadas allí, trabajando con el OA. La sesión tuvo una duración aproximada de 45 minutos. Estuvo bajo la coordinación de los docentes de la cátedra de Programación 1, algunos de ellos son autores de este trabajo.

En la Fig. 4 se listan los pasos realizados por todos los estudiantes en esta experiencia.

² SCORM (del inglés *Sharable Content Object Reference Model*) es un conjunto de estándares y especificaciones que permite crear objetos pedagógicos estructurados

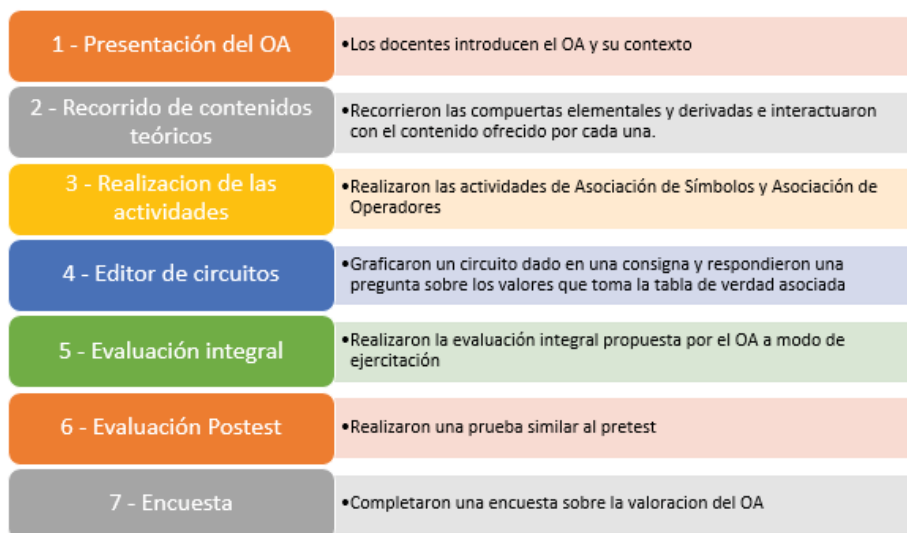


Fig. 4. Ejemplo del editor de circuitos con el generador de tablas automática.

Se analizaron resultados desde dos aspectos: impacto en el aprendizaje mediante la técnica de *pretest* y *postest*, y una valoración de la satisfacción en el uso de este material educativo mediante una encuesta implementada con los formularios de *Google Form*. Los resultados fueron presentados en [14] y permiten vislumbrar una mejora en el aprendizaje, incrementando la comprensión de los temas para aquellos estudiantes que tienen poco conocimiento previo. Particularmente, la valoración obtenida conduce a la conclusión de que los estudiantes tienen un alto nivel de satisfacción con la experiencia realizada utilizando este OA y la interactividad que brinda en el desarrollo de sus actividades. Sin embargo, algunos estudiantes han valorado en una escala intermedia las actividades, y sólo unos pocos como baja. Se busca entonces indagar la incidencia de los estilos de aprendizaje en todas las valoraciones.

Es por ello que en este trabajo se pone el foco en una indagación de estilos de aprendizaje usando el cuestionario CHAEA y la correlación de estos resultados con la satisfacción por las actividades presentadas en el OA. Al principio de la cursada de Programación 1, se administró el cuestionario CHAEA a 73 estudiantes de la asignatura. De estos 73 estudiantes, 22 participaron de la experiencia con el OA Compuertas Lógicas. La misma se realizó con un cuestionario impreso, y luego se procesaron los resultados y se realizó una correlación con las preguntas sobre la valoración de la satisfacción que se describen en la próxima sección.

5 Resultados de la experiencia correlacionando estilos de aprendizaje con la valoración del uso del OA

Una sección de la encuesta que los estudiantes completan luego de realizar la experiencia incluye 3 preguntas sobre las actividades donde tienen que asignarle un

puntaje en una escala de 1 a 5, siendo 1 el menor valor y 5 el mayor valor. Las preguntas son: “¿Cuánto crees que te ayudaron en tu aprendizaje las actividades de Asociación de símbolos?”, “¿Cuánto crees que te ayudaron las actividades de Asociación de operadores?” y “¿Cuánto crees que te ayudó en tu aprendizaje la actividad de Editor de compuertas?”. Las valoraciones obtenidas se observan en el gráfico de la Fig. 5.

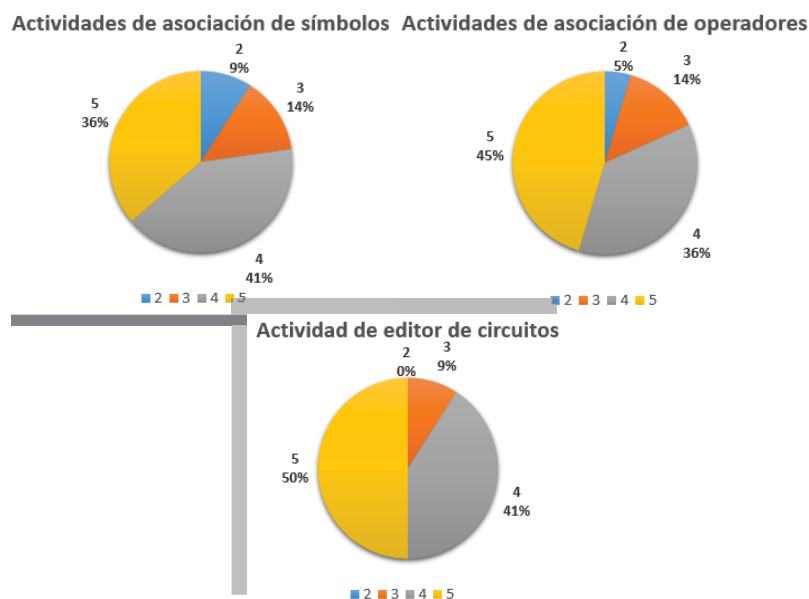


Fig. 5. Valoraciones brindadas por los participantes de la experiencia respecto a la ayuda que brindan las actividades propuestas

A su vez se procesaron los resultados de los 73 cuestionarios CHAEA que respondieron los estudiantes. Los resultados obtenidos indican que los estudiantes involucrados tienen una preferencia por el estilo de aprendizaje teórico (con una valoración de alta), siguiendo por una preferencia por el estilo pragmático (con una valoración de moderada a alta). Estos resultados pueden verse en la Tabla 3.

Tabla 3. Resultados porcentuales obtenidos del cuestionario CHAEA.

	Activo	Reflexivo	Teórico	Pragmático
Muy Baja	12.33 %	10.96%	0.00%	6.85%
Baja	30.14%	26.03%	13.70%	15.07%
Moderada	36.99%	46.58%	23.29%	38.36%
Alta	17.81%	6.85%	38.36%	27.40%
Muy Alta	2.74%	9.59%	24.66%	12.33%

Particularmente los resultados obtenidos respecto al estilo de aprendizaje de los 22 alumnos que participaron en la experiencia del OA, que se observa en la Fig. 5, muestra que:

- La mayoría de los estudiantes ha dado una valoración alta a todas las actividades variando entre 4 y 5. Las actividades del OA parecen entonces ser propicias para el estilo teórico y pragmático que presenta el promedio de estos estudiantes. En las dos primeras actividades de asociación, deben recuperar conocimientos teóricos para poder responder. En la tercera actividad, vinculada al editor de circuitos, se les pide no sólo recuperar conceptos teóricos, si no ponerlos en práctica a través de la creación de un circuito determinado, y analizar su comportamiento para luego contestar una pregunta. Como se presenta en la Tabla 2, los estudiantes con preferencia por el estilo teórico, gustan de indagar, revisar modelos y conceptos que les ayuden a entender, y además, la preferencia por el estilo pragmático, que también caracteriza a este grupo de estudiantes, indica que gustan de actividades que relacionan la teoría con la práctica, y de aplicar lo aprendido. Esto puede explicar la alta valoración por las actividades del OA en cuestión.
- La minoría, que ha indicado una escala promedio de 3 en la satisfacción por las actividades presentadas en el OA, presentan una preferencia de baja a muy baja por el estilo reflexivo. Cabe recordar que la preferencia promedio del grupo por el estilo reflexivo iba de baja a moderada. En los estudiantes que menos valoraron las actividades, se encontró que la mayoría tenía una baja preferencia por el estilo reflexivo. Si bien las dos actividades de asociación, no se vinculan con el estilo reflexivo, la editor sí ya que implicaba la observación de lo ocurrido al cambiar los valores de entrada en el circuito, previo a contestar la pregunta. Esto podría tener alguna incidencia en la valoración de esta minoría de estudiantes.

6 Conclusiones y Trabajos Futuros

En este trabajo se ha abordado un análisis de la posible incidencia de los estilos de aprendizaje de los estudiantes en la realización de las actividades propuesta en un OA. El trabajo no sólo aporta resultados preliminares para el caso en particular, sino que también, permite arrojar luz en relación a las estrategias puestas en juego, tales como la selección de un cuestionario particular para indagar estilos de aprendizaje, y la relación con la satisfacción de los estudiantes en la realización de actividades educativas. Si bien la muestra es pequeña el estudio ha permitido encontrar una correlación positiva en términos de los estilos de preferencia de la mayoría de los estudiantes (teórico y pragmático) y el tipo de actividades planteados en el OA. Al mismo tiempo, existe una pequeña minoría que ha valorado en una menor escala a las actividades del OA y podrían vincularse con una baja a muy baja preferencia por el estilo de aprendizaje reflexivo. Se hace necesario profundizar en estos resultados con una mayor cantidad de estudiantes. A la vez se posibilita la consideración de la creación de actividades en los próximo OA a desarrollar que busquen atender a los estilos de aprendizaje de preferencia de los estudiantes.

Agradecimientos

Este trabajo se vincula con el Proyecto PIIDUZ_16_047 de Innovación Docente 2016-17 de la Universidad de Zaragoza, en el que participan algunos de los autores de este trabajo, y se utilizan aquí algunas de las estrategias previamente adoptadas en este proyecto.

Referencias

1. Sanz, C., Barranquero, F., Moralejo, L. (2016) CROA: a learning object design and creation methodology to bridge the gap between educators and reusable educational material creation, *edulearn16 Proceedings*, pp. 4583-4592.
2. Astudillo, G., Sanz, C. & Santacruz Valencia, L. (2016). Estrategias de diseño y ensamblaje de Objetos de Aprendizaje. Tesis de Magister en Tecnología Informática Aplicada en Educación. Disponible en: <http://sedici.unlp.edu.ar/handle/10915/53442>
3. Sanz, C. (2015) Los objetos de aprendizaje, un debate abierto y necesario. *Bit Byte*. 1, pp. 33-35.
4. Metodología CROA disponible en: <http://croa.info.unlp.edu.ar/> consultado el: 10/07/2017
5. Felder, R. M. (2002). Learning and teaching styles in engineering education [*Engr. Education*, 78(7), pp. 674-681 (1988).
6. Sewall, T. J. (2004). "The Measurement of Learning Style: A Critique of Four Assessment Tools." ERIC ED267247, 1986.
7. Kolb, D. A. (1985). Acerca de la administración de empresas y el proceso de aprendizaje. En D. Kolb, I. Rubin y J. M. McIntyre, *Psicología de las organizaciones: problemas contemporáneos*. pp. 18-34. México: Prentice Hall.
8. Martín del Buey, F., Camarero Suárez, F. (2001). Diferencias de género en los procesos de aprendizaje en universitarios. *Psicothema*, 13(4), pp. 598-604.
9. Franzoni, A. Assar, S., Defude, B., Rojas, J. (2008). "Student Learning Styles Adaptation Method Based on Teaching Strategies and Electronic Media," Eighth IEEE International Conference on Advanced Learning Technologies, Santander, Cantabria, 2008, pp. 778-782. doi: 10.1109/ICALT.2008.149
10. Luna-Ramírez, W. A, Jaimez-González, C. R. (2014). "Supporting structured programming courses through a set of learning objects," International Conference on Information Society (i-Society 2014), London, 2014, pp. 122-126. doi: 10.1109/i-Society.2014.7009024
11. Alonso, C. M. (2008). "Estilos de Aprendizaje, presente y futuro". *Revista Estilos de Aprendizaje*. 1(1). pp. 4-15. Consultado el 19/07/2017
Disponible en:
http://www.uned.es/revistaestilosdeaprendizaje/numero_1/lsr_1_abril_2008.pdf
12. Lugo, J., Rodríguez Hernández, G. y Luna, E. (2012). El cuestionario de estilos de aprendizaje CHAEA y la escala de estrategias de aprendizaje ACRA como herramienta potencial para la tutoría académica. *Revista de Estilos de Aprendizaje*, 10 (10), pp. 148-171.
13. Salazar Mesía N., Sanz C., Gorga G. (2016) Experiencia de enseñanza de Programación con Realidad Aumentada. *Jornadas de la Enseñanza Universitaria de la Informática JENUI 2016*. Almería, España. Julio de 2016. Publicado en las Actas de las Jornadas ISBN: 978-84-16642-30-4 - pp. 213 a 220.
14. Estrebou, C., Salazar Mesía, N., Sanz, C. (2017). Objeto de Aprendizaje para la enseñanza de Computas Lógicas. Experiencia y evaluación. XII Congreso de Tecnología en Educación y Educación en Tecnología. Universidad de La Matanza.

LEGV8, Raspberry Pi 3 y una vieja fórmula

Pablo Ferreyra¹, Agustín Laprovitta^{1,3}, Delfina Velez Ibarra^{1,2}, Gonzalo Vodanovic^{1,4},
Nicolás Wolovick¹

¹ Universidad Nacional de Córdoba, Córdoba, Argentina.

² CONICET, Córdoba, Argentina.

³ Universidad Católica de Córdoba, Córdoba, Argentina

⁴ Universidad Tecnológica Nacional, Facultad Regional Villa María, Villa María, Argentina

Resumen En este trabajo se describe una experiencia sobre la mejora del aprendizaje del Lenguaje Ensamblador (LE) llevada a cabo en la materia Organización del Computador de la Carrera de Licenciatura en Ciencias de la Computación de la Facultad de Matemática, Astronomía, Física y Computación (FAMAF) de la Universidad Nacional de Córdoba (UNC). Los principales objetivos son los de mostrar al alumno la utilidad de adquirir una comprensión de las características y potencialidades de la programación a nivel de LE, y mostrar la factibilidad de incentivar al alumno en el uso del LE usando las capas bajas de abstracción y en contacto directo con el hardware. El trabajo surge a partir del cambio de arquitectura de conjunto de instrucciones (ISA) estudiadas en la materia. Se pasa de la arquitectura MIPS de 32 bits a la LEGv8 de 64 bits, propuesta en el último libro de Patterson-Hennessy. Se aprovecha el hecho que LEGv8 es un subconjunto propio de ARMv8 y la disponibilidad de plataformas Raspberry Pi 3 que incorporan esta ISA. Se diseña un laboratorio con un enunciado mínimo: elaborar una demo gráfica utilizando un *framebuffer* arrancando la máquina sin sistema operativo y ejecutando el código sin ninguna capa de abstracción. En base a los códigos entregados por los alumnos, se realizaron análisis y se generaron estadísticas. Los resultados obtenidos muestran que los objetivos fueron cumplidos, observando un aumento en la participación de la actividad respecto a otras instancias de evaluación.

Keywords: Organización del Computador, lenguaje ensamblador, LEGv8, ARMv8, Raspberry Pi 3

1. Introducción

La materia Organización del Computador (O.C.) en la Licenciatura en Ciencias de la Computación de la FaMAF-UNC nace junto a la carrera en 1993. Durante el periodo 1993-1999 aproximadamente, su dictado está a cargo casi exclusivamente del ingeniero Carlos Alberto Marqués, jefe del Grupo de Desarrollo Electrónico e Instrumental (GDEI) de la FaMAF. Luego, en el periodo 2000-2014 aproximadamente, se desarrolla un trabajo conjunto entre el GDEI, los egresados de la Licenciatura en Computación (que empiezan a colaborar en el dictado de las materias) y otros profesionales, complementando el fuerte perfil electrónico inicial de O.C. con un enfoque más general, basado

en la arquitectura básica de un microprocesador y el estudio de su ISA. En la actualidad y desde hace 5 años aproximadamente, estamos en el periodo “moderno” de la cátedra, conformada principalmente por docentes-investigadores del GDEI y otros de la Sección de Computación. En este periodo se empieza a utilizar el libro de texto de Patterson-Hennessy [1] en base a la arquitectura MIPS [2]. Se realizan laboratorios utilizando el simulador MARS [3] que tiene un entorno de programación rico, además de incorporar dispositivos de I/O interesantes como *framebuffers*, generadores de sonido MIDI, entradas por teclado y acceso al sistema de archivos. Aunque la experiencia es positiva, se detecta que el uso de un emulador oculta la experiencia concreta de los alumnos con el hardware desnudo (*bare-metal*) y los aleja de los problemas reales de la programación de dispositivos embebidos, ya que disponen de un entorno de desarrollo que permite un ciclo de edición-ensamblado-ejecución muy rápido y eso sesga a la programación por prueba-error. También se nota que al ser tan clásico el libro y el simulador, la mayoría de las dudas están resueltas en plataformas de intercambio de conocimientos como Stack Overflow y esto reduce el compromiso de pensar los problemas e intentar resolverlos sin demasiada ayuda externa. En el periodo 2013 a 2015 se incorporan proyectos opcionales de programación de Raspberry Pi 1 en ARMv6 para sensar y actuar dispositivos digitales a través de GPIO, siguiendo las ideas de [4]. Aunque la experiencia es buena, los alumnos tienen que aprender ARMv6 por su cuenta a partir de los conocimientos de MIPS. A mediados del 2015 y gracias al Proyecto de Mejoramiento de la Enseñanza en Carreras de Informática (PROMINF) [5] de la Secretaría de Políticas Universitarias del Ministerio de Educación de la Nación, se adquieren 12 Raspberry Pi 3, que permiten tener una base de plataformas de desarrollo y producir un cambio de ISA en la Cátedra. Los elementos estaban dados por dos componentes fuertes, un libro clásico y de altísima calidad que fue portado de MIPS a LEGv8 [6] y una plataforma de desarrollo para los proyectos con capacidad de ejecutar ensamblador ARMv8 de manera *bare-metal*. Respecto a los recursos humanos disponibles, en la cátedra hay un Doctor especializado en Sistemas Embebidos, un Doctor en Ingeniería especializado en Sistemas Tolerantes a Fallas, un Doctor en Computación especializado en Computación de Alto Desempeño y dos Ingenieros Electrónicos también especialistas en Sistemas Embebidos, es decir, un equipo que es capaz de tomar el desafío de cambiar la ISA y que este cambio no impacte de manera negativa en los alumnos. Actualmente en esta materia se inscriben alrededor de 110 alumnos, pero el conjunto de trabajo real es de aproximadamente 80 alumnos.

2. Plataforma de Desarrollo

La idea de la computadora Raspberry Pi surge en el año 2006 luego que Eben Upton y sus colegas del Cambridge Computer Laboratory en Inglaterra detectan una caída en la comprensión del funcionamiento interno de la computadora entre los ingresantes a la carrera de Ciencias de la Computación. Los ingresantes de 1990 eran hobbistas de la programación con conocimientos profundos de organización y arquitectura de computadoras, mientras que los ingresantes del nuevo milenio apenas tenían alguna idea de diseño web [7].

Surge entonces la idea de re-crear el ecosistema de las microcomputadoras de los 80s que permitían una comprensión total del hardware/software y fomentaban la programación y modificación de la computadora, en vez del simple uso del software. El retroceso de la computación como ciencia en las escuelas estaba causando problemas en el ámbito universitario y había que hacer un esfuerzo para revertir la situación. En 2006 se propone una computadora de 25USD hecha con un microcontrolador ATmega644 con 512 KiB de RAM y salida de video compuesto de 320x200x8 generado por software [8], algo similar a proyectos del 2002 como la SX Game System que estaban basadas en microcontroladores PIC [9]. Aunque Eben Upton deja la academia y empieza a trabajar en Broadcom, la idea no desaparece, sino que muta con la aparición de systems-on-a-chip (SoC) cada vez más poderosos y de precio accesible que permitirían el uso de computadoras a chicas y chicos que inicialmente no solo quieren programar. En 2012 aparece Raspberry Pi 1 alrededor del BCM2835 (ARMv6Z), un SoC para *media players* (Roku por ejemplo) y, aunque de manera lenta, el progreso continúa con Raspberry Pi 2 en 2015, utilizando un BCM2836 (ARMv7A) y con Raspberry Pi 3 en 2016, con un BCM2837 (ARMv8-A) que es ya un procesador de 64 bits en contrapartida a los anteriores que eran de 32 bits. La potencia de cómputo de una Raspberry Pi 3 se equipara a la de un Pentium 4, pero casi dos órdenes de magnitud menos en disipación de calor. Todos los modelos mantuvieron el procesador de video, un Broadcom VideoCore IV que aunque incorpora una unidad de procesamiento SIMD cuádruple (QPU) y es capaz de procesar números de punto flotante de simple precisión a 28 GFLOPS, tiene un *framebuffer* de video fácilmente configurable. La plataforma ARM es tan popular que la mayoría de las distribuciones de GNU/Linux incorporan *cross-compilers*, *cross-assemblers* y *cross-debuggers* tanto para 32 como para 64 bits, `arm-none-eabi-*` y `aarch64-linux-gnu-*` respectivamente. Existen además pruebas de concepto que muestran cómo generar un sistema de archivos FAT32 sobre una tarjeta *microSD* para que la Raspberry Pi 3 cargue en la memoria `0x0000` un programa en código de máquina de ARMv8 que utiliza el *framebuffer* [10] y que se compila en la ubicua plataforma Linux/x86_64.

3. Arquitectura del Conjunto de Instrucciones LEGv8/ARMv8

La materia se divide en dos partes, la primera está a cargo de un Ingeniero Electrónico y produce un recorrido *bottom-up* que parte de compuertas básicas hasta el diseño de circuitos secuenciales, combinacionales y mapeos de memoria. La segunda parte está a cargo de un Científico de la Computación y tiene un recorrido *top-down*, que parte de la ISA de LEGv8, pasando por el código de máquina y finalizando en la implementación de la CPU que ejecuta ese conjunto de instrucciones a partir de todos los módulos vistos en la primera parte. El objetivo fundamental es que los estudiantes comprendan cómo funciona una CPU.

LEGv8 es una ISA de tipo RISC *three-operand/load-store* con un ancho fijo de instrucción de 32 bits, que es un subconjunto propio de ARMv8 con algunas modificaciones para facilitar la generación manual de código de máquina, como por ejemplo los operandos inmediatos de las instrucciones aritméticas y lógicas. Se evita también utilizar el *barrel shifter* para el tercer argumento. Hay más simplificaciones. No existen

los operadores de carga y almacenamiento en memoria escalados ni los pre y post incrementos en estos mismos operadores. La ISA es extremadamente sencilla, a tal punto que en una cartilla A5 se resume todo el conjunto de instrucciones, su semántica y la codificación en bits. Esto se conoce como la *green-sheet* y es la hoja de ruta que los estudiantes deben tener durante toda la segunda mitad de la materia. LEGv8 tiene 31 registros de 64 bits de propósito general (X0 a X30), incorpora un registro de solo lectura que siempre está a cero (XZR), un registro de acceso limitado apuntando a la instrucción que se está ejecutando (PC) y un registro de estado (PSR) que solo se puede leer para controlar el flujo de ejecución a través de saltos condicionales y solo es escrito por las instrucciones aritmético lógicas con el postfijo S de *set flags*. El set de instrucciones tiene la típica división de los procesadores RISC: aritmético-lógicas, memoria y control. Las instrucciones aritmético-lógicas son las usuales tienen dos modificadores independientes de las que surgen cuatro variedades, estos modificadores son si la instrucción tiene operando inmediato y si la instrucción modifica las banderas de estado. También hay instrucciones de desplazamiento. Las instrucciones de transferencia de datos entre el banco de registros y la memoria RAM pueden ser de 64, 32, 16 y 8 bits que se corresponden con double-words, words, half-words y bytes con un registro base y un desplazamiento inmediato. Para modificar el flujo de control hay instrucciones de salto relativo de diferentes rangos con y sin condiciones. También incorpora una instrucción de cambio de PC con almacenamiento previo de la dirección de retorno en X30 o *link register* para implementación de subrutinas con un nivel de anidamiento.

Instrucción	Semántica
ADD X0, X1, X2	$R[X0] = R[X1] + R[X2]$
SUBI X0, X1, #42	$R[X0] = R[X1] - 42$
LSL X0, X1, #16	$R[X0] = 2^{16}R[X1]$
LDUR X0, [X1, #8]	$R[X0] = M[R[X1] + 8]$
STUR X0, [X1, #0]	$M[R[X1] + 0] = R[X0]$
CBNZ X0, 1024	if (X0 ≠ 0) PC = PC + 4096 else PC = PC + 4
B -8192	PC = PC - 32768
BL 6	$R[X30] = PC + 4; PC = PC + 24$

Aunque no existen ensambladores y desensambladores de LEGv8, si se tiene el suficiente cuidado en las diferencias, se puede utilizar las herramientas de desarrollo del GNU Compiler Collection: *as*, *objdump*, *strip*, *gdb*, *gcc*. Un ejemplo típico para ensamblar y desensamblar:

```
$ echo "ADD X0, X1, X2" | aarch64-linux-gnu-as -- && aarch64-linux-gnu-strip ./a.out
→ out && aarch64-linux-gnu-objdump -d ./a.out

./a.out:      file format elf64-littleaarch64

Disassembly of section .text:
0000000000000000 <.text>:
0: 8b020020      add    x0, x1, x2
```

La ISA es moderna, ya que fue anunciada en Octubre de 2011 y aún está penetrando en el mercado de los teléfonos inteligentes, *media players* y computadoras de bajo consumo. La perspectiva es que todos los dispositivos ARM de 32 bits, salvo las arquitecturas de procesadores orientadas a microcontroladores (Familia Cortex-M), se pasarán a esta ISA, por lo tanto el estudio de ARMv8 es una apuesta a futuro.

4. Proyecto

Con el fin que los alumnos puedan visualizar en pantalla los resultados de la ejecución del código ensamblador, se diseña un laboratorio donde se aprovecha el puerto HDMI de la Raspberry Pi 3 y, de esta forma, se prescinde de la incorporación de electrónica de interfaz. Para esto se hace uso del *framebuffer* a bajo nivel y se altera la secuencia de inicio del sistema, cargando el código generado en reemplazo del kernel del sistema operativo. La plataforma Raspberry Pi 3 soporta *framebuffer* para el manejo gráfico en su Video Core (VC). Para esto, hay que realizar una inicialización del VC por medio de un servicio implementado para comunicar el CPU con el VC llamado *mailbox*. Un *mailbox* es uno o varios registros ubicados en direcciones específicas del mapa de memoria (en zona de periféricos) cuyo contenido es enviado a los registros correspondientes de control/estado de algún periférico del sistema. Luego del proceso de inicialización del VC via *mailbox*, los registros de control y estado del VC pueden ser consultados, como por ejemplo la dirección de memoria donde se ubica el inicio del *framebuffer*.

Vale recordar que la plataforma Raspberry Pi 3 está basada en un SoC BCM2837, que incluye un CPU ARM Cortex A53 de cuatro núcleos. Para determinar cuál de los cuatro núcleos ejecutará nuestro código utilizamos el identificador del núcleo mediante el uso de un registro de función especial (MPIDR, *Multiprocessor Affinity Register*) al principio del programa y ponemos a ejecutar un lazo infinito todos los núcleos salvo el 0, donde correrá la demo. Dado que excede los contenidos de la materia, se entrega a los alumnos un código que realiza las tareas de inicialización explicadas anteriormente, junto con un programa de ejemplo a modo de plantilla. Este código inicializa el *framebuffer* en 512 píxeles por 512 píxeles y formato de color RGB de 16 bits. En el formato RGB (ver Fig. 1), el color negro se representa con el valor `0x0000`, el blanco con `0xFFFF`, el rojo con `0xF800`, el verde con `0x07E0`, y el azul con `0x001F`.



Figura 1: Formato de color RGB de 16 bits

En base a la configuración dada, el *framebuffer* queda organizado en palabras de 16 bits (2 bytes), cuyos valores establecen el color que tomará cada píxel de la pantalla. La palabra contenida en la primer posición del *framebuffer* determina el color del primer píxel (indicado con 0 en la Fig. 2), ubicado en el extremo superior izquierdo de la pantalla, incrementando su numeración hacia la derecha en eje X (columnas) hasta llegar al píxel 511. De esta forma, el píxel 512 representa el primer píxel de la segunda línea (fila 1).

Fila\Columna	0	1	2	509	510	511
0	0	1	2	509	510	511
1	512	513	514	1021	1022	1023
2
...
509
510
511

Figura 2: Distribución de píxeles en una pantalla con resolución de 512×512

Debido a que la palabra que contiene el color de cada píxel es de 16 bits, la dirección de memoria que contiene el estado del píxel N se calcula como se indica en Eq. 1, siendo N el número de píxel, mientras que X e Y indican el número de columna y fila en pantalla respectivamente.

$$\text{colorpixel}_N = M[\text{framebuffer}_{\text{base}} + 2N] \quad \text{donde } N = 512Y + X \quad (1)$$

El proyecto que se entrega a modo de ejemplo (Fig. 3), además de la configuración del *framebuffer*, contiene el código necesario para escribir la palabra `0xF800` en todas las posiciones de memoria correspondientes al *framebuffer*. Al compilar dicho código y renombrarlo como `kernel8.img` en la Raspberry, se genera una pantalla de 512×512 píxeles de color rojo:

```
.globl app
app:
    //----- EJEMPLO -----
    // Registro X0 contiene la direccion base del framebuffer

    mov w10, 0xF800    // X10[15..0] = 0xF800 -> color ROJO
    mov x2, 512        // X2 = 512 -> Tamano en Y (filas)
loop1:
    mov x1,512        // X1 = 512 -> Tamano en X (columnas)
loop0:
    sturh w10, [x0]    // M[X0] = X10[15..0] -> Se establece el color del pixel
    add x0, x0, 2      // X0 = X0 + 2 -> Se posiciona X0 en el proximo pixel
    sub x1, x1, 1      // X1 = X1 - 1 -> Decrementa el contador de columna
    cbnz x1, loop0     // Si no es el fin de la columna, salta a loop0
    sub x2,x2,1        // X2 = X2 - 1 -> Decrementa el contador de filas
    cbnz x2,loop1     // Si no es la ultima fila, salta a loop 1
    //-----
    // Loop Infinito
InfLoop:
    b InfLoop
```

Figura 3: Código de ejemplo para poner la pantalla en rojo

La primer consigna que se da a los alumnos, organizados en grupos de tres, es la de escribir un programa en assembler ARMv8 sobre el código de ejemplo dado, que pinte la pantalla completa dividida en tres secciones horizontales de igual tamaño, una roja, una verde y una azul. Con esta primer consigna simple se espera que los

alumnos comprendan y asimilen la organización y funcionamiento del *framebuffer*. El proyecto culmina con el desarrollo de un programa en ensamblador ARMv8 donde se espera que se genere un efecto gráfico llamativo con movimiento, con el espíritu de la *demoscene* de los 80s, una forma de expresión cultural a través de medios digitales [11]. Para evaluar, se introduce un mínimo de restricciones: la duración de la secuencia debe ser de al menos 10 segundos, y el patrón generado debe ser repetitivo y no aleatorio. Además, se considera deseable la utilización de la gama de colores completa (paleta de $2^{16} = 65536$ colores) y se valora la relación entre el efecto logrado y el tamaño del código generado.

5. Resultados

El laboratorio tuvo un total de 10 horas de clases asistidas por los docentes, donde los estudiantes trabajaban y consultaban. El periodo de trabajo fue de un mes. Las 12 plataformas Raspberry Pi 3 se pusieron a disposición, para que los estudiantes, además de las 10 horas disponibles de laboratorio, pudieran continuar trabajando en sus casas. La demanda fue grande y se tuvieron que prestar por turnos, siguiendo un esquema *round-robin* con un tiempo límite por grupo. Finalmente se presentaron 21 trabajos de grupos de hasta tres personas, donde aprobaron 57 estudiantes.

Se observó gran participación y motivación de los estudiantes. Los principales indicadores fueron la demanda de las 12 Raspberry Pi 3, el entusiasmo que se veía en las 10 horas de trabajo asistido por los docentes y sobre todo que la cantidad de estudiantes que presentaron el laboratorio (57) superó a la cantidad de estudiantes que rindieron el segundo parcial (49), que evaluaba habilidades similares respecto al lenguaje ensamblador. También se pudo ver la motivación en que al menos cinco estudiantes adquirieron la plataforma para desarrollar más cómodamente la demo y luego seguir trabajando en ella.

El análisis de los trabajos entregados muestran casos destacados, como la implementación de el Conway's Game of Life, la impresión de dígitos 7-segmentos para mostrar mensajes, impresión de dígitos 7-segmentos con el número binario del color del mensaje, barras de colores siguiendo la secuencia de Fibonacci y complejos dibujos bitmap como el logo de Raspberry Pi superpuesta a la demo gráfica. Algunas de ellas se pueden ver en la Fig. 4. Claramente estos casos superaron la expectativa y mostraron que el laboratorio capturó la metáfora *low-floor/high-ceiling/wide-walls* [12].

Se realizó un análisis cuantitativo del uso de la ISA a fin de buscar evidencia del desempeño del proceso de enseñanza/aprendizaje. En la Fig. 5 se muestra un desglose de las instrucciones de la ISA utilizadas, ordenadas por frecuencia de aparición, tanto para el primer ejercicio de la bandera tricolor como para el segundo de la demo. Se observan varios fenómenos, el primero es que en el segundo ejercicio se cubrió más la ISA, otro importante es que en este ejercicio aparecen picos en instrucciones específicas: `add`, `bl`, `movk` y `lsl`. Todas menos `bl` tienen explicación en la necesidad de mayor cantidad de cálculos para lograr efectos más vistosos.

Este fenómeno también se puede ver en la Fig. 7 donde mostramos el uso de las instrucciones por tipo, dividido en aritmético/lógicas, memoria y flujo de control.



Figura 4: Cuatro demos en acción

En la Fig. 6 graficamos la diversidad de uso de la ISA, mostrando el porcentaje utilizado en los proyectos de instrucciones que no pertenecen a LEGv8 y si a ARMv8. Aunque no fueron dadas en clases, las instrucciones de ARMv8 [13, 14] fueron investigadas y utilizadas por los estudiantes.

Finalmente notamos que el objetivo de la consigna “el efecto logrado vs. la cantidad de líneas de código”, se logró a través del uso de modularización de código mediante el uso del par de instrucciones `bl` y `br X30` para saltar a una subrutina y retornar (Fig. 8). Resulta importante que la modularización surgió de manera natural como mecanismo para disminuir la cantidad de líneas de código y manejar la complejidad de la codificación.

6. Conclusiones

Para verificar el cumplimiento de los objetivos planteados inicialmente, todos los alumnos que realizaron el proyecto debieron pasar por una presentación y defensa oral del mismo. En dicha presentación, se realizaron preguntas para determinar si habían adquirido una comprensión de las principales características y potencialidades de la programación a nivel de LE y su utilidad. También se realizaron observaciones para determinar el grado de motivación y predisposición de los alumnos para usar de las capas bajas de abstracción y en contacto directo con el hardware. En base a éstas herramientas y el análisis de los códigos en LE desarrollados por los alumnos, estamos en condicio-

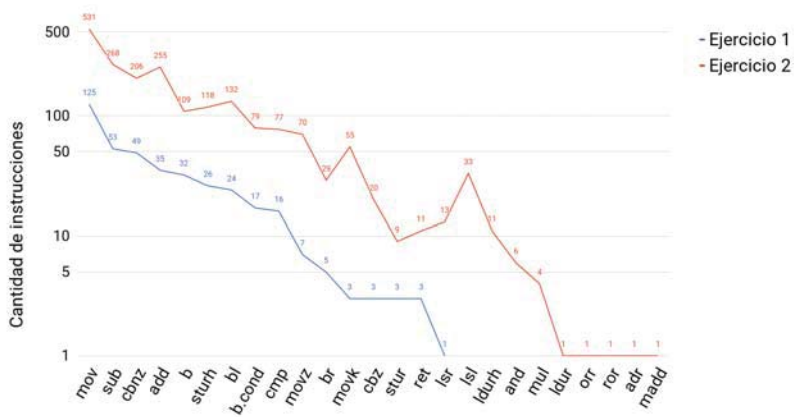


Figura 5: Histograma de cantidad de instrucciones ordenadas por frecuencia

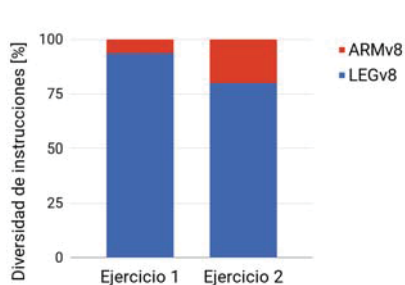


Figura 6: Diversidad en el uso de la ISA

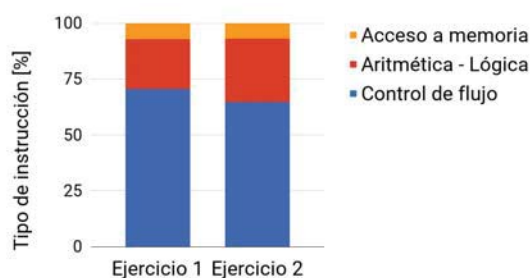


Figura 7: Tipos de instrucciones de la ISA

nes de afirmar que los objetivos mencionados se cumplieron. Otros aspectos a resaltar incluyen, pero no se limitan, a los siguientes:

- Se utilizó una nueva ISA con mucha perspectiva de aplicación a los sistemas embebidos actuales y futuros.
- Se trabajó por medio de experiencias prácticas y concretas, planteando un desafío interesante para los estudiantes.
- La plataforma de desarrollo también resulta altamente motivacional para los alumnos, porque para la mayoría de ellos resultó ser el primer contacto y experiencia con hardware corriendo directamente sobre el metal y a la vista, es decir, con un sistema embebido.

En perspectiva y en resumen, se alcanzó el mismo objetivo que planteaba Upton en su propuesta original de 2006 [8]: hacer una demostración gráfica en LE directamente sobre la CPU y sin el agregado de ninguna capa de software. Tal vez la razón del éxito es

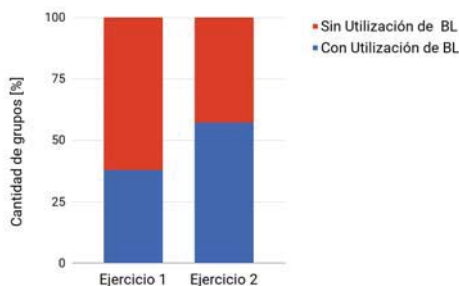


Figura 8: Uso de *branch-and-link* en los grupos

que la vieja receta sigue vigente, aún cuando usamos una ISA moderna como ARMv8 y una plataforma atractiva y poderosa como la Raspberry Pi 3.

7. Agradecimientos

A Martín Marcucci de UCC por la ayuda para compilar `gdb` para ARMv8 en X86_64. A Felipe Manzano que nos ayudó a desnudar los binarios de toda información de debugging. Al PROMINF [5] por los fondos para comprar las Raspberry Pi 3.

Referencias

1. Patterson, D.A., Hennessy, J.L.: Computer Organization and Design, Fifth Edition: The Hardware/Software Interface. 5th edn. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (2013)
2. Przybylski, S.A., Gross, T.R., Hennessy, J.L., Jouppi, N.P., Rowen, C.: Organization and VLSI implementation of MIPS. Technical report, Stanford, CA, USA (1984)
3. Missouri State University: MARS (MIPS assembler and runtime simulator). <http://courses.missouristate.edu/KenVollmar/mars/>
4. Clements, A.: ARMs for the poor: Selecting a processor for teaching computer architecture. (2010)
5. Ministerio de Educación de la Nación: Proyecto de mejoramiento de la enseñanza en carreras de informática (PROMINF). <http://www.educacion.gob.ar/secretaria-de-politicas-universitarias/seccion/168/mejora-de-la-ensenanza> (2016)
6. Patterson, D.A., Hennessy, J.L.: Computer Organization and Design, ARM Edition: The Hardware/Software Interface. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (2016)
7. Raspberry Pi Foundation: About an ARM GNU/Linux box for \$25. Take a byte! <http://web.archive.org/web/20120220051631/https://www.raspberrypi.org/about/> (2012)
8. Raspberry Pi Foundation: Raspberry Pi – 2006 edition. <https://www.raspberrypi.org/blog/raspberry-pi-2006-edition/> (2017)
9. Gunée, R.: SX Game System. <http://www.rickard.gunee.com/projects/video/sx/gamesys.php> (2002)

10. Lemon, P.: Raspberry Pi Bare Metal Assembly Programming. <https://github.com/PeterLemon/RaspberryPi/tree/master/HelloWorld/CPU> (2017)
11. Polgár, T.: Freax, The History Of The Computer Demoscene. CSW-Verlag (2005)
12. Resnick, M., Silverman, B.: Some reflections on designing construction kits for kids. In Eisenberg, M., Eisenberg, A., eds.: IDC, ACM (2005) 117–122
13. ARM Limited: ARMv8 Instruction Set Overview (2011)
14. Franchin, M.: ARMv8-A A64 ISA Overview, 64-bit Android on ARM. Campus London (2015)

Milleannizando el CMMI: El “Mito” de los Estilos de Aprendizaje Hecho Realidad

Marcelo Campo

Analía Amandi

Núcleo de Inteligencia Comportamental Empresarial (NICE)

Facultad de Ciencias Exactas-UNICEN

CONICET

Tandil, Bs. As., Argentina

mcampo@exa.unicen.edu.ar amandi@exa.unicen.edu.ar

Abstract. Mucho se ha escrito sobre los *Millennials*. Ganaron particularmente notoriedad en el campo de la tecnología. Para muchas empresas innovadoras, fueron una oportunidad importante debido a su "forma de ver la vida". Para otras, más convencionales y más grandes, se convirtieron en una profunda preocupación. Fue necesario Darwinismo para hacer frente a este repentino cambio de comportamiento en el personal técnico. El Darwinismo también es necesario en el campo de la educación en Ingeniería de Software. El interés por lo efímero que los caracteriza juega un papel importante para mantener su atención a lo estructurado (y no demasiado divertido) pero esencial como el CMMI. En este artículo contamos una experiencia en un curso que se dictó recurriendo a un método no muy convencional para *mejorar el proceso de aprendizaje* del CMMI por parte de los estudiantes. Un enfoque calificable como *irreverente* de mejora basado en Lean, aplicando un método de desarrollo "Agile" para estructurar la enseñanza y el concepto "mítico" de *estilo de aprendizaje* para personalizar las interacciones *aprendiz-entrenador*. Son resultados, tal vez obvios pero para nosotros sorprendentes que deseamos compartir con la comunidad de Ingeniería de Software.

1 La conclusión

Este artículo debe comenzar con una advertencia. Una advertencia que, normalmente, sería una nota al pie de página. Pero, si fuera una nota al pie de página, sería inmediatamente interpretado como una forma de disculpa científicista. ¿Por qué? Este artículo no está escrito con los modos científicos "políticamente correctos" usuales, sino que está escrito para enfatizar la mejor manera que encontramos para interactuar con los estudiantes, que en nuestro caso son parte de la generación denominada *millennial*¹, utilizando un enfoque ágil de acuerdo su "estilo de aprendizaje".

Normalmente, deberíamos introducir aquí la noción de *estilos de aprendizaje* [1]. Sin embargo, según el tratado "democrático" escrito por Paul Kirschner, *Stop propagating the learning styles myth. Computers & Education 106: 166-171 (2017)*, nos genera temores. Los editores sabiamente publicaron este artículo (que no cita a Felder-Silverman) carente de fundamentos serios junto a otros científicamente serios que demuestran el valor del concepto, como una clara muestra de lo que debe ser la ciencia, diseminación de las opiniones, mismo sean necias, pues también existen los que la citan en redes sociales anunciando la verdad de algo que los *preocupa*, sin haber leído los otros artículos... Sea como sea, Kirschner se ganó una cita y logra la diseminación del *mito*.²

Así, decidimos arriesgarnos y la cuestión detrás de este artículo no es sobre evolución, sino sobre crudo darwinismo. La cuestión no pasa por cómo la educación universitaria tradicional debe adaptarse a

¹ Si el revisor es un *millennial*, por favor no se ofenda.

² Disculpas, no podemos evitar reírnos de ciertas cosas y particularmente de los que las *diseminan* en las redes sociales.



Fig. 1. Diapositivas de la Retrospectiva del curso presentadas por los alumnos

los cambios. La verdadera pregunta es: ¿cómo pueden sobrevivir los cursos universitarios de Ingeniería de Software hoy en día?

Más importante aún, ¿cómo pueden presentarse temas como CMMI a estudiantes que no tienen experiencia en los múltiples problemas que surgen, más allá de la programación y, peor aún, ni siquiera aparecen en *What's Up!*? Este tipo de aspectos, que simplemente no existen en su *mundo*, es uno de los desafíos que enfrentamos. Un desafío más, pero demasiado importante. Un desafío que (hoy parece recurrentemente mediático) resulta obvio decir que requiere creatividad, innovación, o incluso un comportamiento *payasezco* (mejor no entrar en el mundo Rasti o Lego de algunos consultores Scrum)². Eso puede ser cierto, pero lo que realmente requiere es *empatía*, no *simpatía* y este es el desafío que debemos *animarnos* a enfrentar. Y la *irreverencia* nos ha demostrado ser un camino más que posible sino una necesidad.

En este contexto, tuvimos que dictar un curso de Ingeniería de Software basado en CMMI varios años después de no dictar clases regularmente. ¿Qué hacer ahora?, pensamos, teniendo en cuenta que la efemeridad es una característica común entre *millennials*. CMMI es un manual muy importante de buenas prácticas para el desarrollo de software de calidad. Pero, es... un manual, particularmente aburrido (quizás la razón por lo que es tan importante y valioso).

La agilidad está de moda y esto se correlaciona bien con la idea de efemeridad de los requisitos de software, por lo que decidimos explorar la utilización de técnicas ágiles. La primera lección aprendida: pronto comprendimos que hoy debemos referirnos a Agile como una subclase de Scrum.

Para hacer honor a nuestro nuevo estilo *millennial* de enseñanza, comenzamos con las conclusiones. Este es un punto de partida interesante que nos permite resaltar los resultados muy interesantes que motivaron escribir este artículo y tal vez una forma en que el lector puede encontrar una motivación para seguir leyendo toda la historia y las valiosas impresiones de lo que aprendimos durante el proceso.

La Figura 1 presenta tres diapositivas de las conclusiones de los alumnos sobre el curso presentadas por los Project Owners de cada equipo Scrum, como se organizó el curso, en la Retrospectiva. Como es bien conocido las Retrospectivas son una ceremonia esencial en Scrum. Según la visión CMMI, serían una catarsis grupal indocumentada sobre errores cometidos durante el proyecto, y si están documentados, ciertamente los documentos no serán utilizados en el próximo proyecto. Esto no es necesariamente cierto; este artículo es un ejemplo. Vamos a analizar lo que las diapositivas realmente dicen.

Diapositiva 1: Habilidades Adquiridas

Los estudiantes resumieron de manera asombrosa las habilidades que adquirieron durante el curso. Esta simple diapositiva representa, de hecho, la clave de nuestro éxito al reflejar las intenciones exactas que perseguimos. La pirámide que muestran es suficiente. En primer lugar, el desarrollo de software es esencialmente una actividad de grupo y concluyeron que la interacción "forzada" intra e inter grupo que imponíamos era una novedad de la que aprendieron mucho más de lo esperado (de hecho dijeron que eso fue lo primero que causó repulsión y temores, pero finalmente fue la experiencia más enriquecedora). En segundo lugar, entendieron que se encontraron con la necesidad de organizar y coordinar su tiempo. Al principio eso molestó, pero finalmente terminó en una nueva experiencia bienvenida de equilibrar el tiempo y el compartir la responsabilidad. Tercero y lo más importante, se dieron cuenta de que aprender sobre los detalles de las prácticas específicas de CMMI no era tan importante. Cuando comprendieron que la estructura se trata de un *framework*, el manual se convirtió

en un manual que conocen y por lo tanto, serían capaces de encontrar las respuestas que necesitan justo en el momento preciso que la necesitan. ¡Música para oídos *millennial!*. Sólo necesitan sus *cells* para conectarse con el SEI y *voilà*.

Estábamos "humildemente" contentos de escuchar estas conclusiones. Sin embargo, el ambiente agradable, relajado y respetuoso del aula fue el logro más importante. Fuimos capaces de romper su "sensación de seguridad de comunicación de teléfono celular", y esto fue observado en las dos siguientes diapositivas.

Diapositiva 2: Sobre la materia...

1. Hacer llevadero algo que es tan tedioso como el manual del CMMI
2. Adaptación al cambio permanente
3. Se logró mejorar la comunicación entre compañeros
4. Flexibilidad en el orden de las exposiciones

Como se puede notar, observan los aspectos que fueron "pesados" al principio como las experiencias más positivas. Los puntos 2.2 y 2.3 son objetivos clave porque implican la tendencia más marcada en el comportamiento *millennial*. En primer lugar, por lo general prefieren las actividades bien estructuradas con el fin de organizar su valioso tiempo libre. La mayoría de ellos consideran estudiar como parte del "trabajo", por lo que se sienten incómodos con los cambios en las horas de clase o actividades que deben realizarse. De hecho, mostraron un odio profundo por los cambios continuos al principio, pero cuando descubrieron que el profesor le gustaba su capacidad de soportar esa "tortura", empezaron a sentirse cómodos, y al final reconocieron que era una experiencia muy positiva. Reconocieron que fue una ayuda a entender que el trabajo en el desarrollo de software implica una presión permanente en el tiempo y que necesitan trabajar sobre su pánico al fracaso (según Sinek³ característico en los *millennials*, como hemos confirmado).

El punto 3.3 resultó confuso, por qué era una preocupación real para ellos? La sola idea de trabajar en un grupo con compañeros que no solían *chatear* era extraña y horrible. El prejuicio era muy marcado en la mayoría de ellos. Confiesan que se sorprendieron al final que esa cosa tan rara les resultó fructífera, remarcando que era sorprendente descubrir que podían negociar entre los líderes de grupo el orden de las presentaciones, lo que conduce a la satisfacción expresada en el punto 2.4.

Diapositiva 3: Puntos a mejorar...

1. Organizar un calendario al inicio del curso teniendo en cuenta los feriados.
2. Organizar mejor los tiempos de las presentaciones y establecer un acuerdo previo para evitar los inconvenientes en la asistencia.
3. Mantener una comunicación con la cátedra de una universidad como UNICEN a través de un canal informal como Facebook, que no era una de nuestros medios favoritos.

Esta diapositiva refleja un momento divertido. Se avergonzaron porque preguntamos si sólo tres cosas eran cuestionables y particularmente tan contradictorias consigo mismas. Presionamos para más de una manera amable, después intentamos con provocación... nada. Pero el punto más revelador es 3.3 que es realmente sorprendente, pero vamos a ver por qué. Que un *millennial* diga que Facebook no es una forma adecuada de comunicación es raro y cuando insistimos en preguntar por qué, tratando de entender ese mito caído, uno de ellos tomando coraje dijo: Creemos que se están metiendo en nuestra vida privada... Preguntamos por qué, con preocupación verdadera... Porque Facebook es el canal que estamos acostumbrados a usar para interactuar con nuestros amigos, no para trabajar ... y ese sentimiento fue compartido por la mayoría de los estudiantes. Pero es un grupo cerrado, no uno público, les explicamos, todavía tratando de entender. ¿Cuándo tuvieron la oportunidad de hacer preguntas a un profesor en cualquier momento del día sin esperar hasta la próxima clase? --Expresión pensativa-- Hummm... no importa, sentimos eso, pero ahora estamos acostumbrándonos, así que no te preocupes - fue la respuesta fresca. Muy *millennial*, por cierto.

Algunos días después de esta charla, el Banco Interamericano de Desarrollo (BID) publicó un

³ <https://www.facebook.com/NTN24/videos/1467440253275020/>

estudio en el que se informa que la mayoría de los *millennials*, y particularmente los argentinos, en América Latina no les gusta usar los canales de comunicación electrónicos para el trabajo, a pesar de que pasan por lo menos cuatro horas diarias en *chats*. Eso parece ser cierto, pero la experiencia real nos muestra que no es tan así, dependiendo de la forma que Facebook se utilice para comunicarse con ellos.

En este punto llegamos a una encrucijada. O, el lector debería estar lo suficientemente intrigado para averiguar con impaciencia cómo logramos estos interesantes resultados, así que deberíamos seguir explicando lo que es un estilo de aprendizaje. O, si no, queremos agradecer a los revisores anónimos que mucho nos ayudaron con sus comentarios educados que sugieren que debemos pensar seriamente en comenzar a planear un cómodo retiro.

2 Estilos de Aprendizaje y Educación Ágil

Si todavía están aquí, vamos a presentar brevemente lo que es un *estilo de aprendizaje*. Básicamente, un estilo de aprendizaje es la forma en que un estudiante recopila y procesa la información que debe aprender. Como se basa en la prueba de personalidad de Myers-Briggs, la noción de estilo de aprendizaje se refiere a las *preferencias* individuales no al comportamiento *absoluto*. El modelo de estilos de aprendizaje Felder-Silverman (ILS) caracteriza a cada alumno según cuatro dimensiones:

$$\text{ILS} = \{\text{Percepción, Comprensión, Procesamiento, Entrada}\}$$

que comprende 16 estilos de aprendizaje. La Tabla 1 muestra las dimensiones de los estilos de aprendizaje en este modelo: los aprendices Activos aprenden *aprehendiendo las cosas* y trabajando con otros mientras que los Reflexivos aprenden *pensando las cosas* y trabajando solos. Los estudiantes Sensitivos aprenden a través de *ejemplos concretos* y tienden a ser *prácticos*, mientras que los aprendices Intuitivos prefieren *material abstracto* como teorías y sus significados y tienden a ser más *innovadores* que los Sensitivos. Los aprendices Visuales (Vis) recuerdan mejor lo que han visto mientras que los aprendices Verbales (Ver) consiguen más de palabras, sin importar si son habladas o escritas. Los aprendices Secuenciales (Seq) aprenden en pasos lineales y prefieren seguir caminos lineales escalonados mientras que los aprendices Globales (G) aprenden en grandes saltos y se caracterizan como holísticos.

De acuerdo con esta descripción, el cuestionario ILS propone una lista de elementos efectivos para identificar el estilo de cada alumno. El índice de preferencia resultante para cada dimensión se expresa por un número impar que varía [-11, +11] puesto que se plantean 11 preguntas para cada una de las cuatro dimensiones. Para cada pregunta están disponibles 2 respuestas posibles, la de valor +1 y la de valor -1. Por ejemplo, al responder a una pregunta con una preferencia activa, la puntuación del alumno se incrementa en +1 mientras que para la preferencia reflexiva la puntuación se reduce en 1 (es decir, se añade -1).

Más formalmente, se obtiene el grado de preferencia por cada pareja de dimensiones acopladas en el caso en el que todas las preguntas han sido contestadas, como

$$i(\text{DIM}, q) = \sum_{q \in \text{DIM}^+} q_i - \sum_{q \in \text{DIM}^-} q_i$$

donde cada i , $\text{DIM} = \{A/R, S/I, V/V, S/G\}$ indica todas las dimensiones de *pair wise coupled styles* cuyo conjunto de índices es dado por $I = \{i_{A/R}, i_{S/I}, i_{V/V}, i_{S/G}\}$, el conjunto de todas las preguntas para cada dimensión es dado por

$$Q_{\text{DIM}} = \{q_1^{\text{DIM}}, \dots, q_{11}^{\text{DIM}}\}$$

cada q_i indica la contribución dada por la pregunta nro i en las once relacionadas a cada DIM para la detección de preferencias, y

$$q_i \in \text{DIM}^+ \text{ if } q_i = +1, q_i \in \text{DIM}^- \text{ if } q_i = -1.$$

Los datos son expresados en una escala binaria; las salidas están expresadas en una escala

ordinal en la cual sólo la preferencia dominante (esto es. la diferencia de las dos sumas de $i(DIM, q)$ es explícitamente dada⁴.

Dimensión	Estilos de Aprendizaje	Estilos de Enseñanza
Percepción	Sensitivo	Concreto
	Intuitivo	Abstracto
Comprensión	Secuencial	Secuencial
	Global	Global
Procesamiento	Activo	Activo
	Reflexivo	Pasivo
Entrada	Visual	Visual
	Verbal	Verbal

Tabla 1. Dimensiones de estilos de aprendizaje de Felder

Este modelo también conduce a un modelo analógico para estilos de enseñanza, conocido como la hipótesis de mallado, que establece que si los estudiantes reciben una instrucción adecuada para sus preferencias de aprendizaje (por ejemplo, para un "aprendiz activo", haciendo hincapié en las tareas prácticas) Los estudiantes podrían mejorar su experiencia de aprendizaje [2].

Hemos trabajado en estos temas míticos durante los últimos diez años con el enfoque en los diversos aspectos que implica el aprendizaje de Ciencias de la Computación, con el objetivo de la detección automática de estilos de aprendizaje para la auto-adaptación de las plataformas de enseñanza [3, 4]. Basándonos en los resultados obtenidos, fuimos más allá para explorar las actividades de desarrollo de software más complejas estudiando su utilidad en el comportamiento de los estudiantes en los procesos de software ágil [5] [6], incluyendo la interacción a distancia utilizando los entornos de realidad virtual. Elegimos los métodos ágiles, no sólo por su influencia en el movimiento de desarrollo de software, sino por su énfasis en el comportamiento del equipo y las habilidades personales.

En esta investigación exploratoria, encontramos ideas muy interesantes que demuestran la importancia de los Estilos de Aprendizaje en la práctica real y que fueron muy inspiradores para el trabajo descrito en este artículo, como se muestra en las siguientes secciones.

3 Una visión reflexiva de la Enseñanza de Procesos de Software

Era una noche agradable de verano en Tandil. Marcelo estaba cómodamente acostado en su hamaca reflexionando cosas intrascendentes, como de costumbre. De repente, un tema apareció en su mente: "Pensé cuando estaba estudiando mi primer año en la carrera de Ciencias de la Computación a principios de los 80, cuando gastaba mi tiempo con TIMBA (un lenguaje diseñado para enseñar programación estructurada, un esfuerzo voluntarista en aquellos tiempos difíciles en Argentina), Alan Kay estaba ideando el Dynabook y Smalltalk empezó a cambiar el mundo de la informática... que juventud desperdiciada" dijo retóricamente. Analía, profundamente concentrada como siempre leyendo un artículo sobre Estilos de Aprendizaje pensaba: "bien, paciencia, ya empieza...". Y así fue. ¿Por qué en los cursos de Ingeniería de Software enseñando Procesos de Desarrollo de Software no se utiliza un Proceso de Enseñanza y nunca se utiliza la información disponible que tenemos hoy? Analía, sin levantar la vista del artículo, simuló estar sorprendida por una reflexión tan profunda, respondió- ¿Aha? Mirá vos...?

El diseño del curso había comenzado. ¿Cómo enseñar a *millennials* un tema aburrido como CMMI? ¿Y por qué no utilizar un proceso ágil para estructurar el curso en el que los estudiantes tengan como proyecto enseñar CMMI? Eso les permitiría poner en práctica un Proceso Ágil formal para aprender CMMI y aprender también cómo organizar una presentación profesional, crucial para

⁴ El concepto de *preferencia dominante* fue introducido por Jung, y es la base del Myer-Briggs Personality Type Indicator. Una explicación más detallada de este aspecto está fuera del alcance de este artículo, y puede ser encontrada fácilmente en Internet.

desarrollar una carrera exitosa y perder el miedo a hablar en público, y esencialmente derrotar el pánico inherente al fracaso que por lo general, *millennials* tienen. Analía se inquietó un poco, Marcelo había comenzado otra de sus aparentemente *alucinaciones recurrentes*. Se va a poner pesado, pensó... y tenía razón.

Scrum es la palabra clave obligatoria que debe utilizarse. La representación continua de CMMI encaja perfectamente bien para organizar el estudio como cuatro *Sprints*, cada categoría conforma el *Backlog* natural del proyecto y las áreas de proceso dentro de cada categoría de forma natural son el *Sprint Backlog* compuesto por los objetivos específicos como Épicas divididas en *User Stories* dadas por las Prácticas Específicas. Eso está bien, pero son *un número* de estudiantes, ¿cómo organizar eso? Eso es mejor aún porque el proyecto puede ser diseñado como un Multi-Scrum en el que cada Equipo Scrum tendrá asignada para enseñar un Área de Proceso específica, y luego deben presentar los avances en Scrum Semanales y la necesidad de preparar la presentación semanal con Diapositivas requeriría algún

tipo de Scrum diario. La catarata que terminó en menos de un minuto se muestra en la Fig. 2.



Fig. 2. Organización del Curso como un Multi-Scrum usando Representación Continua de CMMI

Para lograr satisfacer este objetivo la práctica CMMI, "Obtener el compromiso de los interesados" es esencial. Pero, ¿cómo podemos conseguir esto con *millennials*? - Aquí, los estilos de aprendizaje juegan un papel central en ayudar en la personalización del curso, dijo Analía -

pareciendo ahora interesada. Esto se convierte en la parte más divertida y desafiante de la experiencia, pero se necesita una planificación cuidadosa.

3.1 Planeando el curso

Como toda nueva experiencia, el curso necesitó ser cuidadosamente planeado porque requería la misma atención que cualquier proyecto de desarrollo de software necesita para Agile, y más.

Existen principalmente dos riesgos:

- ¿Cómo cambiar la visión de los estudiantes acerca de lo que solía ser un curso normal?
- ¿Cómo debemos comunicarnos con ellos manteniendo la efemeridad como un factor potenciador?

El primero no es menor. Una actitud irreverente podría ser malinterpretada como un signo de falta de seriedad. Como después supimos a través de los propios estudiantes, ellos piensan que algunos profesores se comportan como si fueran dispositivos utilizados para transportar su cerebro brillante (tal como dice Sir. Ken Robinson en TEDx [8]), y de hecho, a veces pensamos lo mismo pero no es *educado* decirlo de esta manera.

El segundo es estratégico y podría obligarnos a trabajar más tiempo de lo habitual. Por lo tanto, no tenemos otra alternativa que tratar de evolucionar desde la forma clásica académica a ser más... *millennials*.

3.1.1 Facebook y un gancho esotérico para la conformación de grupos usando estilos de aprendizaje

Como primer paso, primero obtuvimos los correos electrónicos de los asistentes potenciales, creamos una página de Facebook para el curso y los invitamos a unirse. A los dos días casi todos se unieron a la página, encontrando un mensaje de bienvenida que les pedía que informaran su signo en el zodiaco Maya para conformar los grupos de acuerdo a las preferencias descritas en ese zodiaco disponible en Internet. La explicación era que estábamos llevando a cabo una investigación sobre la verdad científica de tales creencias de compatibilidad grupal... En este punto, el lector tal vez piensa que estábamos bromeando (o peor, locos), pero no fue así. Para nuestra sorpresa, en veinte minutos tuvimos la lista completa de pequeños animales que se convirtieron en sus avatares.

Luego, les pedimos que llenaran el cuestionario del Test de Felder que nos ayudaría a hacer el

trabajo científico. Durante el día tuvimos las respuestas al cuestionario y diseñamos grupos de 6 personas mezclando diferentes perfiles ILS. Para dar credibilidad, postulamos que los Halcones tenían el perfil de los *Project Owners* y los Zorros el de los *Scrum Master*, mientras que la combinación entre los Jaguares dando agilidad, combinada con la sabiduría de la lenta Tortuga y la visión de los Búhos era una mejor combinación. Y si tienen una Ardilla en el grupo, será ciertamente más dinámico. Creasé o no esto, ellos revelaron una faceta inesperada para nosotros que podría ser explotada.

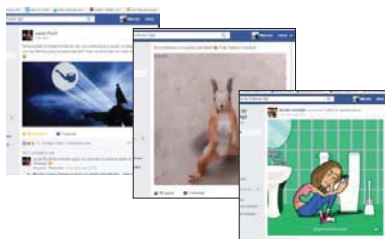


Fig. 3. Algunas publicaciones bromeando sobre el profesor

inmediatamente vistos por casi todos, y los "like" aparecieron lentamente. La confianza fue creciendo hasta que algunos de los más activos e intuitivos comenzaron a responder con bromas sobre sí mismos o sobre el *Old Scrum Master* (Marcelo, que acostumbra a pasearse arrogantemente con su pipa) (Fig. 2)

3.2 La historia "salvaje" de la evolución de los procesos de desarrollo de software

¿Cómo mostrar la importancia de un Proceso y la necesidad de Agile a estudiantes que sólo conocen programas pequeños? Eso debería explicarse de una forma obvia, mejor, debería ser visualmente mostrado en términos de cosas comunes que los estudiantes puedan asociar fácilmente con los conceptos subyacentes del CMMI. Los alumnos clasificaron muy alto en la dimensión Visual y el par Sensitivo-Secuencial, por lo que podríamos explotar la forma en que la mayoría del grupo procesa la información. La imaginación y el darwinismo vinieron al rescate con una apuesta arriesgada: "La historia salvaje de la Evolución de Procesos de Desarrollo de Software", una película que editamos la cual explota las interacciones visuales, habituales en la comunicación de *millennials* de cosas divertidas, usando cualquier información disponible en YouTube.



Fig. 4. Vista darwiniana de la evolución de los procesos ágiles

espíritu no es suficiente para hacer frente a la constante evolución de la tecnología (la caída de King Kong) y que la mejora del proceso es constantemente necesaria y debe ser el foco (Iron Man) que combina todas las cualidades de ingeniería expresadas por, casualmente, CMMI.

En realidad utilizamos el test para diseñar grupos con diferentes combinaciones de perfiles ILS, intentando que al menos cada grupo tuviera dos Intuitivos (supuestamente más abstractos) y dos Activos (supuestamente más participativos). Esto no era posible, porque había pocos intuitivos, pero era posible con activos. En un caso juntamos los pocos reflexivos para estudiar los diferentes comportamientos.

Además, Facebook mostró el potencial de interacción y de mantener su atención. Empezamos a subir algunos videos alegóricos de Les Luthiers haciendo una analogía con comportamientos *millennials* que fueron

La figura 4 muestra una secuencia de *snapshots* de la película que presenta una analogía entre Agile y una secuencia de depredadores (equipos de desarrollo) que cazan presas usando la velocidad (guepardo) como en XP para cazar presas de tamaño medio, pero más grandes que él. Las tácticas de grupo (leones) para explicar elementos esenciales de Scrum de desarrollo en equipos, torres humanas para mostrar problemas de escalabilidad que se resuelven con King Kong (FDD) que muestra que la fuerza, la robustez y el

Sorprendentemente, la presentación fue un completo éxito y obtuvo la atención del grupo a la presentación más formal del tema. Varios factores psicológicos pueden estar involucrados, pero creemos que la correcta explicación de la analogía durante la película, que atrajo su tendencia visual, potenció la comprensión subliminal de la narración verbal de la historia dada por Marcelo, esto es, explotó la parte verbal de su perfil de aprendizaje. De todos modos, la primera experiencia pasó sin problemas y nos animó a profundizar el enfoque.

3.3 La parte más reveladora: los *millennials* se involucran

Después de la aceptación de los videos alegóricos fuimos más allá empujando esta idea sugiriéndoles (no forzándolos) a utilizarla. Los resultados fueron increíbles. Se convirtió en un problema central para ellos porque pensaban que era lo que queríamos que hicieran, pero en realidad para ellos era un juego que pronto se convirtió en una especie de competencia de quien era el más inspirado e inteligente (**¡Error! No se encuentra el origen de la referencia.**).

Es importante prestar atención a este punto. Se obligaron a leer y realmente entender el área de proceso CMMI asignada para encontrar una analogía que no hable de software para poder transmitir el concepto. ¡De una manera indirecta, e inconsciente, conseguimos que pongan atención y aprendan! Esto no es un logro menor. Fue sorprendente el comportamiento de muchos de ellos preguntando cómo hacíamos para encontrar analogías, pensando que debería existir un manual de inspiración o como se llame esta habilidad.

Este efecto nos lleva a centrarnos en los ejemplos que eligen. La mayoría de los ejemplos relativos a la negociación, la planificación, la gestión de riesgos y otros aspectos sociales en las prácticas de CMMI, se referían a situaciones delictivas, como películas de la mafia, traficantes de drogas, robo u otras cosas cuestionables. Sin embargo, el ILS de nuevo da algunas pistas interesantes que después confirmó la prueba MBTI. Los estudiantes muy Introspectivos, con el nivel más alto en las dimensiones Sensitivo y Secuencial de aprendizaje, presentaron ejemplos sublimes procedentes del proceso de cocina en un restaurante. Un grupo nos sorprendió con la película de *Ratatouille* que describe el papel de un buen director de proyecto en una situación estresante. Hubo otras analogías muy interesantes, que las razones tiránicas de espacio no nos permiten describir en detalle. El resultado del segundo sprint es destacable: las presentaciones fueron mucho mejores y enfocadas.

Repasando, encontramos una manera en la que los *millennials* encontraran un canal para disfrutar de sus inherentes tendencias lúdicas en una competencia en la que el aprendizaje de CMMI era "simplemente" un derivado valioso. En este caso, el fin justifica los medios. Por otra parte, la suposición de que sujetos altamente intuitivos-globales tendrían un mejor desempeño no parece ser tan cierto.

3.4 Un "Magical Mystery Tour" *millennial* hacia la comprensión de los Objetivos Genéricos

Este podría considerarse el clímax del curso. Siguiendo la tendencia lúdica, publicamos un anuncio animado divertido con grupos de animales que trabajan juntos para salvarse de los ataques de depredadores (hormigas construyendo una bola enorme para cerrar la boca del oso hormiguero, pingüinos en la parte superior de un bloque de hielo desplazándose en equipo para levantar una punta para que se estrelle una orca, etc.) y preguntamos qué práctica específica de CMMI aplican. Al principio sólo unos pocos respondieron, pero cuando agregamos la emoción de una competencia de que las primeras respuestas correctas serían parte de la calificación de examen, el juego tomó una velocidad increíble.

Notando esto comenzamos a dar pistas como una película de suspenso (de hecho, a Marcelo le encanta jugar ese tipo de chistes) y comenzaron a sugerir Áreas de Proceso que aún no habíamos discutido! Marcelo "olía sangre" y empezó a jugar llevándolos a jugar en razonamiento cuántico preguntando por el Boson de Higgs y mostrándoles la Ecuación del Todo. A esa hora no creíamos la catarata de respuestas *alocadas*, hasta que uno de ellos dio una respuesta que estaba muy cerca del concepto de Framework utilizando la Representación Estructurada de CMMI. Dimos otra pequeña pista y el alumno (con mayor ILS en todas las dimensiones) terminó la discusión diciendo que estábamos hablando de los Objetivos Genéricos de CMMI y las Características Comunes. El hecho era que describía esto como el modelo de Einstein de malla del espacio, en el cual estas metas genéricas cruzan todas las Áreas de Proceso con sus prácticas. ¿Qué más...? La Fig. 5. muestra una serie de instantáneas de la evolución del juego en Facebook.

En este punto, consideramos que el curso había terminado y... en la mitad del tiempo planeado. De esta manera, pudimos avanzar en temas no planeados como el análisis de un tema candente como es la integración de Agile con CMMI y la Definición de un Proceso Agile según los estándares CMMI para el próximo curso, un Taller de Ingeniería de Software en el que tienen que crear una *start-up* simulada llamada *MillennialWare*.

3.5 El factor "small town"

La mayoría de los estudios sobre el comportamiento de los *millennials* se basan en grandes ciudades. Lugares en los que el nivel socioeconómico juega un papel importante para encontrar diferencias. El último estudio del BID enfatiza que la mayoría de los *millennials* prefieren la educación universitaria clásica y no les gusta usar tecnología en el trabajo y esto es más marcado en aquellos de bajos recursos. No estamos de acuerdo con esta afirmación.

La mayoría de nuestros alumnos provienen de pequeñas ciudades cercanas o incluso de pueblos rurales. Se consideran en su mayoría miembros de la llamada clase media. Por lo menos eso expresan cuando se le pregunta acerca de las actividades de los padres. Una cuestión fundamental a comprender es el perfil sociocultural, que en nuestra experiencia es mucho más influyente que el socio-económico. ¿Por qué? Están muy influenciados por las expectativas de los padres de tener un título universitario, casi como un trofeo por sus esfuerzos en la vida. Más del 80% de los estudiantes terminan reconociendo este hecho cuando son empujados para dar una respuesta⁵.

Un ejemplo interesante surgió cuando un estudiante estaba presentando una práctica de CMMI relativa con la negociación con proveedores. Marcelo notó que estaba relativamente seguro con el tema y le preguntó si trabajaba en alguna empresa. Él, con orgullo responde, por supuesto. ¿Dónde? Y con una inspiración notoria dijo "Huge", donde más... Marcelo olía una oportunidad y con una admiración Wauh! preguntó, en qué área?. Dijo entonces Atención al cliente. En un centro de llamadas?. Noo! en desarrollo -- ¿Y quién es el cliente? -- Sony Corp., dijo él con orgullo -- Pero, toda la empresa parece demasiado -- No, en un proyecto (sonriendo con condescendencia). La PlayStation 4, dijo, realmente, muy, muy orgulloso --Muy interesante, estas manejando un proyecto muy grande -- Bueno, no, estoy en el módulo de inicio de sesión (ahora se dio cuenta de que había ido demasiado lejos). Bueno, eso es bueno también. Y quién es tu cliente al final? Dos personas, una americana y una japonesa... En este momento la conversación fue hacia las relaciones humanas y la forma en que interactuó con el personal de Sony; representaba un ejemplo interesante para la clase. Pero el *small town millennial* apareció y dijo, "El único problema es que el americano es demasiado *confianzado*"... Marcelo lo escuchó confundido porque el término suena algo extraño para un *millennial* y pregunta "¿qué querés decir con *confianzado*? ¿Realmente estás viviendo en el mundo real? "El muchacho no entendió, e insistió: Por supuesto, él me interrumpe en cualquier momento por What's Up, me pregunta por qué el proyecto se

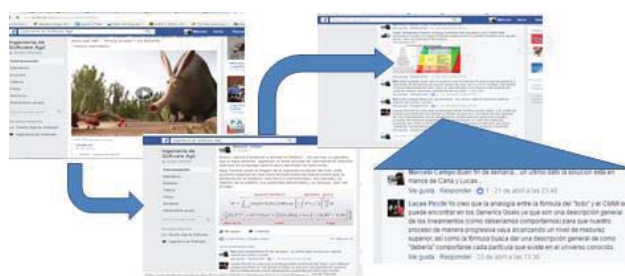


Fig. 5. Secuencia del razonamiento cuántico hacia las Generic Goals en Facebook retrasa!, incluso cuando estoy tomando una cerveza con mis amigos, y eso es demasiado confianzado,

⁵ "Movilidad social ascendente" se utiliza a menudo para referirse a obtener un título universitario. Esto refleja una rara manera de pensar, como una manera de alcanzar una posición aristocrática. Una visión anacrónica e inútil de la "lucha de clases". Preferimos hablar de un camino hacia la prosperidad y satisfacción personal, intelectual y quizás económica.. interpretación personal de .

terminó de manera condescendiente y con sonrisa *canchera*...

Marcelo pasó el resto de la clase tratando que el grupo entienda que estaba trabajando en un ambiente internacional, y lo que implica, etc. Se sintió algo triste; definitivamente será difícil para estos jóvenes adaptarse a la revolución 4.0 como concluye el BID, pero... por razones culturales o idiosincrásicas, no económicas.

4 Observaciones finales

Presentamos brevemente lo que podría ser clasificado como una idea loca. Pero, una muy exitosa. Se pueden extraer muchas conclusiones. Sin embargo, lo más importante para nosotros fue abrir una nueva área de investigación que se puede ampliar mucho estudiando en profundidad la relación con el enfoque MBTI, por ejemplo. Además, que los estudiantes no pasan por sólo otro tema en su carrera; como algunos de ellos dijeron después: "empezamos a ver nuevos horizontes para nuestro desarrollo personal". ¿Qué más?

Si realmente queremos disfrutar mejor de la enseñanza, no necesitamos aprender de las nuevas generaciones. Tenemos que romper con formas estructuradas de parecer "serios". Deberíamos ser un poco más *millennials*, para evitar decir lo que Les Luthiers hace en el sketch de los Jóvenes de Hoy en Día: "¿...Por qué la pasan tan bien ...?".

Al lector interesado, puede contactarse con nosotros por e-mail. Las razones de espacio son tiránicas, pero tienen su sentido y lo respetamos (no de *buena* gana como todos).

5 Agradecimientos

Agradecemos profundamente a los pacientes y entusiasmados estudiantes que nos dieron una de las experiencias de enseñanza más enriquecedoras y agradables que hemos tenido... hasta ahora.

Y un cariñoso recuerdo a Jorge Boria, ese viejo adelantado *clown millennial*... de la Ingeniería de Software argentina.

6 Referencias

- [1] Felder, R.M., Silverman, L.K., 1988. Learning and teaching styles in engineering education. *Eng. Educ.* 78 (7), 674–681.
- [2] Pashler H., McDaniel M., Rohrer D., and Bjork R.. Learning styles concepts and evidence. *Psychological Science in the Public Interest*, 9(3):105–119, 2008. SAGE Publications
- [3] García P., Amandi A., Schiafino S., and Campo M. Evaluating bayesian networks' precision for detecting students' learning styles. *Computers & Education*, 49(3):794–808, 2007. Elsevier
- [4] Feldman, J., Monteserin, A., Amandi, A., 2014. Detecting students' perception style by using games. *Comput. Educ.* 71, 14–22.
- [5] Scott, E., Rodríguez, G., Soria, A., Campo, M., 2014. Are learning styles useful indicators to discover how students use Scrum for the first time? *Comput. Hum. Behav.* 36, 56–64.
- [6] Scott E., Rodríguez G., Soria A. and Campo M., Towards better Scrum learning using learning styles, *Journal of Systems and Software*, ISSN: 0164-1212, vol. 111, pp. 242-253, Elsevier Science, 2016
- [7] Rodríguez G., Soria A., and Campo M.. Virtual scrum: A teaching aid to introduce undergraduate software engineering students to Scrum. *Computer Applications in Engineering Education*, 2013. Wiley Online Library.
- [8] Ken Robinson, <https://youtu.be/iG9CE55wbtY?list=PL70DEC2B0568B5469>
- [9] McFarlane, A. (2015). *Authentic learning for the digital generation*. London: Routledge.
- [10] Eynon, R., & Malmberg, L. (2011). A typology of young people's internet use: Implications for education. *Computers & Education*, 56(3), 585-595.
- [11] Ito, M., Baumer, S., Bittanti, M., Cody, R., Herr-Stephenson, B., Horst, H. A., et al. (2010). *Hanging out, messing around, and geeking out*. Cambridge, Massachusetts: The MIT Press.
- [12] Kennedy, G., Judd, T., Dalgarno, B., & Waycott, J. (2010). Beyond natives and immigrants: Exploring types of net generation students. *Journal of Computer Assisted Learning*, 26(5), 332-343.

O Aluno monitor potencializando o uso das salas de informática nas escolas da rede municipal de ensino de Santa María

Universidade Federal de Santa Maria/UFSM, Rio Grande do Sul, Brasil - Programa de Pós-Graduação em Tecnologias Educacionais em Rede, Mestrado Profissional em Tecnologias Educacionais em Rede

Eunice Pereira Azenha 1, Maritê Medianeira Moro Neocatto 1, Karla Marques da Rocha 2
{1 Mestrandas do PPGTER/UFSM eunice.azenha@gmail.com, mmmneocatto@gmail.com;
2 Prof.^a Dr.^a Orientadora do PPGTER/UFSM karlamarquesdarocho@gmail.com

Resumo: Este artigo pretende apresentar o Projeto Aluno Monitor da Sala de Informática como potencializador do uso desse ambiente nas escolas da rede pública municipal de Santa Maria, como experiência inovadora no que se refere ao protagonismo dos alunos, pois ao mesmo tempo que aprendem, tornam-se incentivadores do uso das tecnologias educacionais como prática cotidiana, incluindo digitalmente professores e alunos da rede municipal de ensino. O referido Projeto foi uma ação idealizada pelo NTEM Santa Maria, em 2015, para solucionar a subutilização das salas de informática das escolas municipais, pela falta de professores de informática educativa designados pela mantenedora. A ação apoia-se na relevância da integração das tecnologias educacionais à prática pedagógica dos professores municipais, como estratégia para transposição de desafios de ensino-aprendizagem e para o manuseio das tecnologias existentes nas escolas, alicerçada em uma abordagem qualitativa e nos procedimentos da pesquisa-ação.

Palavras chave: aluno monitor, tecnologia educacional, aprendizagem colaborativa, NTEM Santa Maria

Abstract. The article intends to present the Monitor Student Project of the Computer Room as a potentiator of the use of this environment in Santa Maria municipal public schools, as a new experience regarding the protagonism of the students, because at the same time that they learn, it becomes Encouraging the use of educational technologies as a daily practice, that would digitally include teachers and students of the municipal school network. This project was an action designed by NTEM Santa Maria in 2015 to solve the underutilization of the computer rooms of municipal schools, due to the lack of educational computer teachers designated as a sponsor. The action is based on the relevance of the educational integration technologies to the pedagogical practice of municipal teachers, as a strategy for transposing teaching-learning challenges and for the management of existing technologies in schools, based on a qualitative approach and research- action.

Keywords: student monitor, educational technologies, collaborative learning, NTEM Santa Maria

1 Introdução

Um dos grandes desafios postos à educação atualmente, é inserir as tecnologias, especialmente as digitais, no ambiente escolar, a fim de que a prática educativa seja revitalizada, permitindo a incorporação de aspectos qualitativos que venham contribuir para o desenvolvimento integral do aluno, destacando que os equipamentos, por si só, não determinam se haverá um bom proveito deles, sua utilização deve estar vinculada a uma proposta bem mais ampla, que abarque toda a complexidade envolvida. Emerge a necessidade do compromisso e de um esforço coletivo que aponta para a importância de redefinição de políticas públicas, que impliquem na inserção efetiva das Tecnologias da Informação e Comunicação (TIC) no processo educativo, com vistas à aprendizagens colaborativas e significativas.

O governo brasileiro possibilitou que a escola pública, e sua cultura, se aproximasse dos avanços que a sociedade vinha desfrutando, ao criar diversas políticas públicas, para o uso das TIC na educação. É o caso do Programa Nacional de Tecnologia Educacional - ProInfo, que distribuiu computadores e periféricos às escolas públicas brasileiras. A chegada dos computadores à escola pública, propiciou a criação de um novo espaço: a Sala de Informática. Esse novo ambiente, criado com o objetivo de pensar novas práticas educativas, novas dinâmicas, novas possibilidades de aprendizagem, entre outros, conduz à ideia de que seja necessário um professor disposto e capacitado para tal, o Professor de Informática Educativa.

Dentro desta perspectiva surgem os Núcleos de Tecnologias Educacionais - NTE (Rede Pública Estadual) e NTEM (Rede Pública Municipal), com a finalidade de formar esses profissionais, bem como proporcionar assessoramento técnico aos ambientes informatizados das escolas. No ano de 2005, foi criado o Núcleo de Tecnologia Educacional Municipal (NTEM), em Santa Maria/RS, como contrapartida à adesão do município ao Programa Nacional de Informática Educativa (PROINFO). Inicialmente, o NTEM formava os professores e a mantenedora designava-os para atuação na função de mediador das TIC, nas salas de informática das escolas. Com o passar do tempo, os mesmos foram sendo realocados para as salas de aula regulares e a salas de informática foram ficando subutilizadas.

Como alternativa a essa situação, o NTEM propôs a formação de alunos, dos anos finais do ensino fundamental, para atuarem como monitores suporte, fazendo a ponte entre a sala de informática e os professores dos anos iniciais, potencializando o uso desse espaço de aprendizagem tão importante. No município, cinquenta (50) escolas de ensino fundamental, receberam os equipamentos oriundos do ProInfo, para as quais o NTEM busca manter assessoramento técnico e pedagógico.

Este artigo pretende apresentar o Projeto Aluno Monitor da Sala de Informática como potencializador do uso desse ambiente nas escolas da rede pública municipal de Santa Maria, como experiência inovadora no que se refere ao protagonismo do aluno, pois ao mesmo tempo em que aprende, torna-se incentivador do uso das tecnologias educacionais como prática cotidiana, incluindo digitalmente professores e alunos da rede municipal de ensino. O referido Projeto foi uma ação idealizada pelo NTEM Santa Maria, em 2015, para solucionar a subutilização das salas de informática das escolas municipais, pela falta de professores de informática educativa designados pela mantenedora. A partir dessa ação e do ingresso das professoras formadoras do NTEM

no Mestrado Profissional em Tecnologias Educacionais em Rede, do Programa de Pós-Graduação em Tecnologias Educacionais em Rede (PPGTER) da Universidade Federal de Santa Maria (UFSM), este, tornou-se um projeto de extensão da UFSM. Partindo desse princípio, a ação apoia-se na relevância da integração das tecnologias educacionais em rede à prática pedagógica dos professores municipais, a partir da implementação, pelo NTEM, do Projeto Aluno Monitor da Sala de Informática como estratégia para transposição de desafios de ensino-aprendizagem e para o manuseio das tecnologias existentes nas escolas, alicerçada em uma abordagem qualitativa e nos procedimentos da pesquisa-ação. Utilizou-se, também, os dados da pesquisa realizada pelo NTEM Santa Maria, por meio da Plataforma Survio, para conhecer o perfil dos poucos Professores de Informática Educativa mantidos pela administração municipal, para a função de mediar do uso das TIC junto aos seus pares, nas salas de informática.

A formação de alunos monitores pretende auxiliar a integração das tecnologias digitais à prática pedagógica dos professores e na introdução da aprendizagem colaborativa na sala de informática da escola, como também, ser capaz de proporcionar competências, habilidades e atitudes para que o aluno monitor consiga transpor os desafios do ensino-aprendizagem de professores e colegas, constituindo-o em colaborador e mediador entre seus pares, potencializando o uso dos recursos tecnológicos existentes na escola. Portanto, o principal objetivo define-se em integrar as tecnologias educacionais em rede à prática pedagógica dos professores, a partir da implementação do Projeto Aluno Monitor da Sala de Informática, visando contribuir para a cultura da aprendizagem colaborativa na escola, para a inclusão digital e para a multiplicação da ação a toda rede municipal de ensino de Santa Maria, RS.

2. O NTEM Santa Maria e a formação de Professores de Informática Educativa

No Brasil, a Informática na Educação teve início, nos anos 80, com o Projeto EDUCOM (Educação com Computador), implementado pelo Ministério de Educação e Cultura (MEC), do governo federal. O objetivo era realizar pesquisa multidisciplinar e capacitar recursos humanos para subsidiar as decisões de informatização da educação pública brasileira. Em 1987, o MEC lançou o Projeto FORMAR, que promovia cursos de especialização, visando preparar professores para atuar na formação de outros professores, habilitando-os para o domínio das tecnologias e as concepções subjacentes ao uso da Informática na Educação, criando propostas de disseminação para o uso do computador em suas instituições de origem.

A partir de 1989, o MEC desenvolveu ações para capacitar professores e técnicos, implantar centros de Informática na Educação - os Núcleos de Tecnologia - adquirir equipamentos computacionais, produzir, adquirir, adaptar e avaliar software educativos. Essas ações permitiram ao governo brasileiro lançar, em 1997, o Programa Nacional de Informática na Educação – PROINFO (reformulado a partir de 2007, chamou-se ProInfo Integrado e ampliou a implantação de ambientes tecnológicos e recursos digitais, disponibilizou conteúdos educacionais, soluções e sistemas de informação (Linux Educacional), além de conexão dos ambientes tecnológicos à rede mundial de computadores). O PROINFO teve a finalidade de disseminar o uso pedagógico das TIC nas escolas públicas de educação básica em

todo o território nacional. Sua operacionalização exigiu a implantação de Núcleos de Tecnologia Educacional (NTE) nos estados e (NTEM) nos municípios, dotados de infraestrutura de hardware e software, de uma equipe de educadores, especialistas em tecnologia, para promover o assessoramento pedagógico e técnico das escolas, como contrapartida dos estados e municípios ao aderir ao Programa. Nesse contexto, o município de Santa Maria aderiu ao PROINFO, em 2005, com a criação do Núcleo de Tecnologia Educacional Municipal – NTEM, através do Decreto Executivo nº 003/2005. O Regimento Interno do NTEM, homologado pelo Conselho Municipal de Educação, em 2007, o referenda como uma estrutura permanente de apoio/assessoramento ao uso das TIC, orientadas exclusivamente para a Educação, atuando no planejamento, suporte técnico, formação e certificação dos professores, equipes pedagógicas e administrativas, quanto ao processo de inserção das TIC na ação docente das escolas públicas da rede municipal de Santa Maria. O objetivo principal do NTEM é contribuir para a inclusão digital dos profissionais da educação, refletindo sobre o impacto das TIC na sociedade, orientando seu uso pedagógico e avaliando sua contribuição efetiva na qualidade educacional do município. De modo que, a finalidade precípua do NTEM é a formação de Professores de Informática Educativa para atuar como multiplicadores nas salas de informática das escolas.

A Informática Educativa ou Informática na Educação abrange o estudo sobre a aprendizagem, a filosofia do conhecimento, o domínio das técnicas e tecnologias digitais e a prática pedagógica. E, se pauta em estimular a formação de Professores de Informática Educativa para atuar no contexto escolar com a multiplicação das TIC, fazendo a ponte entre saberes, de forma interdisciplinar, entre currículos, de maneira a integrar os recursos tecnológicos à prática pedagógica estabelecida no contexto escolar, dando o suporte metodológico à avaliação, produção e utilização das tecnologias nos processos de ensino e aprendizagem dentro da escola.

Na era das TIC, o uso da Informática Educativa nas escolas é imprescindível, mas a utilização da mesma, como recurso pedagógico puro e simples, não causará mudanças. Kenski [8] reforça a importância da conjugação de todos os envolvidos no processo educacional quando afirma que *“para que as novas tecnologias não sejam vistas como apenas mais um modismo, mas com a relevância e o poder educacional transformador que elas possuem, é preciso refletir sobre o processo de ensino de maneira global. Antes de tudo, é necessário que todos estejam conscientes e preparados para assumir novas perspectivas filosóficas as quais contemplem visões inovadoras de ensino e de escola.”*

Neste sentido, destaca-se o importante papel do Professor de Informática Educativa, o qual, entre outras, tem a função de intervir nos processos de ensino-aprendizagem e não apenas deter conhecimentos a serem transferidos aos professores e aos alunos.

A indicação de que seja um professor a assumir esse papel denota a importância deste profissional ser possuidor de uma formação pedagógica com experiência em sala de aula e não ser apenas um técnico em informática. Ele será o mediador, coordenador do ensino-aprendizagem com as tecnologias, fazendo a ponte entre o potencial dos software educativos existentes no laboratório e os conceitos a serem desenvolvidos e contribuirá para a formação de seus pares, apresentando os infinitos recursos tecnológicos necessários para impulsionar as engrenagens do processo, por meio da atualização continuada de si e das TIC, que mudam constantemente.

Essa função, nas escolas públicas municipais de Santa Maria, foi assegurada pelo

Decreto Ejecutivo nº 313/2007, e definida como Assessoramento Pedagógico nos termos da Lei Federal nº 11.301/2006, a qual, para efeitos legais, considera as funções exercidas por titulares dos cargos efetivos, configuradas como apoio pedagógico que contribuam de forma concreta para o projeto pedagógico da escola. Desse modo, foi legitimado o Professor de Informática Educativa nos quadros escolares, para realizar o apoio pedagógico à docência dentro dos laboratórios de informática educativa (LIE) ou salas de informática, como também são chamados e para os quais compete ao NTEM manter a assessoria pedagógica e técnica e formar os professores para multiplicar as TIC entre seus pares na escola.

Em 2017, buscando resgatar o uso das tecnologias nas escolas do município, o NTEM realizou, um levantamento no quadro efetivo de cada escola, para verificar quantos Professores de Informática Educativa estão designados para o LIE das escolas. Ficou constatado, através da efetividade do mês de dezembro de 2016, que *apenas dezoito (18)*, num universo de *cinquenta (50) escolas*, que receberam as salas do ProInfo/MEC, contam com este profissional em seu quadro efetivo. Foi elaborado um questionário, online, na plataforma Survio (<http://www.survio.com/br/>) com questões semiabertas. Após, o link do instrumento de coleta de dados, (<https://my.survio.com/H0G4K8Z4S2C2V2L9D9V1/data/index>) foi encaminhado para o e-mail dos dezoito (18) professores lotados na função de Informática Educativa. O questionário continha catorze (14) questões com os seguintes itens investigativos: o tipo de formação e a carga horária no LIE; a opinião sobre o uso das TIC na escola, as dificuldades encontradas para o desenvolvimento do trabalho com as tecnologias educacionais, como é feito o planejamento de trabalho para o LIE e como o mesmo é realizado, assim como sugestões dos professores para que o NTEM possa melhor contribuir para a inclusão digital das escolas públicas da rede municipal de ensino de Santa Maria/RS. Foram analisadas as *dez (10)* respostas que retornaram e interpretadas a partir do entrecruzamento das mesmas, pois permitiam escolha múltipla, respostas abertas, em ordem de importância, etc. Quanto ao tipo de formação, dos professores que atuam no LIE das escolas, verificou-se que estes têm um *alto nível* de formação, destacando que *seis (06)* professores têm Pós-Graduação, *três (03)* possuem Mestrado, *um (01)* tem Doutorado; no item Qual outra formação na área?, *um (01)* professor disse estar com Doutorado em andamento, *um (01)* professor disse ter formação em Técnico de Informática e *outro (01)* possui Cursos de formação no Sistema Operacional Linux e Montagem e Manutenção de Computadores; porém aqui destaca-se que *nove (09)* professores dizem ser egressos dos Cursos de formação básica do NTEM, ressaltando seu papel de pólo formador. Quanto a carga horária dos professores lotados nas salas de informática, a pesquisa revelou as horas semanais trabalhadas no LIE: *sete (07)* professores com 20 horas, *dois (02)* professores com 10 horas, *um (01)* com 40 horas o que denota dificuldade para a disseminação das TIC, pois são *cinquenta (50)* escolas que possuem salas de informática oriundas do ProInfo, e apenas *dezoito (18)* possuem o professor habilitado lotado, sendo a carga horária um fator limitante das possibilidades de interação com os outros professores, prejudicando o andamento do trabalho, pois é dificultado, ao Professor de Informática Educativa, o planejamento de suas atividades junto aos demais professores da escola. Percebe-se que não há reuniões de planejamento para um trabalho conjunto, como confirmam alguns depoimentos (os professores foram identificados com P seguido de número), conforme o P1 - *“Atualmente, por falta de reuniões pedagógicas e devido*

ao desencontro de horários, os momentos com professores se restringem ao recreio e, quando não há momentos de encontro presencial, conversamos por e-mail ou redes sociais (facebook, whatsapp), o que acaba levando muito tempo no planejamento até partir para execução”, também o depoimento do P2 - “Atualmente não temos oportunidade de planejar juntos, cada professor realiza seu planejamento e trabalhamos em conjunto, pois somente 20 h na escola e professores com carga horária supercarregada inviabiliza um momento para planejamento em conjunto”. Os depoimentos confirmam que não basta ter um professor habilitado, lotado no LIE, se a gestão escolar/mantenedora não entender a relevância do trabalho do mesmo e sua função de multiplicar as TIC entre seus pares. No sentido de demonstrar outras dificuldades para o acesso às TIC no fazer pedagógico das escolas, a questão permitia ordenar por ordem crescente: (1º) equipamentos defasados, (2º) internet lenta, (3º) dificuldade com o sistema operacional, (4º) desinteresse dos professores, (5º) desvio de função do professor do LIE pela gestão da escola, (6º) desconhecimento da real função do professor do LIE por parte da gestão da escola, (7º) falta de formação específica, (8º) não há um professor especificamente habilitado, lotado no LIE da escola, e (9º) no item Outra dificuldade foi relatado o depoimento: “internet lenta em função dos equipamentos defasados e professor lotado no LIE em todos os turnos”.

As respostas da pesquisa corroboram a necessidade da Mantenedora atentar para o upgrade das tecnologias existentes nas escolas e no próprio NTEM, cujos equipamentos estão no prazo de obsolescência programada, como acontece com todas as tecnologias. Essa reestruturação é quesito fundamental para uma boa utilização das TIC, sendo necessário, também, designar Professores de Informática Educativa em todos os turnos de funcionamento das escolas. Segundo, Cysneiros[4] “*ao abordar a introdução da Informática em escolas do Brasil, aponta uma série de problemas para assimilação das novas tecnologias pela escola pública, como a ausência de concepções bem definidas sobre o que é Informática na Educação*”, sendo que as dificuldades envolvem desde aspectos administrativos, financeiros e de recursos humanos, corroboradas pelas respostas acima. Estas dificuldades apontadas em vários estudos são indicativos que demonstram os motivos da ‘não fluência’ tecnológica nas escolas públicas, uma realidade repetida em vários municípios brasileiros. A referida ‘não fluência’ tecnológica da escola pública só faz aumentar a exclusão digital e social o que acaba por reforçar as “*brechas digitais*”, categorizadas por Coll e Monereo,[3] como os ‘*inforicos*’ e os ‘*infopobres*’, um divisor de quem tem e de quem não tem acesso às tecnologias digitais mais atualizadas.

Quanto às sugestões feitas ao NTEM pelos professores investigados, percebe-se que eles têm noção da importância de uma formação específica na área e que reconhecem a formação inicial oferecida pelo NTEM, solicitando que o mesmo continue promovendo o aprofundamento da mesma, como demonstram os depoimentos do P1 - “*Penso que para atuar como coordenador da SI o professor deve ter uma formação mais específica, que no mínimo tenham realizado cursos no NTEM. Com isso, o NTEM oriente as escolas que pretendem designar um professor para atuar na sala de informática, que ele realize anteriormente cursos de formação*”, também o do P2 - “*Continuar com as oficinas que foram ministradas ano passado, para contribuir com a formação dos professores da rede*”. As reflexões realizadas, após o estudo sobre os professores de informática educativa na rede municipal de ensino, apontam para que os Gestores/Mantenedores da Educação Municipal mantenham um olhar atento para

compreender a importância do papel do Professor de Informática Educativa para a inserção das tecnologias digitais no contexto escolar, apoiando o trabalho do NTEM na formação e construção do perfil deste profissional, de forma a proporcionar, através das TIC, a possibilidade de uma aprendizagem contextualizada e significativa, uma vez que, as TIC estão chegando nas escolas e vieram para ficar, sendo preciso, diminuir as “brechas digitais” impostas pela aceleração tecnológica e que atingem mais, diretamente, a escola pública. A falta desses profissionais, no contexto escolar, fez com que o NTEM Santa Maria propusesse formar alunos para potencializar o uso dos ambientes informatizados nas escolas.

3. O Aluno Monitor potencializando o uso das Salas de Informática nas escolas da rede pública municipal de Santa Maria

A ação, idealizada pelo NTEM, para suprir a carência de Professores de Informática Educativa e utilizar as tecnologias existentes nas escolas, de forma mais efetiva, foi a formação de Alunos Monitores. A figura do aluno monitor está imbricada numa metodologia que vem sendo muito estimulada em países europeus, como a Espanha, por exemplo, que usa o termo tutor para este aluno[6]. O aluno tutor/monitor é estimulado a auxiliar os professores na transmissão de conceitos que se aproximam da compreensão dos colegas, como estratégia para a realização de atividades que visem objetivos de aprendizagem onde os resultados esperados serão mais facilmente alcançados se puderem ser experimentados pelos “iguais”, ou seja pela mediação de outro aluno, observados e conduzidos, estrategicamente, pelo professor. Monereo, [10] apontam que *“aprender entre iguais e com iguais remonta à teoria de Vygostsky, [16] e ao conceito, definido por ele, de zona de desenvolvimento proximal (ZDP), tratando-se de um formato interativo marcado por distâncias curtas, (...) onde é possível a construção conjunta do conhecimento na interação que se dá quando é proporcionado interações educativas entre alunos, sob a mediação de professores, alicerçados em métodos de aprendizagem colaborativa como recurso pedagógico”*. O Projeto Aluno Monitor preconiza a autonomia desse aluno, de forma a ser capaz de usar os subsídios da formação recebida, conseguindo fazer a ponte entre a sala de informática e o professor regente dos anos iniciais, tornando-se um facilitador e potencializador do uso deste espaço, contribuindo para o sucesso das atividades pedagógicas ali desenvolvidas e para o desenvolvimento de uma cultura de aprendizagem colaborativa na escola, mediada pelas tecnologias educacionais.

3.1 A formação do Aluno Monitor

O Projeto Aluno Monitor da Sala de Informática foi proposto e estruturado pelo NTEM, em 2015, para solucionar a subutilização das salas de informática das escolas municipais. A partir dessa ação, e do ingresso das professoras formadoras do NTEM, no Mestrado Profissional em Tecnologias Educacionais em Rede, do Programa de Pós-Graduação em Tecnologias Educacionais em Rede (PPGTER), da Universidade Federal de Santa Maria (UFSM), o Projeto em si, ganhou consistência e, por suas possibilidades, tornou-se um projeto de extensão da UFSM. Entendendo-se aqui, extensão universitária como processo de interação, intercâmbio e transformação

mútua entre a Universidade e a comunidade, direcionando as práticas acadêmicas para as questões sociais, políticas, econômicas e ambientais da sociedade. O referido Projeto é divulgado para todas as escolas no início do ano letivo, sendo que cada instituição tem a liberdade de aderir ou não ao mesmo. As instituições participantes selecionam até cinco (05) alunos, dos anos finais do ensino fundamental, e um (01) professor, que acompanha os alunos durante a formação, responsabilizando-se pelo grupo de monitores dentro da escola, pela organização dos horários de estudo e trabalho do grupo na sala de informática e pelo compartilhamento das informações.

O desenvolvimento da proposta se dá por meio de oficinas continuadas, em três etapas: a *primeira etapa* começa com um encontro, *in loco*, nas escolas que aderiram ao projeto e serve para que, alunos e professor selecionados, conheçam os objetivos do projeto, as funções dos envolvidos; a *segunda etapa* tem seguimento com as *oficinas presenciais* de formação para o Aluno Monitor, no NTEM, que foram organizadas em dois momentos. O primeiro momento foi constituído por três oficinas: Primeira Oficina: *Conhecimentos Técnicos Básicos sobre Hardware e Software (Linux Educacional)* - de forma a capacitar os alunos e professores para o uso dos recursos tecnológicos, proporcionando uma visão das tecnologias e suas potencialidades; Segunda Oficina: *Acesso ao Moodle* – Ferramentas de interação, especialmente o fórum, fortalecendo a noção de aprendizagem colaborativa e conhecimento do Moodle com apropriação dos conteúdos da formação [11]. Terceira Oficina: *Aplicativos do Linux Educacional* – Dentre os vários programas educacionais do Linux Educacional, optou-se por explorar o GCompris – um conjunto de jogos educacionais para crianças, nos quais os alunos aprendem brincando.

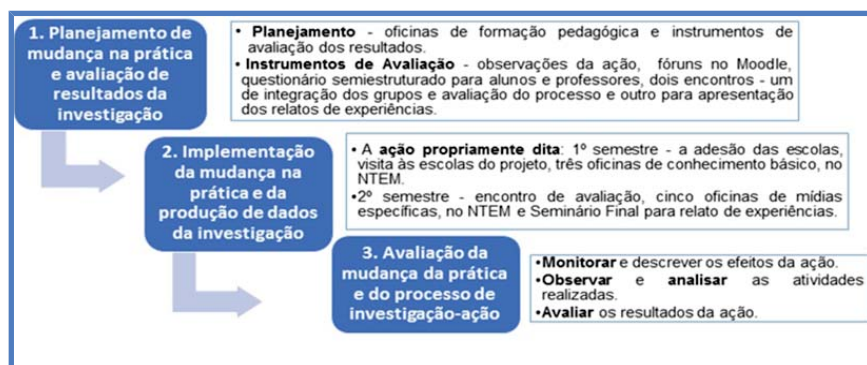
Após esse primeiro momento de formação, propicia-se um encontro entre todas as escolas participantes, com o objetivo de integrar alunos e professores para avaliar como está transcorrendo a implementação da proposta na escola.

No segundo momento de formação são trabalhadas *oficinas de mídias* mais específicas: Primeira Oficina: *Apresentações com o Impress* – Esta oficina baseia-se na construção de conhecimentos e linguagens da mídia-educação, [1] enfocando o uso do Impress na elaboração de slides. Segunda Oficina: *Jornal Escolar* – De acordo com Célestin Freinet [7], o jornal escolar aproxima a escola da vida e dos interesses dos alunos, sendo suporte para uma experiência de vida da criança, que se mobiliza e engaja nos recursos de comunicação, desenvolvendo sua criatividade e autonomia. Terceira Oficina: *Produção de Vídeos* – Oficina para elaborar e produzir vídeos, objetivando instigar a troca de conhecimentos, a criação de roteiros/contéúdos significativos e estimular a autoria dos alunos. Quarta Oficina: *História em Quadrinhos (HQ)* – A HQ trabalha a elaboração de textos, roteiros e argumentos a partir de conteúdos curriculares, de forma lúdica, atraindo a atenção dos alunos para as diferentes formas de leitura. Quinta Oficina: *Lousa Digital* – Essa oficina possibilita aos alunos monitores manusearem a Lousa e auxiliarem os professores para o uso dessa ferramenta em sala de aula. E, por último, ocorre, na *terceira etapa*, no final da formação, um *seminário* para relato das experiências, onde são apresentados atividades desenvolvidas nas escolas, utilizando as TIC abordadas.

4 Metodologia

Constituindo-se como uma pesquisa-ação, o Projeto estrutura-se em etapas ordenadas,

organizadas em três fases, na prática da investigação, caracterizando uma sequência ou ciclo, refletindo mudanças na prática e na ação Tripp,[15] Thiollent,[12].



Fig

ura 1 - Fonte: As autoras - adaptação de Tripp .[15]

Para Thiollent [12], “na pesquisa-ação existe ênfase na ação. As ações são discutidas, analisadas, deliberadas, decididas com pleno consenso ou não. São ações significativas para os atores em situação. (...) São interpretadas pelos pesquisadores com base em diferentes referenciais. (...) são ações portadoras de aprendizagem e conhecimento mútuo, com interações entre observadores e observados”, corroborando a ação que gerou esse Projeto. Portanto, a ação ocorre por meio de oficinas de formação de alunos monitores, com o propósito de integrar as tecnologias educacionais em rede na prática pedagógica dos professores municipais, de forma a instigar nos alunos o perfil protagonista para a utilização das TIC existentes na escola, como prática cotidiana de aprendizagem colaborativa. É esperado que os alunos aprendam a utilizar as tecnologias educacionais como recurso pedagógico e auxiliem os professores na transposição dos desafios da utilização da sala de informática. Considerando que a proposta está em andamento, com observações e coletas parciais de dados, pretende-se que a formação do Aluno Monitor contribua para a integração das tecnologias educacionais em rede à prática pedagógica dos professores, que as atividades auxiliem o desenvolvimento protagonista do aluno, potencializando o uso das TIC, contribuindo para a cultura da aprendizagem colaborativa na escola, inclusão digital e na multiplicação da ação para toda a rede municipal de ensino.

5. Considerações Finais

Atualmente, várias orientações da legislação educacional brasileira, apontam para o uso pedagógico das tecnologias educacionais perpassando os currículos. Quanto a verificação da situação dos Professores de Informática Educativa designados para as salas de informática das escolas públicas da rede municipal de Santa Maria, o estudo apontou a falta desses profissionais, apesar de legislação municipal assegurar sua função na escola, apontando o desmantelamento da proposta inicial do ProInfo, pelo governo municipal. A falta desses Professores, precarizou o atendimento à comunidade escolar nas salas de informática. Outro problema a obsolescência dos equipamentos nas escolas e no NTEM, o acesso a banda larga, o upgrade de

hardware e software educativos quesitos essenciais para um trabalho pleno com as TIC. Esses quesitos dependem de recursos e de escolhas, intencionais, quanto a quais investimentos manter e ampliar e quais preterir, exigindo uma posição da Mantenedora quanto a melhorar as condições tecnológicas das escolas, apoiando o trabalho do NTEM Santa Maria na formação e construção do perfil do Professor de Informática Educativa para integrar as tecnologias educacionais no cotidiano escolar, de forma a diminuir as “brechas digitais”. A carência desse profissional, nas escolas, fez com que surgisse a proposta de formação de alunos monitores para solucionar o impasse das salas de informática subutilizadas. A experiência pedagógica, de 2015, gerou o projeto de dissertação do Mestrado Profissional em Tecnologias Educacionais em Rede – UFSM, tornando-se um projeto de extensão universitária. A ampliação da proposta para todas as escolas municipais, possibilita a formação do Aluno Monitor para a integração das Tecnologias Educacionais em Rede à prática pedagógica dos professores e, com isso, visa a instigar um perfil protagonista nos grupos envolvidos. Deseja-se que a formação de alunos monitores estimule a utilização das tecnologias, perpassando pela apropriação e pelo aprimoramento de competências, habilidades e atitudes que resultam de mudanças na ação não só do aluno, no papel de monitor, mas também do Ser Humano que aprende-ensina-reflete-(re)aprende.

Referências

1. BELLONI, M.L. & BÉVORT, E. Mídia-Educação: Conceitos, História e Perspectivas. Educação & Sociedade, Campinas, vol.30, n.109, p. 1081-1102, 2009. Disponível em: www.cedes.unicamp.br
2. BRASIL. Ministério da Educação. Programa Nacional de Informática Educativa, 1989.
3. COLL, C. e MONEREO, C. Psicologia da Educação Virtual. Aprender e Ensinar com as Tecnologias da Informação e da Comunicação. Porto Alegre: Artmed, 2010.
4. CYSNEIROS, P. G. A Gestão da Informática na Escola Pública. In: XI SBIE 2000 – Simpósio Brasileiro de Informática na Educação. Maceió-AL: Anais. 2000.
5. DECRETO MUNICIPAL Nº 313/07, de 20-de dezembro de 2007, define assessoramento pedagógico nos termos da Lei 11.301/06.
6. DURAN, D. & VIDAL, V. Tutoria. Aprendizagem entre Iguais. Porto Alegre: Artmed, 2007.
7. FREINET, Célestin. O Jornal Escolar. Editorial Estampa, 1974. Disponível em: <https://salaaberta.com.br/wp-content/uploads/2015/06/o-jornal-escolar-freinet.pdf>
8. KENSKI, V. M. Tecnologias e Ensino Presencial e à Distância. Campinas: Papirus. 2006.
9. LINUX EDUCACIONAL. Disponível em: https://pt.wikipedia.org/wiki/Linux_Educacional
10. MONEREO, C. Prólogo. Aprender entre iguais e com iguais. In: Duran, D. & Vidal, V. Tutoria. Aprendizagem entre Iguais. Porto Alegre: Artmed, 2007.
11. SABBATINI, R.M.E. (2007) Ambiente de Ensino e Aprendizagem via Internet: A Plataforma Moodle. O que é Moodle? Disponível em: <http://www.ead.edumed.org.br>
12. THIOLENT, M. (2011) Metodologia da Pesquisa-Ação. 18ª ed. São Paulo: Cortez.
13. THIOLENT, M. & Oliveira, L. (2016). Participação, cooperação, colaboração na relação dos dispositivos de investigação com a esfera da ação sob a perspectiva da pesquisa-ação. In: Atas - Investigação Qualitativa em Ciências Sociais. v.3.
14. TORRES, P.L. (2004). Laboratório on-line de aprendizagem: uma proposta crítica de aprendizagem colaborativa para a educação. Tubarão: Ed. Unisul.15.
15. TRIPP, D. (2005) Pesquisa-ação: Uma introdução metodológica. Educação e Pesquisa, v. 31, n. 3, p. 443-446. SP.
16. VYGOTSKY, L. S. (1988) A formação social da mente. São Paulo: Martins Fontes.

Proactive Independent Learning Approach: A case study in computer arithmetic

Javier Giacomantone, Oscar Bria

Research Institute in Computer Science (III-LIDI) - School of Computer Science
National University of La Plata - Argentina
jog@info.unlp.edu.ar, onb@info.unlp.edu.ar

Abstract. The rapid growth of knowledge and scientific challenges required lifelong continuous education in computational science and engineering. Computer numerical system representation and computer arithmetic are the basis of numerical computing of scientific models. In this work an adapted student centered and problem based learning strategy is presented. Development of problem solving, effective self directed reasoning and communication skills are promoted. A pilot study was conducted to determine the validity of the proposed alternatives. The study aimed to evaluate the performance of students to solve new problems and effectively describe the problems, the theoretical context and the possible solutions. Preliminary results are presented for a particular population from which the sample is actually drawn.

Keywords: Lifelong proactive learning, student centered learning, problem based learning, computer arithmetic, numerical methods, mathematical models.

1 Introduction

Graduates of computer science and engineering programs are expected to face scientific and technological advances. Computational science and engineering is a field that has a high rate of knowledge and technological change. Self-learning, creative thinking and communication skills are needed to solve problems in scientific research and industry [1]. Their acquisition is a gradual and continuous process. It allows the learner to be able to develop a proactive lifelong learning practice and critical thinking. The process of prior knowledge integration is essential to learn any scientific discipline [2][3]. Integration of previous knowledge is fundamental from the first year of any university course to advanced professional interdisciplinary areas in computational science [4][5]. It is a challenge for a teacher not only to develop their subject, but to promote and guide the students to develop these fundamental abilities [6]. It is necessary a framework that promotes knowledge integration and positive feedback [7]. Learning schemes in the first year of a graduate university program in computer science face a few common problems [8][9]. In particular, the quality of education in primary and secondary schools in Argentina has deteriorated in the last decades

[10][11]. As a consequence, first year university student population presents a profound lack of some basic skills expected to be developed in secondary school. The background knowledge heterogeneity is another fundamental problem reported in the national evaluation tests [12]. Any learning strategy to reduce the gap between secondary school and university in Argentina has to consider many factors. It is a complex phenomenon that is out of the scope of this article. Nevertheless, certain aspects have to be considered in any attempt to improve the way computer arithmetic is taught. The aim of this paper is to present an alternative learning approach to achieve specific objectives. Students are encouraged to have more control and responsibility over the learning process [13] in a deeper approach to learning. In particular fixed and floating point representation of numbers and computer arithmetic for model implementation are the main topic of the presented case study. It could be included in a course of computer organization, numerical methods or advanced computing modelling [14]. Results of a pilot experience that has been carried out with students of computer science and engineering are reported.

This paper is organized as follows: Section 2 briefly describes the main ideas behind the proposed approach. Section 3 presents the particular case study. In section 4 results are presented. Finally, conclusions and future works are given in section 5.

2 Proactive Independent Learning Approach

Important developments in computer science and engineering education in recent years has been oriented to Problem-based learning strategy (PBL) [15]. PBL was first applied in the Medicine School at McMaster University (Canada) as an innovative educational proposal [16]. Although it was successfully adopted by other prestigious medical schools like Harvard Medical School, the particularities and necessary adaptations to engineering and computer science programs remains an active research area. A traditional teaching scheme or traditional learning approach (TLA) involves theoretical knowledge first taught to the students. It is followed by practical lectures explaining how to solve problems applying the previously learnt theoretical concepts. Finally the teacher sets an exam to test the basic knowledge and skills acquired by the students. The main characteristics of TLA are to set the teacher as the transmitter of linear and rational knowledge and the student as a passive receiver defining a structure environment of individual learning. PBL can be defined as a learning environment in which the problem solving process involves searching for information and discovering the new knowledge necessary to tackle the problem [17]. It has been shown in the literature that PBL assists to gain skills in problem solving and lifelong learning abilities in contrast to short term surface learning. In a PBL approach, small groups of students work collaboratively to solve a particular problem, with no previous preparation, with the student being the center of the learning process, constructing knowledge as an active participant in a flexible and cooperative environment. The teacher guides and facilitates the whole learning scheme. The

assessment is now shared among the student, the group and of course the teacher. PBL promotes self-learning, developing problem-solving skills, cooperative learning, and improving oral and written communication. An alternative scheme is proposed and named proactive independent learning approach (PILA). This alternative thread is related to problem based learning, active learning and lifelong learning [18]. It comprises a set of objectives, guidelines and constraints. It is not a set of strict steps to follow, but a flexible orientation, applicable to different learning contexts and approaches from TLA to PBL. The main objectives are to promote an active and independent attitude to learning. Independence and critical thinking are essential to develop a proactive approach beyond a particular course or career. Another central objective of PILA is to stimulates students to adopt lifelong and continued education. PILA encourage self discipline habits where learners become responsible for their own learning in university and beyond. No fixed guidelines are proposed, motivation and adaptation are keywords in PILA, both for students and teachers. It implies studying theoretical concepts on an individual basis, intensive practice on problem solving, writing technical essays, oral exposition and discussion of solutions. The proposed approach does not impose collaborative working as a requisite, or any other technique if for a specific student another one suits better. A fundamental step is to learn through solving problems, writing essays or other alternatives that need to integrate previous knowledge to be done. Certain constraints or boundary conditions should be addressed before any attempt to follow this alternative. One of the important constraints is the number of students, it requires relatively small groups if a team collaborative work is going to be adopted. Problems or projects need to be designed carefully involving many resources. A relatively common background knowledge of the students would be convenient as a prerequisite. It has been pointed out that a problem that requires high integration of previous knowledge is recommended from the beginning, reinforcing the main characteristics of the proposed approach. Nevertheless, if PILA is to be applied to a first year undergraduate level, gradual integration would be preferable. A gradual work example strategy [19] and progressive difficulty tasks [20] would be more convenient.

3 Case study: computer arithmetic

The computational and numerical modelling group at III-LIDI seeks to transfer the results of its scientific investigations and experience. In particular to improve teaching strategies in computational science and engineering related subjects [21][22]. The modelling group gathers experimental data of student performance in courses of: computer organization, numerical methods, machine learning, digital image analysis, statistical pattern recognition and digital signal processing. Encouraging excellence in a lifelong attitude towards continuous education in the area [23][24]. Computer numerical system representation and computer arithmetic are the basis of numerical computing [25][26]. A detail knowledge of numerical computer representation is necessary as a previous requirement for the implementation of any scientific numerical model [27][28]. Specific approaches to

teaching computer arithmetic have been proposed [29]. PILA is proposed as an attempt to improve learning of these advanced topics reducing the knowledge gap between basic and advance courses by building solid fundamentals. A pilot experience was carried out with a sample size of 30 students in the experimental group (PILA) and 30 in the control group (TLA). The students were assigned randomly into each group. A detail selection of the 60 students was carefully done to ensure similar previous conditions in background knowledge. Four tests were conducted, named T1 to T4, to evaluate written expression, standard problem solving and integral type problem solving. Each test was evaluated with a scale from 0 to 120 points. The TLA and PILA students frameworks were based on a set of selected topics where students presented mayor difficulties to model and solve problems. The TLA group assisted a 3 hours formal lecture once a week and another 3 hour solving problems class a week. These students learned theoretical concepts and received instruction on how to understand and solve specific related problems. The second group, PILA, assisted also 6 hours a week. Each week an optional, but strongly recommended 2 hours lecture introduced the fundamental topics to cover and problems to be solved. PILA promotes proactive and self directed learning, but it does not impose any limit on the teacher active participation particularly giving feedback as it is required. Both for the experimental and control group optional tasks were proposed. Students were required to submit a description of the problem and a possible solution. In table 1 the characteristics of the groups for test and optional activities are summarized.

Computer Arithmetic				
Type of activity	Tests T1 to T4		Opt. team work	
Group	TLA	PILA	TLA	PILA
Students	30	30	6	6
Teams	-	opt.	2	2
Tasks	-	opt.	1	3

Table 1. Group Characteristics

Optional activities for the control and the experimental groups were suggested. PILA encourage the idea of self directed research in a long time basis, naturally reading and analyzing of diverse bibliography is essential in computer science. Table 2 presents the number of optional activities of each group. The PILA group was faced with a more general real problem, called integral task. The students were challenged to work collaboratively to solve this complex problem. In order to find a solution to the proposed problem they needed to build up the necessary body of theoretical concepts, search bibliography and organize the work among them. At the core of PILA is to respect and encourage to adopt different learning

preferences, like working individually on the integration task. Finally, each group or individual student, had to present their results. The teacher guided the work, but the learning was centered on the student.

Optional activities	TLA	PILA
Suggested readings	6	6
Open optional tasks	1	3
Use of simulation tools	2	2
Collaborative working	1	3
Report presentation	1	3
Integral Task	-	1

Table 2. Number of optional activities

4 Results

The assessment scheme had three main parts: independent evaluation, comparative tests and a survey among students [30][31]. The tests were carefully designed not to alter the normal schedule of the course and not to overload the students subject to the experience with activities [32].

Groups	Test	Mean	SD	df	tvalue	pvalue
TLA	T1	86,56	39,23	58	1,69	0,0972
PILA	T1	72,13	26,12			
TLA	T2	83,79	32,52	58	1,81	0,0765
PILA	T2	96,28	19,53			
TLA	T3	73,81	35,52	58	2,19	0,0319
PILA	T3	88,31	15,62			
TLA	T4	67,14	28,61	58	2,53	0,0142
PILA	T4	82,93	18,67			

Table 3. Comparative tests

Table 3 presents the mean, the standard deviation and two sample independent t hypothesis tests for all the main evaluations. The null hypothesis is that there is no difference in the mean performance between the control and experimental group. T1 is a pre-test that attempt to verify the basic background and

similar skills of both groups at the beginning. T2 determines the performance of the groups to find problem solutions of common types of problems. T3 explores the skills to solve new problems or integral problems. T4 is a global test that involves solving standard problems, integral problems and evaluating writing skills. No initial significant difference was observed between TLA and PILA neither in the pre-test nor in the teacher observations. Figure 1(a) shows the boxplots for the pre-test. Figure 1(b) shows the boxplots for the final evaluation on problem solving for TLA and PILA. Both groups solved the problems without significant difference.

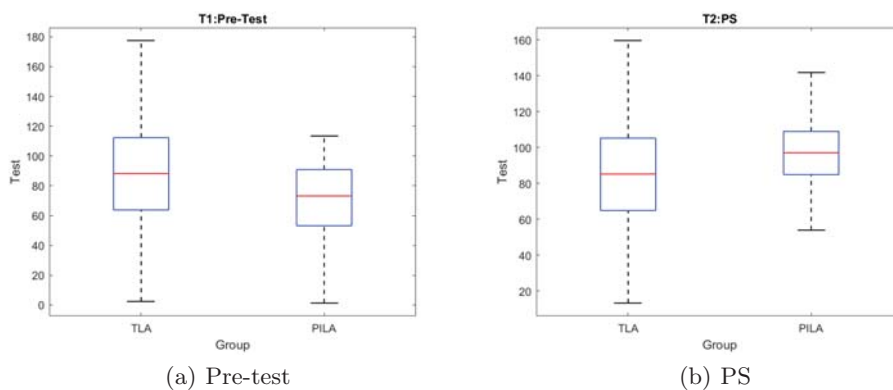


Fig. 1. Pre-test and problem solving test

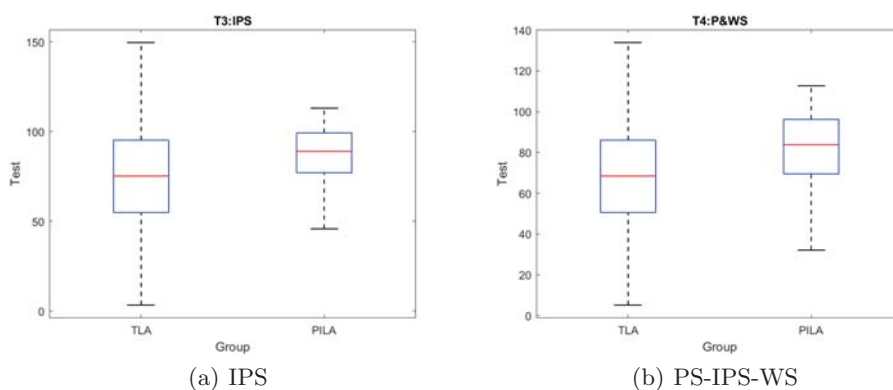


Fig. 2. Problem solving and technical writing skills

Figure 2(a) shows the boxplots of TLA and PILA for the integral problem solving evaluation. Figure 2(b) shows the boxplots for the global evaluation on problem solving, integral solving and technical writing skills and communication. PILA group presented a significant performance difference ($p < 0,05$) in T3 and T4.

Group	TLA	PILA
Open optional tasks	2	2,5
Use of simulation tools	4	4,5
Collaborative working	3	4
Report presentation	2	3
Worked example orientation	4	3,5
Optional integrating task	-	4,5
Overall learning experience	3	3,5

Table 4. Average marks from students opinion.

The information on students learning experience came from the teacher constructive communication with each group and from an anonymous questionnaire completed by the students. A list of statements was presented to the students of both groups. The scale from 0 to 5 indicate their agreement about the contribution to the learning experience from negative to positive respectively. The results are given in table 4.

5 Conclusions

This paper discusses a learning approach, particularly adapted to improve certain skills. It is based on PBL and shared many of its characteristics. The results of the presented case study indicate that PILA can be successfully applied for teaching and learning computer arithmetic at different levels. It has been observed that students in the PILA group were more motivated to develop proactive, independent learning skills. The analysis of results reveals that students of the experimental group displayed more effective abilities to solve problems that require previous knowledge integration. This study aimed at examining the effectiveness of an adapted approach on the performance of adult students. More extensive research needs to be conducted to analyze the performance on critical thinking, written and spoken communication. Further research may examine atypical values in the performance of students under different learning strategies.

References

1. Costa, L.R.J, Honkala, M., Lehtovuori A.: Applying the Problem-Based Learning Approach to Teach Elementary Circuit Analysis. *IEEE Transactions on Education* **50**(1), 41–48 (2007)
2. Bidokht M H., Assareh A: Life-long learners through problem-based and self directed learning. *Procedia Computer Science* **50**(3), 1446–1453 (2011)
3. Tunney M., Bell H. M.: Self-directed learning: preparing students for lifelong learning. *Pharmacy Education* **11**(1), 12–15 (2011)
4. Noor, M.: Implementing a problem based learning for undergraduate course: A first experience. *The Learning Conference*. Institute of Education, University of London, UK 15–18 (2003)
5. Montero, E., Gonzáles, M.J.: Student Engagement in a Structured Problem-Based Approach to Learning: A First Year Electronic Engineering Study Module on Heat Transfer. *IEEE Transactions on Education* **52**(2), 214–221 (2009)
6. Giacomantone, J., Tarutina, T.: Problem Based Learning. A Case Study in Computer Science First Year Students. *Computer Science and Technology Series. XV Argentine Congress of Computer Science Selected Papers*. Edulp 123-130 (2010).
7. Rodrigues A. N., Dos Santos S.: A framework for applying problem-based learning to Computing Education. *IEEE Frontiers in Education Conference* 1–7 (2016)
8. García Famoso, M.: Problem-based learning: a case study in computer science. *Proceedings of the Third International Conference on Multimedia and Information Technologies in Education* 817–821 (2005).
9. Dattatreya, G.R.: A Systematic Approach to Teaching Binary Arithmetic in a First Course. *IEEE Transactions on Education* **36**(1), 163–168 (1993).
10. Etcheverry, G. J.: *La Tragedia Educativa*. Fondo de Cultura Económica, Buenos Aires, (1999).
11. Programme for International Student Assessment. <https://www.oecd.org/pisa/>.
12. Sistema Abierto de Consulta. Ministerio de Educación. República Argentina. <http://www.educacion.gob.ar/secretaria-de-evaluacion-educativa/>.
13. Waters, R., McCracken, M.: Assessment and evaluation in problem based learning. *Proceedings of Frontiers in Education* **2**, 689–693 (1997).
14. Garzón, E. M, García I., Fernández J. J.: An approach to teaching computer arithmetic. *Proceedings of the 5th international conference on High performance computing for computational science* 269–283 (2002).
15. Linge, N., Parsons, D.: Problem-Based Learning as an Effective Tool for Teaching Computer Network Design. *IEEE Transactions on Education* **49**(1), 5–10 (2006)
16. Schmidt, H. G.: Problem-based learning: Rationale and description. *Medical Education* **17**, 11–16 (1983)
17. Schmidt, H. G.: Foundations of problem-based learning: some explanatory notes. *Medical Education* **27**, 422–432 (1993)
18. Ditcher, A. K.: Effective teaching and learning in higher education, with particular reference to the undergraduate education of professional engineers. *International Journal of Engineering Education* **17**(1), 24–29 (2001)
19. Sweller, J.: The worked example effect and human cognition. *Learning and Instruction* **16**(2), 165–169 (2006)
20. Merrill, M. D.: A Task-Centered instructional Strategy. *Journal of Research on Technology in Education* **40**(1), 33–50 (2007)
21. Mikhailovich T. Y., Yurievna A. A.: Application of problem-based learning technology in technical education. *International Forum on Strategic Technology* 474–478 (2016)

22. Korud V., Rendzinyak S., Hamola O.: The problem-based learning in electrical engineering. *International Scientific and Technical Conference Computer Sciences and Information Technologies* 116–118 (2016)
23. Barg M., Fekete A., Greening T., Hollands O., Kay J., Kingston J. H.: Problem-Based Learning for Foundation Computer Science Courses. *Computer Science Education* **10**(2), 109–128 (2000)
24. Samuel A. E.: Student centered teaching in engineering design. *Instructional Science* **15**, 213–238 (1986)
25. Parhami, B.: Generalized Signed-Digit Number Systems: A Unifying Framework for Redundant Number Representations. *IEEE Transactions on Education* **39**(1), 89–98 (1990)
26. Bharade P., Joshi Y., Manthalkar R.: Design and implementation of FIR lattice filter using floating point arithmetic in FPGA. *International Conference on Information Processing* 598–603 (2015)
27. Joldes M., Marty O., Muller J. M., Popescu V.: Arithmetic Algorithms for Extended Precision Using Floating-Point Expansions. *IEEE Transactions on Computers* **65**(4), 1197–1210 (2016)
28. Karthik K. R., Diwakara Reddy K. L., Pawar C. K., Sunitha K., Jayashree H. V.: Reconfigurable floating point arithmetic components for DSP applications. *IEEE Annual India Conference* 1–6 (2016)
29. Navi, K., Molahosseini, A., Esnaeildoust, M.: How to Teach Residue Number System to Computer Scientists and Engineers. *IEEE Transactions on Education* **54**(1), 156–163 (2011)
30. Dos Santos S. C.: An Authentic Assessment Model for PBL-Based Software Engineering Education. *IEEE Transactions on Education* **60**(2), 120–126 (2017)
31. Oliveira A., Salgado A. G., Magalhaes A., Faias J., Miranda L. G.: Perceived quality of life among first-year health students engaged in problem based learning and traditional teaching model: First-year students of allied health sciences. *Iberian Conference on Information Systems and Technologies* 1–7 (2016)
32. Sweller, J.: Cognitive load during problem solving: Effects on learning. *Cognitive Science* **12**(2), 257–285 (1988)

Student-Centered Learning in a University Programming Course

Carlos Fontela, Pablo Suárez,
Fac. de Ingeniería Universidad de Buenos Aires, Argentina
{cfontela, psuarez}@fi.uba.ar

Abstract. In teaching Object Oriented Programming to university students, we have noticed that motivation is vital for the correct transmission of key concepts. In particular, engaging students provides a particular challenge that has changed over the recent years. To address this, in our OOP course in the University of Buenos Aires, we have been incorporating a variation of the flipped classroom approach we call Student Centered Learning. We have seen a good reception of this on behalf of our students, which has given us reason to maintain and improve on this strategy, while looking for more hard data so support the results.

Keywords: learner-centered teaching, problem-based learning, active learning, collaborative learning, object-oriented programming teaching.

1 Introduction

1.1 From teaching to learning

In the last few decades there has been a movement from a teacher-centered educational paradigm to an experienced-based learner-centered one. Definitely, we are shifting the focus from teaching to learning [1]. This movement is referred through different keywords, such as “active learning” [2], “collaborative learning” [3], “cooperative learning” [4], “constructivism”, “learner-centered” [5], “problem-based” [6], “student-centered learning” [7]. Bowman's book is perhaps the most disruptive from outside the world of academia beginning with its challenging title [8]. The goal of all these practices is to accomplish knowledge appropriation naturally through active exploration, switching the responsibility of learning to the students, and having them work in groups in problem-solving activities, while teachers concentrate on mediation and coaching. The focus is on students more than teachers and learning more than teaching. In this article, we will employ the expression Student-Centered Learning (SCL) to summarize all these practices.

We may think there is a revolution taking place in education. Nevertheless, learner-centered education has a long history, from Confucius and Socrates to Hegel and Pestalozzi. However, in the last fifteen years, research on learning, cognitive neuroscience, and other related subjects have given a scientific background to what

had been empirically observed. In addition, these studies stated many more findings, such as the importance of movement, multi-sensory stimuli, and conversation to keep the brain active to facilitate the appropriation of new knowledge.

In any case, since 2014, we have decided to try SCL in a programming course, where students could actually learn by doing, making mistakes with very little cost, and learning from their mistakes. The fact that the computer gives an immediate response to the students' tries is another plus for the practice of SCL. All that said, we think these practices will help students learn more, and even make the activity of teaching more enjoyable, remembering we became teachers to improve students' lives. That year we explained our approach in a local workshop (WISIT 2014) and we gathered new recommendations and feedback [9].

1.2 Context

The course in which we work is the third programming course in the Software Engineering curricula, after one introduction to algorithms and programming with Python or C, and a data structures course. Therefore, the students reach our course with good background in basic programming. Our course is focused in Object Oriented Programming, with some topics of Object Oriented Design and Methodology.

The classes have a very large amount of students, with 150 to 200 in the classroom. The classroom is big, with steps that difficult both the students' and teachers' motion.

In this paper, we will describe how we manage to follow SCL practices, the limitations we are facing and some of the tradeoffs we have to make in our particular environment.

2 Related Works

There are many publications on SCL, its fundamentals, recommendations on applying and other aspects. An interesting book is the one from Doyle [10]. In this text Doyle makes interesting propositions, from SCL fundamentals to practical ways of classroom implementations.

Blumberg [11] offers us another interesting book that makes practical suggestions and tools to achieve SCL specifically in University courses.

Huba and Freed have undertaken the search for innovating in evaluation techniques for SCL educational environments [1]. There are other authors that have followed the same path, such as Saulnier et al. [12].

There are, however, very few, if any, publications which include actual classroom experiences, and even less in Latin America, nevertheless we know of many classes where SCL techniques are being used in teaching Computer Science and Software Engineering.

3 Learner-Centered Teaching in an Object Oriented Programming University Course

3.1 Classes

In the basic class structure we implement, in a course with 2 classes a week, each with a duration of 3 hours, we apply the following guidelines:

- No distinction between lectures and practice sessions.
- Many class start with a group dynamic as an opening activity (often involving students conversing in groups regarding a specific subject), as Bowman recommends [8].
- Two or three teachers per class.
- The students work with their computers in most classes.
- Ideally, two students share a computer doing pair programming.
- Sometimes, we start with a short lecture, involving as much interactive discussion with the students, and then we continue with exercises involving encapsulation, inheritance, polymorphism and some design aspects as well.
- Some classes revolve around the analysis of specific texts, having the students read them in the days prior to the class.
- Other class dynamics include starting with presenting the students with a practical exercise to resolve, and then presenting the theories behind the problem's resolution, introducing TDD, refactoring, polymorphism and other design principles.
- Role play is used for reinforcing certain more abstract subjects, such as delegation and MVC.
- Long lectures (over one hour) are avoided as a rule, as it is our perception that the current generation of students lose focus if practical exercises are not introduced frequently.
- Master classes? Yes, for specific subjects that will not be put into practice, but that require a conceptual explanation. In those cases, we make an effort to have as much interaction as possible with the students, asking questions every 15-20 minutes to reinforce what has been discussed, and trying to introduce real live anecdotes or experiences.

3.2 Practical Work

We have coursework that the students do outside the classroom.

According to Froyd y Simpson “the faculty member has the responsibility for forming teams” because “the teams that are formed influence the learning environment that is created” [7]. We have not observed the same, and believe in giving the students the liberty to build their own teams, reinforcing their involvement in the learning process.

- Groups of three or four students.

- In occasions, we evaluate the individual contribution of a specific student to their group.
- The feedback is given in-person.
- We make the students program games, since we believe this motivates them to look at the full scope of their work, and not just resolving a specific problem isolated from the real use of their work.
- The coursework is developed in an incremental fashion, following TDD guidelines, with feedback from a tutor in each incremental stage of their work.

3.3 Exams

In this course, according the current regulations, we have one mid-term exam and one final exam (called integration exam).

The mid-term exams are traditional tests, and in the following classes we resolve the test together with the students, and when the students get their grades, there is personalized feedback by with the teacher who graded the exam.

Final exams are done in the following manner:

- Students work on a computer (either their own, or one in our laboratories).
- They are given a set of questions and a short portion of working computer code, usually with important design flaws.
- Their task usually consists of identifying the code's shortcomings, suggesting improvements, expressing the improvements in UML diagrams and implementing them.
- They are usually interrupted by a teacher before finishing, and the written exam turns into an oral exam, where the student explain to the teacher his solution.
- The oral exam continues and the student is asked general theoretical questions spanning the semester's curricula.
- The student is given the result of the exam at the end of the oral examination.

We believe this form of examination is a better way to gauge the students knowledge, both giving them an opportunity to express themselves through their computer code, together with a more traditional oral examination.

3.4 Languages

The selection of computer languages is aligned with both the curricula and the particularities of the classroom dynamics.

We start with Smalltalk:

- No type checking.
- IDE is a part of the language platform.
- No configuration or deployment issues.

- This makes for a very contained environment, which helps students concentrate on the fundamental principles they are learning, with little distraction from non-OOP specific problems.

We continue with Java, which apart from being a strongly typed language, has a vast and readily available set of tools, including IDEs, code coverage, deployment frameworks, etc.

With these languages we focus on OOP principles, while in other courses after this one, the students work on more real-world and integrated assignments.

3.5 TDD

Test Driven Development (TDD) is used from the first class, initially using ad hoc tests and then incorporating a unit test framework (SUnit, JUnit).

This, along with refactoring, helps incorporating OOP concepts and understanding the benefits.

Towards the end of the course, reflection is introduced, and the use of frameworks like SUnit and JUnit helps understand the usefulness of reflection.

4 Discussion

4.1 Limitations

Are theoretical concepts actually understood?

Froyd and Simpson [7] found that many adopters indicate that this approach covers less content than traditional teaching methods, but we do not see any evidence of this happening.

Evaluating the students' advance is a challenge, since we consider short artificial exercises do not challenge the students sufficiently.

We do not have classrooms with computers available to accommodate all our students, so we depend on students bringing their own equipment, and sharing.

The classrooms often do not have room to move around, and cannot change the arrangement of tables.

What if students are not inclined to collaborate? This teaching dynamic requires working in small groups, and some students prefer traditional approaches that require less involvement (sitting, listening, taking notes) and allow them to stay within their comfort zone.

This makes it vital for us to explain what we are doing and why, trying to convey to them that this approach is in their benefit.

Although we tailor the course as the weeks advance, according to the characteristics of the group, content coverage is high priority for faculty members.

4.2 Results

At the end of each course we have the students, voluntarily, fill out a survey to gauge their opinion, and we have a retrospective activity during the final class. Through these, the feedback we get shows a preference for more practical classes, group dynamics and the students involvement in the learning process.

These activities do not result in hard data. Objective numbers are difficult to obtain, in part due to how the students are evaluated (we do not have standardized tests, like multiple-choice exams, for example). We believe the evaluation is an integral part of the learning process, and standardized examinations do contribute little to this.

We perceive that students are progressively acquiring better critical-thinking skills and greater conceptual understanding, but have no hard data to support this.

5 Future improvements

In the next semesters, we are considering improvements along the following lines of work:

- Implementing actions to measure if the teaching method is really effective, since we are currently basing our decisions on opinions of teachers and students.
- Changes to the mid-term exam (which is mandatory), since currently it is a traditional written test.
- Explore ways to help students develop teamwork abilities.
- Measuring the rhythm of study, as is proposed by Fontdevila et al. [13].

7 Conclusions

We believe that students can be co-creators together with teachers, collaborating in the construction of knowledge. Fundamentally, we believe they need to be lifelong learners, in all aspects of knowledge and more so in a profession that will require them to learn continuously. In this sense, our approach attempts to favor independent learning, and we believe that in this way the skills and concepts will be better incorporated by our students.

Although we cannot realistically expect to completely change their learning habits in one course, we believe we can introduce them into new ways to approach the learning of new skills or tools.

Our work also leaves us with some important questions, one of the most outstanding is how to measure the benefits of the SCL approach we are implementing. This is one area we plan to work on in the next semesters.

References

1. Huba, M., Freed, J.: Learner-Centered Assessment on College Campuses: Shifting the Focus from Teaching to Learning. *Journal Community College Journal of Research and Practice*, vol. 24, Issue 9 (2000)
2. Bonwell, C. C., Eison, J. A.: *Active Learning: Creating Excitement in the Classroom*. George Washington University Press. (1991)
3. Bruffee, K. A.: Collaborative Learning and the "Conversation of Mankind". *College English*, 46(7), pp. 635-652. (1984)
4. Johnston, S., Cooper, J.: Quick-thinks: Active-thinking Tasks in Lecture Classes and Televised Instruction. *Cooperative Learning and College Teaching*, 8(1), pp. 2-7. (1997)
5. Bilimoria, D., Wheeler, J.V.: Learning-centered education: a guide to resources and implementation. *Journal of Management Education*, 29(3), pp. 402-428. (1995)
6. Blumberg, P.: Problem-based learning: a prototypical example of learning-centered teaching. *Journal of Student Centered Learning*, 3(2), pp.111-125. (2007)
7. Froyd, J., Simpson, N.: Student-Centered Learning. Addressing Faculty Questions about Student-Centered Learning. What is meant by Student-centered Learning (SCL). *Science Education* (1997)
8. Bowman, S. L.: *Training from the back of the room!: 65 ways to Step aside and let them learn*. John Wiley & Sons. (2008)
9. Fontela, C., Páez, N., Suárez, J.P.: Lecciones aprendidas enseñando OO en UBA Ingeniería. *Workshop de Ingeniería en Sistemas y Tecnologías de Información*. (2014)
10. Doyle, T.: *Learner-centered teaching: Putting the research on learning into practice*. Stylus Publishing, LLC. (2012)
11. Blumberg, P.: *Developing Learner-Centered Teaching: A Practical Guide for Faculty*. Jossey-Bass. (2008)
12. Saulnier, B., Landry, J.; Longenecker, H. Jr; Wagner, T.: From Teaching to Learning: Learner-Centered Teaching and Assessment in Information Systems Education. *Journal of Information Systems Education*; West Lafayette 19.2, pp. 169-174 (Summer 2008)
13. Fontdevila, D., Tugnarelli, M., Ismael, S., & Videla, L.: Promoción del ritmo de estudio por feedback colectivo de progreso en trabajos prácticos. In XXI Congreso Argentino de Ciencias de la Computación CACIC. (2015)
6. National Center for Biotechnology Information, <http://www.ncbi.nlm.nih.gov>

VI Workshop Seguridad Informática (WSI)

Coordinadores

Díaz, Javier – Ramón, Hugo – Aciti, Claudio

Comité Científico

Acquistapace, Carlos (Argentina)

Benítez, Carlos (Argentina)

Castro Lechtaler, Antonio Ricardo (Argentina)

Cipriano, Marcelo (Argentina)

Díaz, Javier (Argentina)

Echaiz, Javier (Argentina)

García Garino, Carlos (Argentina)

García Villalba, Javier (España)

García, Edith (Argentina)

Liporace, Julio César (Argentina)

Maiorano, Ariel (Argentina)

Molinari, Lia (Argentina)

Ramió Aguirre, Jorge (España)

Sandoval Orozco, Ana Lucía (España)

Saroka, Raúl (Argentina)

Tarouco, Liane (Brasil)

Venosa, Paula (Argentina)

Análisis de metodologías de recolección de datos digitales en servidores web

Mónica D. Tugnarelli ⁽¹⁾, Mauro F. Fornaroli ⁽¹⁾, Sonia R. Santana ⁽¹⁾,
Eduardo Jacobo ⁽¹⁾, Javier Díaz ⁽²⁾

(1) Facultad de Ciencias de la Administración – Universidad Nacional de
Entre Ríos

(2) Facultad de Informática – Universidad Nacional de La Plata

*e-mail: montug, maufor, ssantana, ejacobo [@fcad.uner.edu.ar]
jdiaz [@unlp.edu.ar]*

Abstract. Cuando se produce un incidente o amenaza de seguridad, en el cual un recurso del sistema queda comprometido o potencialmente expuesto a accesos no autorizados, las técnicas y metodologías de forensia informática deben asegurar que se pueda determinar adecuadamente el qué, quién, cuándo y cómo sucedió el incidente, así como también ocuparse del aseguramiento y preservación de la evidencia recolectada. Este trabajo aborda dos metodologías de recolección de datos digitales, la primera llamada *Enfoque preventivo-Recolección de datos a priori* o *Forensic Readiness* y la segunda *Enfoque reactivo - Recolección de datos a posteriori de un evento de seguridad*, para analizar comparativamente sus prestaciones en base a determinados criterios y puntos de control establecidos sobre servidores web HTTP.

Keywords: seguridad, incidente, forensia, metodologías, HTTP.

1 Introducción

Si una arquitectura de seguridad informática está correctamente definida debe ofrecer un plan y un conjunto de políticas que describan tanto los servicios de seguridad ofrecidos a los usuarios como los componentes del sistema requeridos para implementar dichos servicios. Estas políticas de seguridad se aplican a los activos de información identificados por su relevancia con los objetivos de la organización, conociendo como se gestionan y cuáles son sus riesgos para implementar estrategias y mecanismos que aseguren la confidencialidad, la integridad y la disponibilidad de los mismos [1].

Cuando se produce un incidente o amenaza de seguridad, en el cual un recurso queda comprometido o potencialmente expuesto a accesos no autorizados, esta

arquitectura de seguridad se ve vulnerada. A modo general, como amenazas del entorno, pueden considerarse aspectos que incluyan desde la seguridad administrativa, la seguridad de las comunicaciones, la seguridad informática, la seguridad ambiental hasta la seguridad física. Entonces, la arquitectura de seguridad debe poder afrontar tanto amenazas intencionales como accidentales. Implementar un programa sistemático de monitoreo y gestión de incidentes, basado en el empleo de metodologías, puede proporcionar un enfoque estructurado y organizado para minimizar el impacto del incidente de seguridad y ayudar a proporcionar una respuesta rápida y adecuada.

Diariamente cientos de equipos se encuentran expuestos a potenciales incidentes, consideremos como ejemplo el avance de Internet de las Cosas (IoT) y sus características de trabajo para llegar a dimensionar el grado de posibilidad y el riesgo de ocurrencia de un incidente y su consecuente impacto [2,3]

En este entorno tecnológico las metodologías de forensia informática deben asegurar que se pueda determinar adecuadamente el qué, quién, cuándo y cómo sucedió en relación a ese incidente de seguridad, así como también ocuparse de la correcta preservación y trazabilidad de los datos recolectados.

La definición brindada por la primera Digital Forensics Research Workshop (DFRWS), acuerda que el análisis forense digital o forensia informática es *“El uso de métodos científicamente probados y derivados hacia la preservación, recolección, validación, identificación, análisis, interpretación, documentación y presentación de evidencia digital derivada de fuentes digitales con el fin de facilitar o promover la reconstrucción de los hechos, que pueden constituirse en evidencia legal, o ayudando a anticipar acciones no autorizadas que han demostrado ser perjudiciales para operaciones planeadas.”* [4].

Las fuentes proveedoras de datos son numerosas, abarcan desde computadoras, teléfonos celulares, tarjetas de cámaras digitales, chips embebidos, drones, *snapshots* de memoria hasta consolas de videojuegos, es decir cualquier tipo de dispositivo que produzca datos digitales.

La forensia informática requiere entonces, una correcta aplicación de métodos científicos, técnicas y herramientas para cumplimentar las etapas relacionadas con la identificación, preservación y análisis de la evidencia digital la cual, llegado el caso, puede ser considerada legalmente en un proceso judicial por lo cual además se necesita asegurar la calidad y trazabilidad de estos datos.

Planteado este panorama, el Proyecto de Investigación y Desarrollo PID 7052 realizado en la Facultad de Ciencias de la Administración de la Universidad Nacional de Entre Ríos busca avanzar en el estudio comparativo de metodologías de recopilación de datos asociados a incidentes de seguridad y, particularmente, analizar la performance de estas metodologías en entornos de servidores web. Para ello se han establecido cuatro etapas: Estudio exploratorio de las metodologías de recolección de datos; Análisis del nuevo protocolo HTTP/2; Configuración del entorno de testing y determinación de puntos de control y captura; y por último la construcción de una matriz comparativa entre las metodologías y sus aspectos más relevantes en cuanto a calidad, trazabilidad, disponibilidad de datos y el tiempo de respuesta ante un incidente.

Las tres primeras etapas se han cumplimentado y en los puntos siguientes se presentan de manera resumida los resultados de las mismas. La última etapa de análisis de datos, se encuentra en ejecución y se presentan resultados parciales.

2 Estudio exploratorio de las metodologías de recolección de datos

Actualmente, y a modo general, las metodologías de recolección de datos pueden clasificarse en dos enfoques:

2.1- Enfoque preventivo-Recolección de datos a priori de un evento de seguridad. También conocido como *Forensic Readiness* [5,6,7]. Este enfoque introduce el concepto de resguardar la posible evidencia antes de que ocurra el incidente para cubrir principalmente dos objetivos: maximizar la capacidad del entorno para reunir evidencia digital confiable y minimizar el costo forense durante la respuesta a un incidente. La premisa es que esos datos puedan ser pasibles de ser utilizados no solo como insumo para el análisis de posibles incidentes de seguridad y de recuperación para la continuidad del negocio, sino también como prueba legal lo que involucra el aseguramiento de la prueba a medida que se realiza la recolección activa de los datos.

No solo es un enfoque beneficioso para organizaciones, sino que también puede aplicarse en otros ámbitos, como por ejemplo en los sistemas de voto electrónico, en los cuales no solo hay que ocuparse de la seguridad del software de votación y de los protocolos criptográficos sino que también se requiere de un entorno de confianza que tenga la capacidad de preservar la evidencia para mostrar a los votantes que el proceso funcionó como se esperaba.

Este enfoque, además, plantea que estar preparado para reunir y utilizar evidencia también puede tener beneficios como elemento disuasorio considerando los altos porcentajes de infracción de políticas internas de seguridad.

Algunas de las actividades clave en la planificación de la Forensic Readiness son:

- a. Definir los escenarios o activos que pueden requerir de pruebas digitales;
- b. Identificar las fuentes disponibles y los diferentes tipos de posibles pruebas;
- c. Establecer una forma segura de obtención de pruebas para cumplir con el requisito de admisibilidad legal;
- d. Establecer una política para el almacenamiento seguro y el manejo seguro de las evidencias;
- e. Garantizar el seguimiento para detectar y prevenir incidentes mayores;

- f. Capacitar al personal de modo que todos entiendan su papel en el proceso de pruebas digitales y la sensibilidad jurídica de las mismas;
- g. Garantizar el control jurídico para facilitar la acción en respuesta al incidente.

En resumen, la capacidad de una organización para explotar estos datos y la anticipación de la respuesta a un incidente son el foco de la Disponibilidad Forense complementando y mejorando las actividades de la organización en cuanto a seguridad de la información y evaluación de riesgos.

2.2- Enfoque reactivo - Recolección de datos a posteriori de un evento de seguridad. En este enfoque se trata de recuperar la evidencia luego de la detección de un incidente de seguridad, con el objetivo de realizar un análisis forense para determinar lo ocurrido. Este examen forense debe preservar el escenario y garantizar la admisibilidad de las pruebas. Piccirilli [8] aporta en su tesis doctoral, una descripción de las etapas que se pueden aplicar en todos los casos periciales en los que intervengan elementos vinculados a la informática, las cuales incluyen:

- el estudio y análisis del entorno, para identificar la evidencia digital a obtener;
- el análisis de los puntos de pericia, que establecen el objetivo que debe cumplir la evidencia digital;
- la adquisición de la evidencia digital;
- el análisis de la evidencia obtenida, conforme a los lineamientos del cuestionario pericial ordenado;
- la forma de exponer la evidencia digital obtenida en la investigación realizada;
- la preservación de la evidencia digital tratada (para eventuales futuras etapas de investigación, cuya fuente sería la misma evidencia digital).

Los estándares y recomendaciones más ampliamente considerados, tales como la RFC 3227 [9] y la ISO/IEC 27037 [10], plantean un conjunto de puntos comunes para realizar un análisis forense correcto, entre los que se incluyen: la importancia de preservar el entorno de pruebas, cómo y donde se guardan las pruebas, cómo se analizan para obtener el máximo rendimiento y la importancia de la obtención de informes claros y concisos [11,12].

Si bien muchas organizaciones son conscientes de la necesidad de contar con planes de recuperación y continuidad de negocio, la tendencia es implementar este enfoque, es decir, esperar a que ocurra un incidente, tratar de manejarlo y luego reunir las evidencias.

3 Análisis del Protocolo HTTP/2

Hypertext Transfer Protocol (HTTP) es un protocolo de nivel de aplicación con características definidas para utilizarse en sistemas de información distribuidos, colaborativos e hipermediales. Se caracteriza por ser un protocolo cliente/servidor sencillo y ampliamente aceptado, que define la estructura de los mensajes de requerimiento/respuesta así como también la forma en que se realiza el intercambio de dichos mensajes entre los clientes y los servidores web. Se han introducido

mejoras en sus diferentes versiones sobre todo tendientes a mejorar la performance del mismo, a reducir el consumo de recursos y la latencia y para resolver algunos inconvenientes que tiene la comunicación a través TCP [13,14].

La última versión HTTP/2 [15], presenta un protocolo binario que incorpora multiplexación y el uso obligatorio de TLS¹ conservando la misma semántica y la compatibilidad con las versiones 1.0 y 1.1 El protocolo se implementa si el cliente y el servidor tienen soporte y en el caso de que alguno de los dos no lo tengan, en la negociación de protocolo, se acuerda usar las versiones anteriores. Actualmente la mayoría de los browsers y entornos de servidor cuentan con implementaciones oficiales para la nueva versión.

En la siguiente tabla, a modo resumen, se describen las principales diferencias entre las versiones:

Versión	HTTP/1.0 (1996)	HTTP/1.1 (2000)	HTTP/2 (2015)
<i>RFC</i>	<i>RFC 1945</i>	<i>RFC 2616</i>	<i>RFC 7540</i>
Manejo de requerimientos	Un requerimiento entregado por vez sobre una conexión.	Mecanismo HTTP Keep Alive: varios requerimientos pueden utilizar múltiples conexiones con el servidor para reducir la latencia.	Múltiples mensajes de requerimiento/ respuesta sobre una misma conexión. Permite asignar prioridades a los requerimientos.
Header	Formato texto	Formato texto	Formato binario. - Compresión de header (Algoritmo HPACK).
Multiplexación	No permite conexiones simultáneas	No permite conexiones simultáneas	Permite múltiples solicitudes y respuestas en paralelo usando la misma conexión TCP, enviando cada requerimiento en un stream diferente.
Servidor	Descarga de recursos a solicitud del cliente (primero HTML, luego CSS, JS, imágenes, enlaces)	Descarga de recursos a solicitud del cliente (primero HTML, luego CSS, JS, imágenes, enlaces)	Tecnología server push: Permite cargar los archivos (CSS, JS, imágenes) desde el servidor al cliente sin que éste lo pida.

¹ TLS (Transport Layer Security) es un protocolo criptográfico que proporciona comunicaciones seguras en una red.

Tabla 1.Principales diferencias entre versiones HTTP

4 Configuración del entorno de testing y determinación de puntos de control y captura

Para desarrollar las actividades y pruebas se ha configurado un entorno de trabajo en una red LAN Ethernet que se conforma con un servidor donde corre un Sistema Operativo Ubuntu Server versión 15.10 y un Servidor Web Apache Server versión 2.4.12 que cuenta con una dirección IP pública. Esta red se completa con seis estaciones de trabajo con conexión a la red cableada y a red inalámbrica.

Para la ejecución de pruebas y adquisición de datos se analizaron herramientas de forensia informática de libre distribución [16,17], tales como CAINE [18], BlackArch Linux [19] y KALI Linux [20]. Se optó por Kali Linux versión 64bit 2017.1, que cuenta con más de 300 herramientas y aplicaciones relacionadas con la auditoria y la forensia.

Como guía general para las pruebas y marco de trabajo se seleccionaron los siguientes documentos:

- RFC 3227: publicado por la *Internet Engineering Task Force* (IETF) donde establece directrices para recopilar y almacenar evidencias sin ponerlas en riesgo.
- ISO/IEC 27037:2012 que proporciona directrices para el manejo adecuado de la evidencia digital gobernada por tres principios fundamentales: la relevancia, la confiabilidad y la suficiencia.
- OSSTMM (*Open Source Security Testing Methodology Manual*) [21]. Uno de los estándares profesionales más completos y utilizados para auditorías de seguridad en sistemas.

La primer actividad consistió en Identificar los puntos de control del protocolo HTTP, para lo cual se realizó la captura, análisis y resguardo de:

- a. tráfico entrante y saliente de los puertos 80 y 443 TCP, con el fin de obtener patrones de tráfico.
- b. estado de las conexiones establecidas a los puertos 80 y 443.
- c. archivos log de Ubuntu Linux (*/var/log/*):
 - messages.log: registro de mensajes generales del sistema
 - auth.log: registro de autenticación
 - secure: registro de autenticación
 - utmp/wtmp: registro de login
- d. httpd: archivos log de Apache: error.log y access.log. El primero proporciona información de diagnóstico y registra cualquier error que ocurre en el procesamiento de los requerimientos; mientras que el segundo almacena todos los requerimientos procesados por el servidor.
- e. archivos de configuración del servidor Apache: a fin de determinar modificaciones no autorizadas en la configuración del servidor alterando su funcionamiento.

La recolección de datos contempla las particularidades de cada enfoque:

a. **Enfoque preventivo**

- Monitoreo y recopilación de datos según los puntos establecidos y detallados en la actividad anterior.
- Diariamente se realizan dos copias de los datos recolectados los cuales son almacenados en medios externos con protección de integridad mediante hash (MD5).

b. **Enfoque reactivo**

- Monitoreo de datos según los puntos establecidos y detallados en la actividad anterior.
- Se realiza un backup estándar diario de los datos

5 Análisis de datos recolectados y construcción de matriz comparativa

Si bien esta etapa esta en ejecución, un análisis inicial de los datos recolectados permite arribar a los siguientes resultados parciales:

- Identificar la fuente y tipo de evidencia requerida agiliza la recolección de pruebas y el almacenamiento de las mismas. Esto, indispensable para el enfoque preventivo, complementa el análisis de riesgo de los activos de la organización. En este caso, los puntos de control especificados sobre el servidor web concentran los recursos dedicados a la captura de datos.
- El registro centralizado es clave tanto para una detección eficiente de un incidente como para la implementación de herramientas forenses.
- El resguardo periódico de datos en un sistema externo los aísla de posibles incidentes de seguridad sobre el servidor web. El volumen de almacenamiento obviamente dependerá de los activos y puntos de control necesarios determinantes para la continuidad del negocio. En este proyecto, y por sus características, el resguardo diario cumple con estos requisitos equilibrando volumen (promedio de 2 Gb), accesibilidad y costo de almacenamiento.
- Considerando la metodología Forensic Readiness, donde lo recolectado es evidencia digital pasible de ser utilizada en instancias legales, la cadena de custodia se puede implementar con almacenamiento externo, la generación de hash por bloque de datos y la incorporación de datos administrativos del tipo fecha, hora y personal actuante.

El próximo paso de esta etapa es la simulación de un ataque activo de Denegación de Servicio (DoS) sobre el servidor de prueba, a los fines de analizar de manera completa los diversos aspectos ya mencionados relacionados con la performance de

las metodologías, y fundamentalmente dar respuesta sobre algunas cuestiones pendientes tales como:

- ¿Qué enfoque proporciona mejores tiempos de recuperación operativa?
- ¿Qué metodología proporciona una mejor respuesta ante un incidente de seguridad?
- ¿Qué enfoque brinda un entorno más adecuado para realizar la forensia luego de los incidentes?

Conclusiones y trabajos futuros

En este artículo se han presentado los primeros avances del PID 7052-UNER denominado Análisis de Métodos de Recolección de Datos Digitales iniciado en abril de 2017.

En cuanto a las metodologías de recolección propuestas para el análisis, se realizó un estudio exploratorio de material y publicaciones relacionadas, logrando la identificación de las principales características de los mismos ante un incidente de seguridad.

En una segunda instancia, se analizó la nueva versión del protocolo HTTP con la finalidad de conocer inicialmente sus implicancias en la captura del tráfico de datos. Desde el punto de vista que interesa a este trabajo no se detectaron grandes cambios a nivel aplicación, sí se debería avanzar a futuro en el análisis de su relación con TCP y las restricciones de seguridad con el uso de TLS.

Para la configuración del entorno de trabajo y testing se utilizó software libre y herramientas de licenciamiento libre tales como Ubuntu, Apache y Kali Linux una distro GNU especializada en seguridad informática.

Si bien la tendencia es la respuesta reactiva, es decir, esperar a que ocurra un incidente, tratar de manejarlo y luego reunir las evidencias, los resultados iniciales del análisis de datos recolectados y tratados como evidencia digital muestran indicios de los beneficios tangibles y la importancia que puede tener para una organización adoptar el enfoque preventivo de preparación forense especialmente porque una gran parte de la evidencia requerida está disponible antes del incidente.

Referencias

- [1] Incident Management and Response ISACA <http://www.isaca.org/>
- [2] Internet Crime Complaint Center (IC3). Annual Report 2015 <http://www.ic3.gov/media/annualreports.aspx>
- [3] Ministerio Público Fiscal de la Ciudad Autónoma de Buenos Aires. CyberCrime Informe Final 2013 - Delitos Informáticos. <http://delitosinformaticos.fiscalias.gob.ar/wp-content/uploads/2014/02/CyberCrime-Informe-Final-2013-flip.pdf>
- [4] Digital Forensic Research Workshop (DFRWS). <http://www.dfrws.org/>
- [5] TAN, John. (2001). *Forensic Readiness*. http://isis.poly.edu/kulesh/forensics/forensic_readiness.pdf

- [6] Rowlingson, Robert. *A Ten Step for Forensic Readiness*. (2004) International Journal of Digital Evidence. Volume 2, Issue 3.
- [7] Poee, A., Labuschagne, L. *A conceptual model for digital forensic readiness* (2012) <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?reload=true&arnumber=6320452>
- [8] Piccirilli, Dario. (2016). *Protocolos a aplicar en la forensia informática en el marco de las nuevas tecnologías (pericia – forensia y cibercrimen)*. Tesis de doctorado. Facultad de Informática. Universidad Nacional de La Plata. <http://hdl.handle.net/10915/52212>
- [9] RFC 3227 Guidelines for Evidence Collection and Archiving. <https://www.ietf.org/rfc/rfc3227.txt>
- [10] Guidelines for identification, collection, acquisition and preservation of digital evidence” ISO/IEC 27037:2012
- [11]. U.S. Department of Justice. Electronic Crime Scene Investigation: A Guide for First Responders, Second Edition. <https://www.ncjrs.gov/pdffiles1/nij/219941.pdf>
- [12] Forte, D. *Principles of digital evidence Collection* (2003) <http://www.sciencedirect.com/science/article/pii/S1353485803000060>
- [13] RFC 1945 Hypertext Transfer Protocol - HTTP/1.0 <http://tools.ietf.org/html/rfc1945>
- [14] RFC 2616 Hypertext Transfer Protocol - HTTP/1.1 <http://tools.ietf.org/html/rfc2616>
- [15] Hypertext Transfer Protocol Version 2 (HTTP/2). <https://tools.ietf.org/html/rfc7540>
- [16] Digital Forensic with Open Tools. (2011). DOI: 10.1016/B978-1-59749-586-8.00001-7. Elsevier.Inc
- [17] Tugnarelli, M.; Fornaroli, M.; Pacifico, C. *Análisis de prestaciones de herramientas de software libre para la recolección a priori de evidencia digital en servidores web*. Workshop de Investigadores en Ciencias de la Computación (WICC 2015). ISBN 978-987-633-134-0
- [18] Computer Aided Investigative Environment <http://www.caine-live.net/>
- [19] BlackArch Linux. <https://blackarch.org/>
- [20] KALI Linux. www.kali.org
- [21] Open Source Security Testing Methodology Manual (OSSTMM) <http://www.isecom.org/mirror/OSSTMM.3.pdf>

Aspectos de Seguridad en Internet de las Cosas

Alicia Castro¹, Eduardo Casanovas³, Veronica Gil-Costa^{1,2},

¹ Universidad Nacional de San Luis, San Luis Argentina

² CONICET, Consejo Nacional de Investigaciones Científicas y Técnicas

³ Instituto Universitario Aeronáutico

adcastro@unsl.edu.ar; ecasanovas@iua.edu.ar; gvcosta@unsl.edu.ar

Resumen. Actualmente, los ataques contra la ciberseguridad han producido efectos económicos adversos, afectando no sólo a compañías privadas, sino también a estados gubernamentales y a las personas en sus hogares. La diversidad de ataques que se producen continuamente alrededor del mundo y la diversidad de dispositivos – en la mayoría dispositivos de IoT - con las que se utilizan, hace casi imposible generar un único protocolo o framework de seguridad que sea aplicable en todos los casos. Más aún, estadísticamente se ha reportado que la mayoría de los casos de ataques se producen por la falta de protección básica (como cambios en claves de seguridad) que los usuarios finales deberían proveer a sus dispositivos. En este trabajo se clasifican los dispositivos IoT y las vulnerabilidades. Se realiza un análisis de los protocolos, modelos y arquitecturas de seguridad existentes.

Keywords: Internet de las cosas. Seguridad. Privacidad. Smart Cities. IoT. Arquitecturas de Seguridad IoT. Ciberseguridad.

1 Introducción

En la actualidad encontramos Internet de las Cosas (Internet of Things - IoT) que permite conectar enorme cantidad de dispositivos y brindar servicios en distintos ámbitos de la vida diaria, obteniendo y distribuyendo información por la web, incrementando la apertura y complejidad de las redes, lo que lleva a nuevos desafíos de seguridad. Gartner.Inc estima que habrá más de 20 millones de dispositivos conectados para el 2020 [1].

Los dispositivos IoT pueden ser utilizados en distintos ambientes: industria, ciudades inteligentes (Smart Cities), agricultura inteligente, edificios inteligentes, Salud, finanzas, etc. De acuerdo a la utilización que se dé a estos dispositivos, se pueden categorizar en [1]: dispositivos para consumidores finales, dispositivos utilizados en el área de salud, dispositivos de uso en infraestructuras industriales (IIoT) y dispositivos utilizados en Smart Cities. Dentro de la primera categoría encontramos todos los electrodomésticos inteligentes, que si bien resultan útiles para controlar tareas rutinarias del día a día, también ofrecen a los ciberdelincuentes una nueva puerta de entrada. Respecto a los dispositivos de uso en el área de salud, se pueden observar que su uso se incrementa día a día, desde instrumentación médica con conexión inalámbrica (Estetoscopio inalámbrico, termómetro, monitor de

glucosa), productos farmacológicos (caja de pastillas inteligentes), herramientas de diagnóstico (Diabetes, Azoi), hasta dispositivos IoT implantables (desfibrilador, marcapasos). Los dispositivos de IoT utilizados en ciudades inteligentes (Smart City - SC) se utilizan en distintos entornos, geográficamente dispersos, algunos de ellos continuamente en movimiento. Se podría decir que una SC está conformada por una grilla inteligente de “cosas vulnerables” conectadas que colaboran en el funcionamiento de una sociedad moderna. En sistemas industriales, los dispositivos utilizados ayudan a automatizar la transmisión de datos entre dispositivos mecánicos o eléctricos.

Existen distintos factores que contribuyen a que los dispositivos IoT sean vulnerables: mezcla de malos diseños, entornos no regulados o regulados ineficientemente, entre otros. Los atacantes pueden usar las vulnerabilidades de estos dispositivos para obtener acceso permanente a estos equipos y a la infraestructura de red por la cual están conectados. Durante el año 2017, se han realizado nuevos ataques que utilizan distintos tipos de dispositivos, que ponen en riesgo la seguridad y privacidad de los dispositivos. Un informe brindado por Ponemon Institute y el Shared Assessments Program [2], indica que el 63% de los ataques registrados fueron dirigidos en contra de cámaras IP o grabadores digitales de vídeo y el 20% estuvieron dirigidos a dispositivos de red como módems DSL, routers y equipos similares, afectando a naciones como: China (17%), Vietnam (15%), Rusia (8%) y en un 7% a Taiwán, Turquía y Brasil.

En este trabajo presentamos un análisis y clasificación de los dispositivos utilizados en IoT y sus vulnerabilidades. En particular, nos enfocamos en el aspecto de la comunicación entre los dispositivos para detectar las características que deben ser consideradas en diferentes niveles o capas de seguridad. Para ello, en la Sección 2 presentamos las características de los dispositivos IoT y su arquitectura. En la Sección 3 se muestran las características referentes a la seguridad informática. La sección 4 presenta las conclusiones y trabajos futuros.

2 Características de los dispositivos IoT

Los dispositivos IoT son minicomputadoras. Cuyo hardware, son procesadores (microcontroladores) y plaquetas para uso específico, por lo cual existen diferentes fabricantes y diversidad de modelos. Con respecto al software, algunos disponen de firmware, otros de Sistema Operativo (Google tv, WebOS, Android Things, Tizen, Zephyr) y de aplicaciones de propósito específico.

Existe una gran cantidad de dispositivos IoT que no son interoperables, aunque existe soluciones de software para lograr interoperabilidad entre dispositivos de diferentes fabricantes con distintos sistemas operativos. Por ejemplo, AllJoyn que se utiliza en los televisores LG y en altavoces Panasonic. Se han conformado asociaciones de empresas que definen estándares comunes, uno de ellos es Open Interconnect Consortium [3], que brinda código abierto para crear un estándar común.

A nivel de comunicaciones, se pueden adquirir diferentes módulos que proveen distintos protocolos de comunicación. Se pueden distinguir entre dos grupos, según la aplicación que se pretenda dar: (1) sin restricciones o limitaciones (NTU), con enlaces de comunicación de alta velocidad, en el rango de transferencias con una tasa igual o

superior al Mbit/s y bajas latencias en el nivel de enlace, más afectadas por las congestiones en la red que por las propias características físicas de los elementos, y (2) restringidas o limitadas (NTC) con enlaces de comunicación de baja velocidad, en el rango de transferencias inferior al Mbit/s y grandes latencias, tienen tecnología a nivel físico de baja capacidad y políticas de ahorro de energía. En cuanto a los protocolos de comunicaciones en dispositivos IoT, podemos encontrar diferentes protocolos.

Gran parte de los dispositivos IoT de consumidores, utiliza el entorno web para la transmisión y visualización de los datos, por lo cual se utilizan los protocolos HTTP, websocket, XMPP. Sin embargo, han surgido alternativas a estos protocolos cuyo consumo de recursos es menor, por ejemplo, CoAP (protocolo de aplicación para redes con baja potencia y pérdida). Un ejemplo son las aplicaciones en los dispositivos móviles para monitoreo de actividades físicas que usan HTTPS para transferir comunicaciones con los servidores remotos.

Para redes Inalámbricas de Área Personal (WPAN), con capacidades limitadas en procesamiento, poca memoria, baja potencia, bajo costo y corto alcance, existen protocolos que trabajan bajo la normativa de la IEEE 802.15.4 como (a) LoWPAN, (b) 6lowPAN definido en la RFC 4944 para IPv6, (c) Zigbee, protocolo estándar en redes de sensores. También están los protocolos basados en la especificación de la IEEE 802.15.1 (Bluetooth), que facilita la interoperabilidad entre diferentes dispositivos empleando un enlace por radiofrecuencia en la banda de 2.4 GHz. Reconocido por Ericsson, Nokia, IBM, Toshiba e Intel. También se puede usar IPv6 con Bluetooth de baja potencia (BLE).

Para grandes redes de pequeños dispositivos que necesitan la supervisión o el control de un servidor de back-end en Internet, se puede usar MQTT, protocolo de Telemetría de cola de mensajes muy ligero, útil en redes con bajo ancho de banda y alta latencia o poco confiables, ofrece pocas opciones de control, es eficiente en términos de ancho de banda, independiente de los datos y tiene reconocimiento de sesión continua, no está diseñado para la transferencia de dispositivo a dispositivo, ni para realizar "multidifusión" de datos a muchos receptores.

En entorno industrial, los dispositivos de IoT usan el protocolo Data Distribution Service (DDS), estándar elegido por el Industrial Internet Consortium como plataforma de comunicación para construir sistemas inteligentes en Internet Industrial. En el ámbito de empresas de servicios (energía, gas, agua), se utiliza DNP3 (Distributed Network Protocol) protocolo industrial para comunicaciones entre equipos inteligentes (IED) y estaciones controladores, componentes de sistemas SCADA. DNP3 es actualmente compatible con las especificaciones del estándar de seguridad para infraestructuras de Sistemas de Información para la energía IEC 62351-5. El estándar IEEE 1379-2000 propone una guía para la comunicación entre RTU y dispositivos electrónicos inteligentes (IED) en entornos de energía, incluyendo cifrado y una serie de prácticas que mejoran la seguridad frente a métodos intrusivos conocidos. También está el estándar TC 57, responsable del desarrollo de estándares para el intercambio de sistemas de energía y otros relacionados como SCADA y distribución de automatización y teleprotección.

Para interconectar vehículos (Internet de los Vehículos IoV) se usa el protocolo de comunicaciones de corto rango dedicado (DSRC) que soporta transmisiones rápidas de mensajes entre vehículos, infraestructura y aplicaciones. Para comunicaciones internas de CPU a periféricos, existen protocolos como I2C (Inter-Integrated Circuit)

y SPI (Serial Peripheral Interface). En cualquier ámbito, a nivel de red, existen varios protocolos utilizados en IoT como IPv6 pensado para convertirse en el estándar de comunicaciones en Internet, o Thread [4] protocolo para redes inalámbrica, para interconectar dispositivos de bajo consumo de uso hogareño.

2.1 Arquitectura IoT

Existen múltiples propuestas de arquitecturas para IoT [5], algunas privadas creadas para un fin específico, y otras open source desarrolladas por un consorcio de empresas relacionadas con IoT que pretenden brindar esta arquitectura para uso general. Un ejemplo de una arquitectura global está conformada por el dispositivo final IoT sensor-actuador (“Things”) conectado (generalmente de modo inalámbrico) a otro dispositivo que permite la interconexión a Internet para brindar a un usuario, servicios de terceros o interactuar con servidores Cloud. Disponer de varias arquitecturas para Sistemas IoT, hace difícil la interoperabilidad de estos sistemas y resulta complejo brindar la seguridad y privacidad a los sistemas IoT.

OpenIoT es una plataforma open source [6] que incluye funcionalidades para componer dinámicamente y bajo demanda servicios IoT. Provee la posibilidad de recolectar y procesar los datos obtenidos de cualquier sensor, incluyendo dispositivos físicos, algoritmos de procesamiento de sensor, algoritmos de procesamiento de redes sociales, herramientas para describir los datos del sensor semánticamente según las especificaciones de W3C (Semántica para Redes de Sensores), envío de datos de diferentes sensores a la infraestructura en Cloud, descubrimiento dinámico de sensores y sus datos, visualización de datos IoT en formatos apropiados (gráficos, esquemas, mapas), optimización de recursos con el middleware OpenIoT y la infraestructura Cloud.

La arquitectura abierta de IoT V3.0 [7], generada por un proyecto europeo IoT-A y patrocinado por una gran cantidad de empresas europeas, incluye distintos modelos relacionados a las propiedades de confianza, seguridad, privacidad y confiabilidad.

3 Seguridad en IoT

En esta sección, presentamos un análisis y clasificación de los dispositivos utilizados en IoT y sus vulnerabilidades. Actualmente, es necesario repensar enfoques, ajustar herramientas, métodos y procedimientos para mejorar la seguridad en sistemas IoT.

La diversidad de hardware, de software y de protocolos de comunicación, implica mayor análisis y compromete en mayor medida la seguridad en los sistemas y dispositivos IoT. Muchos dispositivos IoT tienen sistemas operativos portables sin configurar, a menudo con un conjunto de utilidades utilizadas en el ámbito de desarrollo, pero que no deberían estar disponibles en sistemas de producción, por ejemplo, acceso por Shell. En muchos desarrollos de sistemas IoT se usan componentes de software de tercero (como librerías), que pueden incorporar vulnerabilidades. La violación a la seguridad de los dispositivos IoT, puede radicar en: (1) Los dispositivos IoT como víctimas y (2) los dispositivos IoT como herramientas de ataques.

Actualmente, surgen interrogantes sobre quien tiene la responsabilidad de resolver los problemas de seguridad de los dispositivos de IoT conectados a Internet ¿le corresponde al fabricante, al vendedor, al usuario?, ¿debe ser regulado por los gobiernos? No siempre queda claro quién es responsable de las decisiones de seguridad: una compañía diseña un dispositivo, otra provee el software, otra opera la red en la que se lo integra y otra pone a disposición el equipo. No existen normas y estándares aceptados a nivel internacional.

Un estudio [2], reveló que el 94% de los expertos considera firmemente que los dispositivos del IoT que no estén debidamente protegidos podrían detonar un incidente de seguridad "catastrófico". El 76% piensa que cualquier momento en los próximos dos años se producirá un ataque DDoS (Distributed Denial of Service) a través de IoT. El 77% admite que no considera el riesgo de IoT al desplegarlos, ya que delega este tema en las terceras partes involucradas. Además, el 67% no evalúa las prácticas seguridad y privacidad de las terceras partes antes de generar una operación de negocios con ellos. Con respecto al esfuerzo de seguridad, el 94% lo direcciona a un firewall de red tradicional para manejar las amenazas de IoT.

3.1 Amenazas en IoT

En el informe de la empresa Karsperky de junio del 2017, indica que han detectado 7.000 tipos distintos de malware que atacan a dispositivos IoT, de los cuales el 50% han sido desarrollados en 2017 y tienen como objetivo el espionaje, extorsión y chantaje [8]. El malware Mirai, tiene tres componentes principales: un módulo de comando y control que establece la comunicación; un escáner de redes, que permite infectar otros dispositivos de la IoT desde un equipo de retransmisión; y un módulo de ataque, que permite hacer uso y abuso del tráfico legítimo de las redes.. El malware Hajime [9] surgió para evitar que el dispositivo sea infectado por el malware Mirai cerrando los puertos utilizados en los ataques de DDoS. Tiene un módulo de propagación, se expande usando redes P2P descentralizadas (en vez de servidores de control y comando) y utiliza diferentes técnicas para ingresar al dispositivo e infectarlo. Si bien surgió como un software con fines benéficos para la seguridad informática, en abril del 2017 se conocieron ataques empleando este malware.

La Tabla 1 muestra algunas amenazas del procesamiento, comunicación y almacenamiento de la información, presente en cualquiera de las categorías de dispositivos IoT [1] [5].

Tabla 1: Clasificación de las amenazas en IoT

Clasificación	Descripción
Suplantación de identidad	Un atacante obtiene la clave del dispositivo, a nivel hardware o software.
Denegación de servicio	Dejar sin funcionar un dispositivo, impedir la comunicación, o brindar un servicio.
Manipulación de la información	Manipulación o reemplazo de los componentes del sistema operativo. Lectura de datos desde el almacenamiento de información. Manipulación de datos de telemetría, de datos de control de comandos en cola o en la memoria caché, de paquetes de actualización de configuración o firmware

	durante el almacenamiento en caché o en la cola local.
Divulgación de información	Cuando un dispositivo ejecuta un software manipulado que proporciona datos a partes no autorizadas.
Elevación de privilegios	Se puede forzar a un dispositivo que realiza una función específica a realizar otra función.
Suplantación, revelación de información	Los dispositivos tienen a menudo funciones de seguridad, como PIN o contraseña, o se basan totalmente en confiar en la red, lo que significa que concederán acceso a la información cuando un dispositivo se encuentre en la misma red y con frecuencia dicha red sólo esté protegida por una clave compartida.
Alteración	Un atacante puede interceptar o invalidar parcialmente la difusión y enviar información falsa.

A continuación, se muestran los riesgos por categorías de dispositivos IoT. [5]

Dispositivos IoT de Consumidores: Obtener acceso al dispositivo para beneficio personal. Acosadores puede usar la información de localización. Identificación de patrones. A modo de ejemplo se mencionan amenazas detectadas: a) En junio 2016 [10] se detectó un malware tipo ransomware conocido como “flocker” que afectó a Smart TV LG con S.O. Android, con el objetivo de obtener datos, encriptarlos y bloquear la pantalla para pedir “rescate” de dinero para su desbloqueo. b) En Marzo del 2017 [11], la empresa Oneconsult ha logrado emitir un malware capaz de tomar el control del televisor emitido por la frecuencia de TDT (Televisión digital Terrestre). Esto no solo afectaría a los TVs sino a cualquier dispositivo en el rango de acción de la antena. Probado en Smart TV Samsung. c) La empresa Slashgear [12] reporta una vulnerabilidad presente en varios Smart TV's que podría permitir a atacantes vulnerar cientos de dispositivos al mismo tiempo.

Dispositivos IoT para la Salud: Violación a la vida (causar daño a pacientes VIP). Pérdida de privacidad (brindar información confidencial de los pacientes). Diagnóstico no confiable debido a fallas en software o hardware. Modificación de información, generando toma de decisiones nocivas para la salud. Efectos nocivos para el cuerpo humano por ingesta incorrecta de medicamentos

Dispositivos utilizados en Infraestructura Industrial: Destrucción o daño físico causado por el ciber o eco-terrorismo. Interrupción de operaciones por broma o “hacktivismo”. Robo de productos a ser entregados por el dron y/o robo del propio dispositivo en sistemas aéreos sin tripulación.

Dispositivos IoT utilizados en Smart Cities. Daños o destrucción física causados por Ciber-terrorismo.

3.2 Vulnerabilidades

Existe un conjunto amplio de vulnerabilidades en los dispositivos IoT [13] como son: Comunicaciones locales y remotas sin encriptar, almacenamiento sin cifrar, acceso remoto por Shell, cuentas ocultas, acceso a través de las interfaces UART (Universal Asynchronous receiver transmitter) y JTAG, entre otros.

Las redes radio cognitivas [14], cuyo acceso al espectro es dinámico, son extremadamente vulnerables a ataques maliciosos, particularmente porque los

usuarios secundarios no son propietarios del espectro. Al tener una red con alto dinamismo se hace difícil aplicar medidas de seguridad, por lo que los métodos para asegurar un espectro compartido son críticos. La amenaza más crítica es impedir comunicaciones de usuarios secundarios.

A continuación, se mencionan algunas vulnerabilidades reportadas por distintas empresas del ámbito de seguridad Informática.

Vulnerabilidades de monitores de bebé de diferentes marcas, detectadas por la empresa Rapid7 [13] en septiembre del 2015. Presenta problemas de acceso físico al dispositivo, acceso desde una LAN y desde Internet, por ejemplo: a) La vulnerabilidad CVE-2015-2886 permite que cualquier usuario autenticado en el sitio ibabycloud.com puede observar la cámara y los registros de video de otro usuario debido a la vulnerabilidad de referencia de objetos directos. b) La vulnerabilidad CVE-2015-2887, muestra cómo se puede acceder por telnet o UART al dispositivo accediendo con usuario común (admin) hardcodedos en su código. c) La vulnerabilidad CVE-2015-2883, muestra como el servicio web para crear sesiones remotas de streaming es vulnerables a ataques XSS y Session hijacking.

Vulnerabilidades en productos wearable que permiten a las personas estar conectados en todo momento. La empresa Open Effect [15] realizó una investigación mostrando cómo algunas empresas fabricantes de estos productos no tienen en cuenta la privacidad, permitiendo obtener información del dispositivo.

Vulnerabilidades en dispositivos de salud [16]: en bombas de infusión se encuentran fallas de ausencia de autenticación para las sesiones telnet, almacenamiento de claves de acceso en texto plano, trabajan con versiones de servidor web vulnerables, código de credenciales asociados a FTP hardcodedos, los dispositivos cardiacos implantables no tienen implementado una lista de comandos autorizados. En lo que respecta a la autenticación, 3 de 4 fabricantes de dispositivos de monitoreo cardiaco hogareños, tienen hardcodedas las credenciales que permiten autenticar los pacientes en la red.

Vulnerabilidades detectadas en los televisores Smart TV: En diciembre del 2012, detectaron una falla en Smart TV Samsung led 3D, que permite obtener acceso al TV de forma remota y así modificar archivos y modificar la configuración del control remoto. En julio del 2013, se detectó la vulnerabilidad (CVE-2013-4890) en el servidor web de Smart TV Samsung PS50C7700, que permite ejecutar un exploit para resetear el televisor utilizados en un ataque DoS.

Vulnerabilidades en Cámaras IP: En junio 2017, un estudio generado por empresa de seguridad F-Secure, encontró problemas de seguridad en cámaras IP de varias empresas.

3.3 Ataques sobre Dispositivos IoT

Los dispositivos IoT pueden ser atacados por ciberdelincuentes por distintos motivos: entretenimiento, obtener información confidencial, y cibertales utilizando el poder computacional y de comunicaciones del dispositivo y así poder conformar una botnet para atacar a un objetivo específico y dejarlo sin servicio, lo que se conoce como Ataque de Denegación de Servicio Distribuido (DDoS). A continuación se darán algunos ejemplos de estos tipos de ataque.

Ataque de Denegación de Servicio Distribuido usando dispositivos IoT [1]

En julio de 2015, la revista Wired reveló que piratas informáticos habían alterado remotamente la conducción de vehículos Jeep Cherokee. Fiat Chrysler Automóviles NV, tomó medidas de seguridad a nivel de red para evitar este tipo de manipulación remota. También programó una campaña de recuperación preventiva de 1,4 millones de automóviles y camiones equipados con radios vulnerables en los EE. UU.

El 23 de diciembre de 2015, la distribución de energía en Ucrania fue afectada por un ataque que interrumpió el servicio de una gran cantidad de usuarios durante varias horas. Los piratas informáticos usaron el troyano BlackEnergy para acceder al sistema de gestión de la distribución de energía y fueron así capaces de emitir comandos de interrupción del servicio, borrar y sobrescribir datos del sistema y realizar operaciones de apagado

En octubre del 2016, se produjo el mayor ataque de DDoS de la historia contra la compañía de hosting francesa OVH y el proveedor de DNS estadounidense Dyn, que forzó la desconexión de más de 100 sitios web (Twitter, Spotify, Netflix, Amazon, GitHub, PayPal, etc) por varias horas. El ataque se produjo utilizando una botnet conformada por dispositivos IoT infectados con el malware Mirai. Se realizaron tres ataques que involucraron a más de 10 millones de direcciones IP que generaron un tráfico superior a 1 Tbit/s.

Ataques a dispositivos para obtención de información confidencial

La empresa VTech anunció en diciembre del 2015 que sufrió una violación de seguridad exponiendo datos personales de 12 millones de personas. El ataque fue explotando una vulnerabilidad de SQL injection y servicios de registración de usuario sin encriptar (TLS) en dispositivos de juegos electrónicos de enseñanza.

En agosto del 2014 [17], el termostato inteligente Nest de Google, dispositivo que ayuda a aprender sobre los hábitos de calefacción y refrigeración sufrió un ataque que demostró la posibilidad de obtener el control total de Nest en segundos.

3.4 Modelos de seguridad actuales

En esta sección se mencionan algunas propuestas para brindar seguridad a sistemas IoT ya desplegados y se describen los modelos de seguridad propuestos en algunos de las arquitecturas IoT analizadas en las secciones anteriores.

Diferentes organismos gubernamentales han generado lineamientos para que los desarrolladores y usuarios introduzcan la seguridad dentro de la arquitectura. El departamento de Seguridad Nacional de Estados Unidos [18], presenta seis principios estratégicos para asegurar IoT con el objetivo de informar a usuarios, operadores y fabricantes para que tomen decisiones conscientes al trabajar con dispositivos conectados. Estos principios incluyen: (a) Incorporar la seguridad en la fase de diseño. (b) Actualizaciones de seguridad y gestión de vulnerabilidades de avanzada. (c) Construir sobre prácticas de seguridad probadas, implementar la defensa en capas, compartir información relacionada a incidentes y vulnerabilidades. (d) Priorizar las medidas de seguridad según el impacto potencial. (e) Promover la transparencia de IoT. Abarcar a todos los componentes e incluir a desarrolladores y fabricantes. (f) Proponer la implementación de “conexiones intencionales” y selectivas.

Arquitectura de seguridad para IOT.

La arquitectura de seguridad sistemática (IPM) [19], integra conciencia e interacción del mundo real, cibernético, y social dentro del modelo U2IoT, arquitectura típica de IoT compuesta por unidades IoT (redes y sensores IoT, nodos de control distribuido, gestión y centralización de centros de datos) y dispositivos IoT ubicuos (IIoT, local IoT, global IoT, nacional IoT). La arquitectura considera tres aspectos: a) El modelo de seguridad de información considera los requerimientos de seguridad básica y avanzada para tratar los sensores, red, aplicación y atribuciones sociales. b) La seguridad física incluye la infraestructura, ubicación y estado de la entidad, información de trazabilidad y condiciones del mundo real. c) La gestión de seguridad provee recomendaciones estratégicas por jerarquías clasificando escenarios con racionalidad y compatibilidad.

La arquitectura abierta de IoT (IoT-A) [7], incluye distintos modelos relacionados a las propiedades de confianza, seguridad, privacidad y confiabilidad. El modelo de confianza provee integridad de datos y confidencialidad, autenticación en puntos finales y no repudio entre dos entidades que interactúan. Incluye a) dominios de confianza, b) mecanismos de evaluación de confianza, basado en analizar información almacenada previamente del sujeto, c) políticas de comportamiento que regulan la manera en que dos sujetos con el mismo dominio de confianza interactúan; d) definición de niveles de confianza (local, global o centralizado); e) disponer de una federación de confianza que delimite las reglas de las relaciones de confianza entre los sistemas y los diferentes modelos de confianza; f) soportar M2M (Machine to machine) que permite interactuar máquinas autónomas evaluando la integridad de cada una.

Microsoft propone un modelo de seguridad [20] basándose en el modelo de riesgos. Consta de los siguientes pasos: a) Modelar la aplicación, b) Enumerar las amenazas, c) Mitigar las amenazas d) Validar las mitigaciones. Este modelo de riesgos de IoT es el utilizado en la arquitectura de referencia de IoT Microsoft Azure.

4 Conclusiones y Trabajo Futuro

En este trabajo presentamos un análisis de los aspectos de seguridad relevantes para dispositivos IoT. Conocer la forma en que un atacante podría poner en peligro un sistema ayuda a tomar las medidas pertinentes desde el principio, por lo cual es de suma importancia conocer sobre la seguridad de los sistemas y dispositivos utilizados en Internet de las cosas, conociendo cuales son las amenazas, vulnerabilidades y ataques presentes en este ámbito.

Como trabajo futuro, se planea investigar y proponer un modelo de seguridad y privacidad para el diseño de los dispositivos IoT y la implantación de los mismos dentro de sistemas IoT. Considerando la integración de los diferentes protocolos de comunicación, considerando los distintos niveles o capas del sistema de comunicación y de los distintos elementos que forman parte de una arquitectura IoT, para obtener así un sistema IoT seguro. Este modelo debe incluir las propiedades de confianza, seguridad, privacidad y confiabilidad, brindando integridad de datos, confidencialidad, autenticación, disponibilidad y no repudio. Para lo cual es

importante conocer las amenazas a las que puede estar expuesto y agregar las defensas adecuadas durante el diseño de la arquitectura. Es especialmente importante diseñar productos que consideren desde el principio, un modelo de seguridad integral.

5 Referencias

1. Creación de un mundo IoT fiable y gestionado. Instituto Nacional de Ciberseguridad (INCIBE). España. (2017).
2. C. Forrest: <http://www.techrepublic.com/article/94-believe-unsecured-iot-devices-could-lead-to-catastrophic-cybersecurity-attack/>. (17-05-2017)
3. Open Interconnect Consortium. <https://openconnectivity.org/>.
4. Thread Protocol. Threads group. <http://threadgroup.org/news-events/press-releases/ID/20/Introducing-Thread-A-New-Wireless-Networking-Protocol-for-the-Home>.
5. Future-proofing the Connected World. Cloud Security Alliance (CSA). Report. (2016).
6. IoT Open Plataforms. <http://open-platforms.eu/library/openiot-the-open-source-internet-of-things/>. (05-11-2014)
7. M. Bauer, M. Boussard, N. Bui, F. Carrez, C. Jardak y J. D. Loof: Internet of Things – Architecture IoT-A – Final architectural reference model for the IoT v3.0. Report (2013).
8. Trampas para el Internet de las cosas. <https://securelist.lat/honeypots-and-the-internet-of-things/85165/>. (19-06-2017)
9. Hajime, the mysterious evolving botnet. <https://securelist.com/hajime-the-mysterious-evolving-botnet/78160/>. (25-04-2017)
10. E. Duan: <http://blog.trendmicro.com/trendlabs-security-intelligence/flocker-ransomware-crosses-smart-tv/> (14-1-2016)
11. D. Goodin: <https://arstechnica.com/security/2017/03/smart-tv-hack-embeds-attack-code-into-broadcast-signal-no-access-required/> (31-03-2017)
12. R. Waugh: <https://www.welivesecurity.com/la-es/2014/06/10/peligro-smart-tv-vulnerabilidad-ataques-en-masa/> (10-06-2014)
13. T. B. Mark Stanislav: HACKING IoT: A Case Study on Baby Monitor Exposures and Vulnerabilities. Rapid 7. <https://goo.gl/Uh7y4e> (Septiembre 2015)
14. S. Kim: Cognitive Radio Anti-Jamming Scheme for Security Provisioning IoT Communications. (2015).
15. Every Step You Fake A Comparative Analysis of Fitness Tracker Privacy and Security. Open Effect. (2016).
16. B. Rios y J. Butts: Security Evaluation of the Implantable Cardiac Device Ecosystem Architecture and Implementation Interdependencies. WhiteScope. (2017).
17. Y. Jin, G. Hernandez y D. Buentello: <https://www.blackhat.com/us-14/briefings.html#smart-nest-thermostat-a-smart-spy-in-your-home> (2014)
18. Strategic Principles for securing the Internet Of Things (IoT). U.S. Department of Homeland Security. Version 1.0. (15-11-2016).
19. H. L. Huansheng Ning: Cyber-Physical-Social Based Security Architecture for Future Internet of Things. Scientific Research, (2012).
20. Microsoft.com. <https://docs.microsoft.com/es-es/azure/iot-suite/iot-security-architecture>. (2017)
21. Privacy and Security in a Connected Life: A Study of US, European and Japanese Consumers. Ponemon Institute LLC. (Mazo 2015).

Buenas prácticas para la recolección de la evidencia digital en la Argentina

Nicolas Armilla¹, Marisa Panizzi¹, Jorge Eterovic¹, Luis Torres¹.

¹ 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
nicolasarmilla@hotmail.com, marisapanizzi@outlook.com, jorge_eterovic@yahoo.com.ar,
torreslu@ar.ibm.com

Resumen. En la República Argentina se evidenciaba la ausencia de un manual, un procedimiento o de un código sobre la recolección de la evidencia digital. Esto conlleva a que una gran cantidad de casos quedasen inconclusos y sin resolución, hasta la creación de la Guía de obtención, preservación y tratamiento de evidencia digital de la Procuración General de la Nación Argentina, en Marzo del año 2016. En este trabajo se realiza una revisión sistemática de guías y buenas prácticas de nivel internacional y nacional, identificando los aportes y áreas de vacancias. Se presenta un conjunto de buenas prácticas para la recolección de la evidencia digital en Argentina, logrando subsanar las áreas de vacancia identificadas. Para la correcta validación del conjunto de buenas prácticas, aplicaremos el conjunto a un caso de estudio aplicado a la realidad que consta de la recepción de un mail malicioso en una computadora de escritorio.

Palabras Clave. Informática forense, perito informático, evidencia digital, buenas prácticas, procedimientos en la informática forense.

1 Introducción

Se ha realizado una investigación exploratoria documental respecto a definiciones de informática forense, antecedentes actuales en el ámbito internacional y nacional.

Darahuge define la Informática Forense como el conjunto multidisciplinario de teorías, técnicas y métodos de análisis, que brindan soporte conceptual procedimental a la investigación de la prueba indiciaria informática [1][2].

Kovacich define la Informática Forense como la aplicación legal de métodos, protocolos y técnicas para obtener, analizar y preservar evidencia digital relevante a una situación en investigación [3].

Gómez define la Informática Forense como aplicación de técnicas científicas y analíticas especializadas a infraestructura tecnológica que permite identificar, preservar, analizar y presentar datos que sean válidos dentro de un proceso legal. O también lo define como una ciencia que busca reproducir científicamente con una metodología estricta de los hechos acontecidos y su correlación para determinar el grado de impacto, y posteriormente establecer en coordinación con otros entes intervinientes, mecanismos tendientes a evitar nuevamente su ocurrencia, que van desde el marco normativo hasta la utilización de mecanismos técnicos [4].

Listek en el Diario La Nación plantea que el Gobierno quiere normas claras para obtener pruebas digitales en los procesos judiciales [5].

La Procuración General de la Nación menciona que uno de los temas que puede tocarse desde ahora es el relativo a la evidencia digital, ya que su adecuada obtención, conservación y tratamiento es un elemento clave, entre muchos otros, para asegurar el éxito de las investigaciones [6].

Luego de revisar los antecedentes en nuestro país, nos planteamos como problema de esta investigación la escasa maduración de procedimientos para la recolección de la evidencia digital en la informática forense en la República Argentina.

En los últimos años los peritos informáticos se basaron en procedimientos y buenas prácticas de otros países tales como Canadá, Estados Unidos, Reino Unido y Hong Kong. En la actualidad, a partir del año 2016 se cuenta con la nueva resolución de la Procuración General de la Nación [6].

Utilizando el método de revisiones sistemáticas de Argimón [7] se ha realizado una investigación documental sobre los procedimientos y buenas prácticas que se detallan a continuación:

- Guía de buenas prácticas para evidencia digital. [8].
- Computación Forense - Parte 2: Mejores Prácticas. [9].
- Guía para recolectar y archivar evidencia - RFC 3227. [10].
- Investigación en la escena del crimen electrónico. [11].
- Guía de obtención, preservación y tratamiento de evidencia digital. [6].

Se ha detectado que el inconveniente de basarse en procedimientos y buenas prácticas de otros países presenta diferencia de factores tecnológicos, sociales, culturales y legales respecto a los de nuestro país.

En este contexto, este artículo presenta las guías y buenas prácticas consideradas para el análisis comparativo (Sección 2), se presenta el conjunto de buenas prácticas para la recolección de la evidencia digital (Sección 3), se presenta un caso de estudio para la validación de la propuesta de solución (Sección 4) y se formulan conclusiones y futuras líneas de trabajo (Sección 5).

2 Guías y Buenas Prácticas consideradas

En esta sección se presentan las guías y buenas prácticas que se han contemplado para el estudio comparativo, la Guía de buenas prácticas para evidencia digital [8]

(sección 2.1), Computación Forense - Parte 2: Mejores Prácticas. [9] (sección 2.2), Guía para recolectar y archivar evidencia - RFC 3227. [10] (sección 2.3), Investigación en la escena del crimen electrónico [11] (sección 2.4) y Guía de obtención, preservación y tratamiento de evidencia digital. [6] (sección 2.5). Por último, se presenta el resultado del análisis realizado (sección 2.6).

2.1 Guía de buenas prácticas para evidencia digital.

La Asociación de Jefes de Policía (ACPO – Association of chief police officers) del Reino Unido mediante su departamento de crimen basado en informática, publicó esta guía en el año 2012 [8].

El propósito de este documento no es solo proveer una guía para asistir a la ley sino asistir a la investigación de la seguridad informática e informática forense tanto en escenas de crímenes como incidentes [8].

Es necesario aclarar que esta guía no pretende ser una receta desde A-Z de la forensia digital, o un manual de instrucciones específico que indique como realizar todas las tareas. Debe abarcar un cuadro general y proporciona una estructura subyacente a lo que se requiere en las Unidades Forenses Digitales [8].

2.2 Computación Forense - Parte 2: Mejores Prácticas

El ISFS, Information Security and Forensic Society (Sociedad de Seguridad Informática y Forense) creada en Hong Kong, publicó “Computación Forense - Parte 2: Mejores Prácticas” (Computer Forensics – Part 2: Best Practices). Esta guía cubre los procedimientos y otros requerimientos necesarios involucrados en el proceso forense de evidencia digital, desde el examen de la escena del crimen hasta la presentación de los reportes en la corte. Su estructura es:

- a) Introducción a la computación forense.
- b) Calidad en la computación forense.
- c) Evidencia digital.
- d) Recolección de Evidencia.
- e) Consideraciones legales (orientado a la legislación de Hong Kong).
- f) Anexos [9].

2.3 Guía para recolectar y archivar evidencia

El propósito de esta guía es proveer a los peritos un sistema de pautas sobre la recolección y archivo de evidencias digitales, para un incidente de seguridad determinado [10].

No hace falta insistir en que todos los peritos del sistema estrictamente deben seguir estas pautas cada vez que tienen un incidente de seguridad, lo importante es proveer una guía sobre que deberíamos hacer si ellos eligen recolectar y proteger la informaron relacionada con un intruso [10].

Dicha recolección representa un esfuerzo considerable por parte del perito. En los últimos años se han realizado grandes progresos para acelerar la reinstalación del Sistema Operativo y facilitar la reversión de un sistema a un estado “conocido” haciendo de este modo, la “opción fácil” aún más atractiva. Mientras tanto, poco se ha realizado para suministrar formas fáciles para archivar la evidencia (la opción difícil).

Además las capacidades de memoria y de disco en aumento y el uso más difundido de cautela y de tácticas de cubrir huellas por parte de los atacantes han exacerbado el problema [10].

Si la recolección de evidencia se realiza correctamente, es mucho más útil para aprehender al atacante y representa una oportunidad mucho mayor en ser admitida como hecho en un juicio [10].

Se deberían utilizar estas pautas como una base para formular los procedimientos de recolección de evidencia de sitio y se deberían incorporar los procedimientos de sitio en una documentación de manejo de incidente [10].

Una vez que se hayan formulado los procedimientos de recolección de evidencia de sitio, deberían tener la aplicación de la ley para tu jurisdicción confirmando que son adecuados [10].

2.4 Investigación en la escena del crimen electrónico

La investigación en la escena del crimen electrónico fue creada por el Departamento de Justicia de los Estados Unidos de América en 2001 [11].

Dentro de las principales responsabilidades se encuentran la de preservar la escena crimen electrónico, recolectar y resguardar la evidencia digital [11].

Se ocupa de las situaciones encontradas en la escena del crimen y evidencia electrónica digital [11].

Tratándose de pruebas digitales, los principios forenses y procesales generales se deberían aplicar en:

- El proceso de recolección, aseguramiento y transporte de pruebas digitales, las mismas no debieran cambiar.
- Las pruebas digitales sólo deberían ser examinadas por los entrenados expresamente con ese objetivo.
- Todo lo hecho durante el transporte y el almacenaje de pruebas digitales se debería documentar, conservarse y encontrarse disponibles para la revisión [11].

2.5 Guía de obtención, preservación y tratamiento de evidencia digital

La obtención, conservación y tratamiento de la evidencia digital es un elemento clave, entre muchos otros, para asegurar el éxito de las investigaciones, eje central de preocupación de la comunidad internacional para la investigación transfronteriza eficaz de estos delitos [6].

No pretende abarcar la totalidad de procedimientos a tener en cuenta, ni ahondar en cuestiones técnicas reservadas a los expertos en seguridad y en informática, sino

brindar recomendaciones utilizadas a nivel mundial para incautar, analizar y preservar evidencia digital que deben ser tenidas en cuenta por los operadores judiciales [6].

2.6 Dimensiones consideradas para el análisis

En esta sección se plantean las dimensiones que se han considerado para el análisis de las buenas prácticas y procedimientos a nivel nacional como internacional. Las dimensiones consideradas son:

- **Evaluación de escena:** El perito informático debe tomar medidas para garantizar la seguridad de todas las personas en el lugar de los hechos y para proteger la integridad de todas las pruebas, tanto tradicionales como electrónicas [11].
- **Herramientas y equipamientos:** Generalmente, y dependiendo del problema a analizar, se sugiere aplicar una combinación de herramientas, para asegurar la efectividad que se debe tener en estos casos donde la libertad de las personas puede estar comprometida, y ello podría depender del resultado de una pericia informática aplicando herramientas de forensia. [12].
- **Dispositivos electrónicos:** La mayoría de los dispositivos electrónicos referidos a informática forense se aplica a computadoras y dispositivos digitales en general. Pero también existen otros dispositivos que requieren consideraciones adicionales [9].
- **Recolección:** La recolección de la evidencia digital, como cualquier otra evidencia, debe manejarse cuidadosamente y de una manera que preserve su valor probatorio. Esto se refiere no sólo a la integridad física de un artículo o dispositivo, sino también a los datos electrónicos que contiene. Por lo tanto, ciertos tipos de pruebas informáticas requieren una recolección especial [11].
- **Almacenamiento y transporte:** Las acciones no deben agregar, modificar o destruir datos almacenados en una computadora u otros medios. Las computadoras son instrumentos electrónicos frágiles que son sensibles a la temperatura, humedad, choque físico, electricidad estática y fuentes magnéticas. Por lo tanto, se deben tomar precauciones especiales al empaquetar, transportar y almacenar evidencia electrónica [11].
- **Análisis:** El análisis forense digital se corresponde con un conjunto de técnicas destinadas a extraer información valiosa de dispositivos, sin alterar el estado de los mismos. Esto permite buscar datos que son conocidos previamente, tratando de encontrar un patrón o comportamiento determinado, o descubrir información que se encontraba oculta [1][2].
- **Reporte:** Los reportes son información escrita con una terminología determinada, haciendo referencia a los detalles específicos de un caso particular [9].

En la Tabla 1, se presenta el grado de cumplimiento de las dimensiones de análisis consideradas para análisis de las buenas prácticas y procedimientos a nivel nacional e internacional.

Tabla 1. Tabla Comparativa de buenas prácticas y procedimientos a nivel internacional y nacional

	Internacional				Nacional
	Guía de buenas prácticas para la evidencia digital (ACPO, 2012)	Computación Forense - Parte 2: Mejores Prácticas (ISFS, 2009)	Guía para recolectar y archivar evidencia – RFC3227 (RFC, 2002)	Investigación en la escena del crimen electrónico (NLJ, 2001)	Guía de obtención, preservación y tratamiento de evidencia digital (Procuración General de la Nación, 2016)
Evaluación de Escena		■		■	■
Herramientas y equipamientos		■	■	■	■
Dispositivos electrónicos				■	■
Recolección	■	■	■	■	■
Almacenamiento y transporte	■	■	■	■	■
Análisis	■				■
Reporte	■	■			

Los resultados de análisis a los cuales se ha arribado con la Tabla Comparativa de (Tabla 1.) permiten formular las siguientes conclusiones: ninguna de las guías y buenas prácticas analizadas custodian las dimensiones analizadas en su totalidad.

A partir de este análisis y de la identificación de las áreas de vacancias detectadas, se propone un conjunto de buenas prácticas para la recolección de la evidencia digital que considere todas las dimensiones analizadas (sección 3).

3 Propuesta del conjunto de buenas prácticas para la recolección de la evidencia digital en Argentina

El conjunto de buenas prácticas propuesto se denominará “Conjunto de Buenas Prácticas para la Recolección de la Evidencia Digital, cuya abreviatura será Co. Bu. P.R.E.D.A.

Este conjunto de buenas prácticas se compone de siete (7) etapas, en este contexto cada una de las etapas equivale a la dimensión de análisis planteada en la Tabla 1:

1. Evaluación de escena
2. Herramientas y equipamientos
3. Dispositivos electrónicos
4. Recolección
5. Almacenamiento y transporte

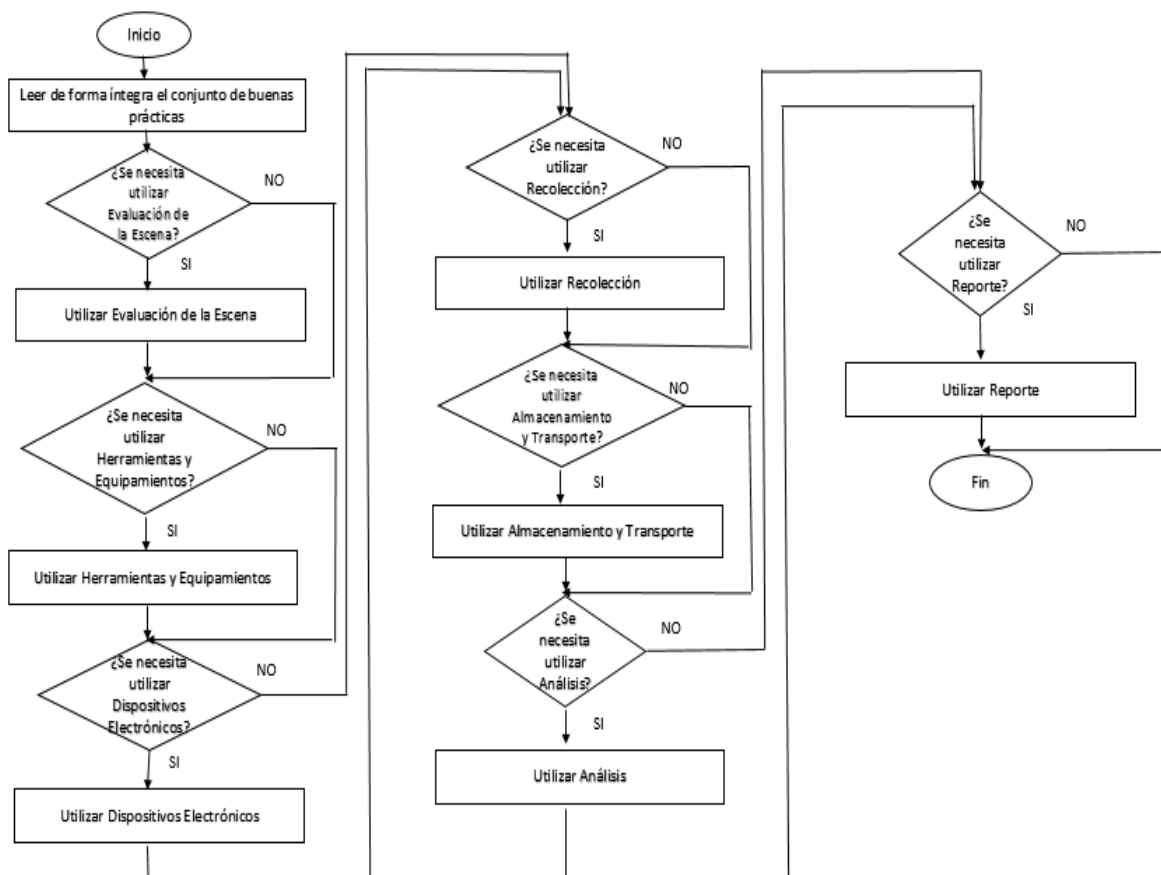
- 6. Análisis
- 7. Reporte

Cada etapa propone prácticas y herramientas, las prácticas hacen referencia a las actividades que tiene que realizar el perito en cada etapa y las herramientas refieren a los conocimientos, softwares y artefactos que necesitan para poder llevar a cabo las prácticas en cada etapa. Es importante aclarar que este conjunto de buenas prácticas contempla los factores tecnológicos, legales, sociales y culturales en Argentina.

Este conjunto de buenas prácticas está dirigido a los peritos con el propósito de facilitarles su tarea de recolección de la evidencia digital. Este conjunto puede ser utilizado en su totalidad o ajustándolo a la necesidad de cada uno de los escenarios planteados. Es recomendable que antes de utilizar un conjunto de buenas prácticas sea leído en su totalidad para comprender cuáles son los puntos que se adaptan al caso en particular y cuales no para luego aplicarlos de forma óptima.

En la Figura 1 se presenta un diagrama de flujo, el cual sintetiza la manera de utilización del conjunto de buenas prácticas propuesto.

Figura 2. Diagrama de flujo para utilizar Co. Bu. P.R.E.D.A



4 Caso de estudio para la validación de la propuesta

El caso de estudio seleccionado para validar la propuesta del conjunto de buenas prácticas consiste en la recepción de un mail con una supuesta amenaza que recibe un usuario final en una computadora de escritorio con sistema operativo Windows 10, la cual se encuentra conectada a la energía eléctrica.

La selección de este caso se debe a que es un problema común en el ámbito de la informática forense.

En la Tabla 2, se presenta la aplicación del Conjunto de Buenas Prácticas para la Recolección de la Evidencia Digital en la Argentina (Co. Bu. P.R.E.D.A.) para analizar la supuesta amenaza de un mail malicioso dentro de la computadora de escritorio.

De la aplicación de Co. Bu. P.R.E.D.A. en el caso de estudio (Tabla 2.) se arribó a la conclusión que ha sido necesario considerar las siete (7) etapas del conjunto de buenas prácticas junto con sus prácticas y herramientas.

5 Conclusiones y futuras líneas de trabajo

Se ha presentado una revisión sistemática de guías y buenas prácticas de recolección de evidencia digital que ha permitido identificar las vacancias de los procedimientos existentes.

Se ha logrado la construcción de un conjunto de buenas prácticas compuesta por siete (7) etapas contemplando prácticas y herramientas que se adaptan a los factores tecnológicos, sociales y culturales de Argentina.

Se ha logrado validar el conjunto de buenas prácticas en un caso de estudio.

Como futuras líneas de trabajo se identifican:

- La experimentación del Conjunto de Buenas Prácticas para la Recolección de la Evidencia Digital en Argentina (Co. Bu. P.R.E.D.A.) en casos relacionados a fraudes de telecomunicaciones, violencia doméstica, investigaciones referidas a estupefacientes, amenazas y/o acoso vía correo electrónico, homicidios, copia ilegal de software, abuso infantil y pornografía.
- Se evidencia un área de vacancia en las copias bit a bit del Sistema Operativo en dispositivos móviles (Android, IOS, Windows phone) y manuales de recolección de evidencia digital para los mismos.

Tabla 2. Aplicación de Co. Bu. P.R.E.D.A. en caso de estudio.

Etapas	Prácticas	Herramientas
Evaluación de la escena	-Preservación de la computadora -Aislar la computadora de personas ajenas a la investigación	-Cámara de fotos. -Cinta del lugar del crimen. -Guantes.
Herramientas y equipamientos	-Actualización sobre herramientas y equipamientos relacionados con la informática forense.	- Cámara de foto, cinta del lugar del crimen, guantes, instrumentos no magnéticos, bloc de notas, cajas de cartón, registros, etiquetas, marcador y bolso antiestático. -Conocimiento sobre Windows 10, hardware, periféricos, redes y seguridad informática. -Programas para hacer copia bit a bit, para examinar estado del sistema y para generar imágenes.
Dispositivos electrónicos	-Evitar pérdida de información volátil. -Interés principal por Windows 10, mails, historial de internet y logs.	-Conocimiento sobre Windows 10. -Conocimiento sobre hardware. -Conocimiento sobre periféricos.
Recolección	-Etiquetar, documentar, marcar, fotografiar, filmar y rotular la computadora. -Individualizar todos los cables de la computadora. -Verificar registro para ver si se eliminó algún dato.	-Guantes. -Instrumentos no magnéticos.
Almacenamiento y transporte	-Procedimiento de embalaje de computadora. -Procedimiento de transporte de computadora. -Procedimiento de almacenaje de computadora. -Cadena de la custodia.	-Bloc de notas. -Cajas de cartón. -Guantes. -Registros del inventario de pruebas. -Etiquetas adhesivas de pruebas. -Bolso antiestático. -Marcador permanente.
Análisis	-Copia bit a bit del sistema operativo Windows 10. -Estrategia forense (concentrarse en los mails maliciosos). -Situación de datos en listado de la evidencia digital.	-Conocimiento en Windows 10. -Programas para examinar procesos. -Programas para examinar el estado del sistema. -Programa para hacer copias bit a bit. -Programas para generar imágenes esenciales y para poder examinarlas.
Reporte	-Generación de reporte técnico. -Conclusiones alcanzadas.	-Conocimiento en Windows 10. -Conocimiento en seguridad informática. -Conocimiento en redes.

Referencias

1. Darahuge Maria Elena – Arellano González Luis Enrique, Manual de informática forense 1, Buenos Aires, 2011.
2. Darahuge Maria Elena – Arellano González Luis Enrique, Manual de informática forense 2, Buenos Aires, 2012.
3. Kovacich Gerald, High-Technology Crime Investigator's Handbook: Working in the Global Information Environment, United States of America, 2000.
4. Gómez Luis A., La informática forense: una herramienta para combatir la ciberdelincuencia, Buenos Aires, 2012.
5. Listek Vanesa, El gobierno quiere normas claras para obtener pruebas digitales en los procesos judiciales, Diario La Nación - Argentina, viernes 19 de agosto de 2016. <http://www.lanacion.com.ar/1929918-EL-GOBIERNO-QUIERE-NORMAS-CLARAS-PARA-OBTENER-PRUEBAS-DIGITALES-EN-LOS-PROCESOS-JUDICIALES>
6. Procuración General de la Nación, Guía de obtención, preservación y tratamiento de evidencia digital, publicada en la Resolución PGN-0756-2016-001, 31 de marzo de 2016.
7. Argimón J. 2004. Métodos de Investigación Clínica y Epidemiológica. Elsevier España, S.A. ISBN 9788481747096.
8. ACPO: Association of Chief Police Officers, Good Practice Guide for Digital Evidence, Reino Unido, 2012.
9. ISFS: Information Security and Forensic Society, Computación Forense – Parte 2: Mejores Prácticas, Hong Kong, 2009.
10. RFC: Request for Comments, RFC 3227: Guía para recolectar y archivar evidencia, 2002.
11. NIJ: National Institute of Justice, Electronic Crime Scene Investigation: A Guide for First Responders - Second Edition, Washington, 2001.
12. Piccirilli Dario, La forensia como herramienta en la pericia informática, Buenos Aires, 2013.
13. Piccirilli Dario. PROTOCOLOS A APLICAR EN LA FORENSIA INFORMÁTICA EN EL MARCO DE LAS NUEVAS TECNOLOGÍAS (PERICIA – FORENSIA y CIBERCRIMEN), La Plata – Prov. Buenos Aires, 2015.
14. ENFSI: European Network of Forensic Science Institutes, GUIDELINES FOR BEST PRACTICE IN THE FORENSIC EXAMINATION OF DIGITAL TECHNOLOGY, Europa, 2009.
15. Acurio Del Pino Santiago, Manual de Manejo de Evidencias Digitales y Entornos Informáticos. Versión 2.0, Ecuador, 2009.
16. Darahuge Maria Elena – Arellano González Luis Enrique, La cadena de custodia informático forense, Buenos Aires, 2016.
17. Darahuge Maria Elena – Arellano González Luis Enrique, Aplicaciones Creativas e Innovadoras en Informática. Desarrollos informáticos creativos e innovadores, Buenos Aires, 2016.

Efficient Votes Storage in a Non-Interactive Dining Cryptographers (NIDC) Protocol

Pablo García¹, Silvia Bast¹, Germán Montejano^{1,2}

1. FCEyN (UNLPam). Uruguay 151, Santa Rosa, La Pampa, Argentina
{pablogarcia, silviabast}@exactas.unlpam.edu.ar.

2. FCFMyN – (UNSL). Ejército de Los Andes 950, San Luis, Argentina
gmonte@unsl.edu.ar

Abstract - This paper shows the behavior of a storage technique for anonymous data, based on parallel channels implementation. At the beginning, it was conceived to apply on electronic vote. However, it may be generalized to whatever situation that requires anonymity and demands very high level of security respect from loss of information. Furthermore, a formula mistake that was slipped in previous publications has been corrected here. Finally, results of simulations are shown, to analyze the behavior of the equations proposed.

Key words: E-Voting - NIDC – Anonymity – Parallel Channels.

1. Introduction

There are many real-world problems that require very high levels of security with respect to information use. A good example of an application, whose level of demand in that sense is highest, is the electronic voting. In such an application, the condition of anonymity of a voter indefinitely is crucial. But indisputably it must be ensured that the result of the ballot accurately reflects the will of the electorate.

It is obvious that the interests involved are very transcendent and that, therefore, attempts to perform a fraud may occur with high probability. And if consider in detail, many aspects of the process, whose safety must be ensured, appear. In this case, the behavior of an alternative proposal for the storage of votes, based in the implementation of parallel channels of slots will be analyzed.

The results shown in this paper belong to a research line that began in 2013 [1]. Within this scope, it is searched to define the exact assurance level requested for anonymity in an electronic voting scheme. In [2] it is concluded that it is necessary to give unconditional security for the privacy, because it must be protected indefinitely. Otherwise, votes must be kept for a finite period of time.

Consequently, protocols that verify that condition reach most importance. Particularly one of the most interesting is Dining Cryptographers, which is described in detail in [3]. This protocol is resourceful and it covers the requirement to guarantee unconditional privacy.

The model may be described as follows:

“Three cryptographers share a dinner in a restaurant. When the time to pay comes, the waiter tells them that the addition has already been paid and that who did

it, does not want that his identity is revealed. Cryptographers want to know if any of the guests was the one who made the payment, or if it was paid by someone external to the group of diners. They only want to know whether any of them paid or not. ”

Raised in this way, the solution found is:

“Each of the diners throws a coin. He looks at the result and shares with its neighbor on the left. Then, each of them looks exactly two currencies, self and neighbor who shares with him. Finally, each one should indicate whether the two currencies that could be observed are ”equal” or ”different” with the condition that if any of them paid the addition, he should lie about his statement.”

In the conditions described, if the number of cryptographers proclaiming “different” is odd, the payer is in the group of diners. An even number, otherwise, indicates that the payer is external to the group.

Dining cryptographers presents unconditional security levels, with regard to anonymity associated with the issuance of certain information, through public channels. The initial problem is based on three participants exchanging only one bit of information, but it generalizes to any number of participants and any volume of information naturally and without significant complications.

It is necessary to highlight some points:

1) It is considered that the currencies used provide a truly random experiment result with respect to “throw the coin”, so that, $\Pr (“Heads”) = \Pr (“Tails”) = \frac{1}{2}$.

2) This scheme gives correct results only if it has the honesty in the response of all participants. If someone paid but he does not lie in his statement the model does not guarantee right conclusions. The same applies if a diner, who did not pay the bill, does not tell the truth when comparing both currencies.

3) This scheme works correctly if a single payment is made for dinner. If the waiter had accepted two or more anonymous payments, the conclusions will not be correct. This point relates to the above: the original model works properly only if all participants show an honest behavior.

4) If the required conditions are right, none of the diners get any information on the identity of the payer. As stated, if the payer is external, anonymity is assured. If the payer belongs to the group, it is easy to analyze cases to conclude that a cryptographer who did not pay the bill does not receive any information that allows deducing the identity of the payer. This property makes it very attractive scheme, since the central goal (anonymity) is obtained by default, as included in the scheme without additional effort.

5) One element of great value in Dining Cryptographers is the unconditional security given to the anonymity. A system which allows transmission of messages ensuring the unconditional anonymity of the source is provided. For any information exchange scheme in which privacy is desired, dining cryptographers becomes a very attractive scheme.

To understand the underlying reason why anonymity is ensured, it is necessary to put in place a diner who has not paid the bill (C). Obviously, the case in which the payer is external guarantees anonymity based on the assumptions of the scheme. Therefore, one must analyze the case where the payer belongs to the group of cryptographers. This can only have two alternatives:

- *The two currencies that C looks are equal.* In this case, necessarily, one of the other participants indicated “equal” and the other “different”.

Therefore, if the currency that he could not observe were equal to the one he saw, the participant who said "different" is the payer. And, conversely, if the unknown currency was different from which he could see, the payer is who expressed "equal". However, both states are equally probable in the scheme. Therefore, it is not possible to obtain information related to the payer.

- *The two currencies that C looks are different.* In this case, it is inevitable that the other two cryptographers express coincident results. If both say "different", the payer will be the participant who is closest to the currency matching the result of the hidden coin. And if both proclaim "equal", the payer is who is closer to the currency that differs from the value of the hidden coin. As in the previous case both situations are equally probable. Consequently, any information that betrays the identity of the payer may not be obtained.

To complete the analysis, the concept of view, which will allow us to demonstrate the safety of the proposed scheme, is defined:

A view is a random variable describing what set of information has a particular participant when the process is finished. For example, at the end of the election act if it is a scheme of electronic voting.

If we may prove that the view obtained by any user may not determine the choices made by the other participants in any case, anonymity is ensured. In particular, to analyze the scheme Dining Cryptographers, we may distinguish the following elements and analyze which ones are visible for each participant:

- *Coins:* according to the mechanics of the DC model, each participant sees its own currency and its neighbor's on the left. We denominate $x_i \in \{Tails, Heads\}$ to the value obtained in the action of the coin toss i .
- *Investment information:* The value of this element will be $m_i \in \{True, False\}$. If m_i is True, it implies that the participant paid the bill and, therefore, lies about the obtained result when comparing the two currencies that can be observed. A value False implies the opposite.
- *Information comparing two currencies:* For this data $r_i \in \{Equal, Different\}$ will be used. Obviously, a value Equal implies that the participant i declares that the values of the two currencies which he may see are coincident.

Therefore, in these terms, initial security scheme may be observed. It is sufficient to analyze the views that a particular participant has for all possible cases, since the symmetry of the scheme ensures that the conclusions may be generalized for all participants.

The views that the participant A_i could have available is then analyzed. The cases are:

- The payer is external. In this case, anonymity is guaranteed.
- A_1 is the payer. This situation is also trivial.
- A_2 is the payer. The view that A_1 has available is the following:

$$VI = (x_1, -, x_3, m_1, -, -, r_1, r_2, r_3) \quad (1)$$

- A_3 is the payer. In this situation, A_1 observes:

$$V_1 = (x_1, -, x_3, m_1, -, -, r_1, r_2, r_3) \quad (2)$$

It is clear that the views of the last two cases are coincident. Specifically, the values of x_1 , x_3 , m_1 and r_1 have the same probability distribution in both cases; r_2 and r_3 , instead, present opposing values depending on who paid the bill. However, that does not give additional information to A_1 , because such values depend on x_2 , value he does not know and has the same probability to take either of two possible values.

Therefore, if any of his colleagues paid dinner, A_1 cannot tell who it was, because:

$$\Pr(A_1) = \Pr(A_2) = \frac{1}{2} \quad (3)$$

2. Non - Interactive Dining Cryptographers (NIDC)

The analysis is focused on a derivative one called Non Interactive Dining Cryptographers (NIDC, [4]), that relaxes the condition of concurrency online for all participants, condition that is present in multiple real-world problems.

The idea is that, through the use of blind signatures introduced by Chaum [5], the voter obtains a valid vote of the authorities of the process election. In particular, the protocol presented in [6] could be used. This protocol allows the voter communicate with authorities to send a blind vote. They respond by signing (blindly) the vote and resending it to the voter. It should be noted that the process is perfect and that all options are mathematically equal, the reason explaining why authorities cannot deduct any information related to the voter's option.

Obviously, the authorities should record each vote, so that no voter may cast vote several times. Similarly, both parties must sign their messages and keep records of them for the purpose of solving any subsequent dispute.

Upon receiving the message from the authorities, the voter retrieves it and he can verify that it contains a valid and signed vote.

NIDC uses a storage model based on a single vector of slots. The anonymity is guaranteed by random position where a vote is stored. The fact of randomness brings the outcome of collisions. A collision occurs when two or more votes are stored in the same slot. That results in the loss of the coincident votes. In this context, the proposed model in Figure 1 may be explained by Birthday Paradox [7] which states:

"In a group of 23 people, the probability that there are at least 2 who share the same birthday is very close to $\frac{1}{2}$."

This assertion is little auspicious for the purposes of this research. Associated vector size is relatively large respect to the sample; however, the associated security level is far from what could be acceptable in practical applications. An E-voting system in which the probability of loosing at least one vote is next to $\frac{1}{2}$ lacks importance.

Graphically, the original scheme is shown in Figure 1.

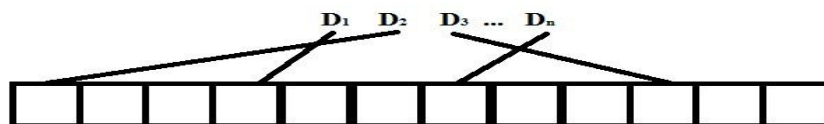


Fig1: Original Storage Scheme NIDC

The tendency of behavior of a scheme like Birthday paradox keeps little interesting values whatever the parameters were. The amount of positions that must maintain the vector to obtain acceptable security levels is very significant. For example in the typical case (365 days and 23 people) the probability of a collision approaches $\frac{1}{2}$ although, there exists 342 dates where nobody was born.

It is possible to think about a scheme that improves this redundancy in an efficient way. This alternative is described in the next section.

3. NEW PROPOSED STORAGE SCHEME TO NIDC

According to what is stated in the previous section, it becomes of great interest to analyze deeply the possibility to find alternative methods that improve the use of storage, since the single vector scheme requires significant number of positions in order to ensure an appropriate security level. In [8] an alternative proposal is exposed. It is shown in Figure 2.

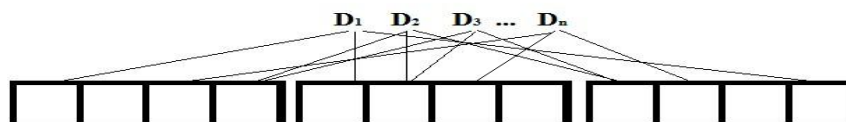


Fig 2: New Alternative For Storage.

In [9] another approach is shown, related to NIDC. It is based on parallel or serial channels. Conversely, this document goes forward to what was exposed in [10], where a series of equations are exposed. Those equations describe behavior of the model, based on the following parameters:

T : #total slots to implement.

S : #slots on each channel.

Q : #parallel channels to implement.

Q_i : #parallel channels to implement. (Theoretical).

Q_p : #parallel channels to implement. (Practical).

N : # voters.

V_i : i -th vote.

R_{ij} : event which indicates that V_i occupies j -th slot.

C_{ijk} : event that occurs when V_i collides with V_j in channel k .

B_{ij} : event when V_i loses in channel j .

A_i : event which indicates that at least a multiple collision is produced.

X : event which indicates that no vote is lost simultaneously in all channels.

L : #votes that are lost in all channels simultaneously.

I : event which indicates that at least a multiple collision is produced.

The equations exposed in [10] are the following:

For a fixed number of voters N , the recommended number of slots (S) for each parallel channel is given by the formula:

$$S = \left\lceil \frac{N}{\ln 2} \right\rceil + 1 \quad (4)$$

For given values of T and N , there exist an optimal number of parallel channels. Such value is expressed by:

$$Q_t = \ln 2 \frac{T}{N} \quad (5)$$

That formula should be the next integer.

$$Q_p = \lceil Q_t \rceil + 1 \quad (6)$$

As explained previously, it is rounded to the next integer, since the value of Q applied must be necessarily integer.

The expected value for the PLV variable (Percentage of Lost Votes) is obtained by applying equation:

$$|PLV| = (1 - e^{-\frac{N}{S}})^Q \quad (7)$$

An appropriate lower bound for the probability of “no votes are lost” is obtained by computing equation:

$$\Pr(X) > 1 - \left(\frac{1}{S}(N-1)\right)^Q \quad (8)$$

This last equation corrects the one published in [8] where an error was slipped. Indeed, the formula is correctly developed, as follows:

For $Q = 1$:

If V_1 drops in slot 1, the probability to collide with V_2 , consists on that both drop in slot 1:

$$\Pr(C_{121}|R_{11}) = \frac{1}{S} \frac{1}{S} = \frac{1}{S^2} \quad (9)$$

Then, the probability that votes V_1 and V_2 collide in whatever slot is:

$$\Pr(C_{121}) = \frac{1}{S^2} S = \frac{1}{S} \quad (10)$$

$$\Pr(C_{121}) = \Pr(C_{1j1}) \forall j \in \{3..N\} \quad (11)$$

Let $I =$ “Multiple Collisions occur”, where, “Multiple Collision” means that three or more votes are stored in the same slot. Consequently, the probability that V_1 is lost in the only channel is given by the expression:

$$\Pr(B_{11}) = \frac{1}{S}(N-1) - \Pr(I) \quad (12)$$

From the previous equation, another one, which is more appropriate is derived. This equation is based on the probability V_I is not lost.

$$Pr(\overline{B}) = 1 - \frac{1}{S}(N - 1) + Pr(I) \quad (13)$$

The value for $Pr(I)$ is low but positive, then:

$$Pr(\overline{B}) > 1 - \frac{1}{S}(N - 1) \quad (14)$$

The previous formula lets explain to a voter the probability that his vote is lost or not in a single channel scheme. The accuracy of the formula increases when $(S \rightarrow \infty) \wedge (N \rightarrow \infty)$.

When more channels are added, it can be asserted:

$$Pr(A_i) = Pr(B_i)^Q > \left(\frac{1}{S}(N - 1)\right)^Q \quad \forall i \in \{1 \dots N\} \quad (15)$$

Finally, an appropriate lower bound for the probability of $X = \text{"no vote is lost simultaneously in all channels"}$, is obtained applying the following equation:

$$Pr(X) = Pr(\overline{A_1}) \cap Pr(\overline{A_2}) \cap \dots \cap Pr(\overline{A_N}) \quad (16)$$

But:

$$Pr(\overline{A}) = Pr(\overline{A_2}) = \dots Pr(\overline{A_N}) \quad (17)$$

Moreover, they are self independent events, then:

$$Pr(X) = Pr(\overline{A})^N \quad (18)$$

Or what is the same:

$$Pr(X) = 1 - Pr(A)^N \quad (19)$$

Finally, it can be asserted:

$$Pr(X) > 1 - \left(\frac{1}{S}(N - 1)\right)^{Q \cdot N} \quad (20)$$

Precisely, at this point, the formula published in [8] has an error, consisting on the location of a parenthesis. The formula described above (20) is the correct form.

This lower bound is very useful to describe the probability of losing votes in a real voting with mentioned parameters. Definitely, it allows knowing previously the exact probability of no losing votes during the process.

4. SIMULATIONS

Given formulas above and knowing about error in formula ([8]), a simulator has been implemented which have two main aims:

- 1) To verify the correctness of formulas.
- 2) To bear out that the approach of storing in parallel channels optimizes the results in terms of several variables which can be considered.

The simulator is implemented allowing the following inputs:

- 1) Total number of slots to implement (T).
- 2) Number of voters (N).
- 3) Quantity of parallel channels to implement (Q).
- 4) Quantity of election acts that will be simulated by session (R).

The simulator verifies that the total number of slots (T) is a multiple of quantity of parallel channel, because the quantity of slots in each channel (S) must be an integer number.

When the simulation is complete, the following information can be obtained:

- 1) Total of successful votes (SV).
- 2) Total of lost votes (LV).
- 3) Quantity of runs where at least one vote is lost (R)
- 4) Quantity of runs (Votings) without lost votes (RWL)
- 5) Quantity of runs (Votings) with lost votes (RL)
- 6) Best case, that is to say, how many votes were lost in the most successful run (BC).
- 7) Worst case, that is to say, how many votes were lost in the less successful run (WC).

For all runs the following values for the parameters were selected:

$T = 480$ slots.	$N = 120$ voters.	$R = 1.000.000$
------------------	-------------------	-----------------

The successive simulations were executed with $Q = 1..5$. Q_t , using the appropriate formula, gives 2,772588722 channels. Consequently, it is expected that the best values will be obtained using $Q_p = 3$.

A. Verifying correctness of the formulas

Multiple simulations were executed to bear out the correctness of formulas (5) y (6). As it was explained previously, (5) indicates optimum theoretical value for Q_t y (6) the next integer. In all simulations the best results were obtained dividing t slots in Q_p parallel channels. For the purpose of illustrating the situation, figures 3, 4 are examples of the results obtained with selected values of T y N .

Hundreds of simulations were executed and in all cases optimum results occurs when Q_p is used.

Figure 3 shows percentage of lost votes for $Q = 1..5$. The lower loss occurs when $Q=3$.

Figure 4, shows quantity of successful votes. Also, in this case the best result occurs when $Q=3$.

B. Behavior of parallel channels technique.

At this point there are several facts that validate the method.

- 1) When 1.000.000 runs are fulfilled with a single channel, in the best

case, 3 votes were lost and in the worst case, 40. All the simulations with $Q=1..5$ give results. However best results occur with $Q= 3$ where in the worst case, 24 votes are lost but in the best case, no lost votes are registered.

2) Taking as a reference the simulations with $Q=1..5$ the worst results are obtained with $Q = 1$ in all the measured parameters.

Table I show the results obtained for runs of the simulator with $Q=1..5$, with $T=480$ slots, $N=120$ voters and $R=1.000.000$ repetitions. Similar conclusions were obtained for different values of parameters.

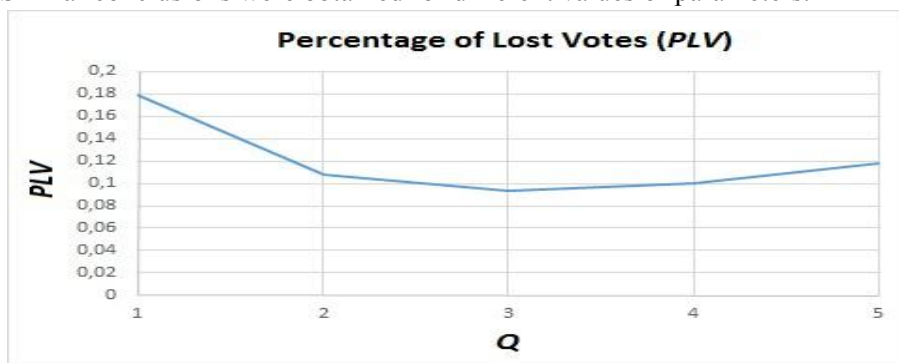


Fig 3: Percentage of Lost Votes

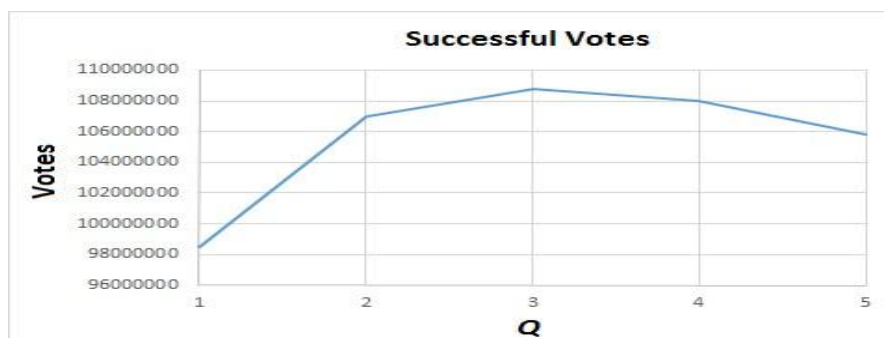


Fig. 4: Simulation Results: SV

Q	SV	LV	PLV	BC	WC
1	98291231	21508769	0,17924	2	40
2	107029947	12970053	0,108084	1	27
3	108763817	11236133	0,0936344	0	24
4	108001192	11998808	0,0999909	1	26
5	105794687	14205313	0,118378	1	30

Table 1: Results of Simulator's Runs

5. CONCLUSIONS

It is considered demonstrated that the proposed technique increases efficiency of storing anonymous data resulting in an advantageous variant respect from the use of a single array. This fact allows setting the security level for the desired value. The empirical results suggest that the behavior is higher in all aspects selected.

The quantity of votes that recovered successfully is optimized dividing T in Q_p slots. The quantity of runs without lost votes is highest when Q_p is used.

The best case (BC) optimizes with the use of Q_p . This means that the run with fewer lost votes occurs when Q_p is used. Similarly this occurs with the variable WC (Worst Case).

Grounds to suggest that the new proposal is a significant improvement for the storage of votes in a scheme Non - Interactive Dining cryptographers. The proposed formulas also allow an easy and accurate administration of the security levels to the system user.

References

- [1] Uzal R., van de Graaf J., Montejano G., Riesco D., García P.: "Inicio de la Línea de Investigación: Ingeniería de Software y Defensa Cibernética". XV Workshop de Investigadores en Ciencias de la Computación (WICC). Ps. 769 - 773. ISBN: 9789872817961.
- [2] van de Graaf J., Montejano G., García P.: "Optimización de un Protocolo Non-Interactive Dining Cryptographers". Congreso Nacional de Ingeniería Informática / Sistemas de Información CoNaIISI 2013. Córdoba, Argentina.
- [3] Chaum D.: "The Dining Cryptographers Problem: Unconditional Sender and Recipient Untraceability". Journal of Cryptology. 1988.
- [4] van de Graaf J.: "Anonymous One Time Broadcast Using Non Interactive Dining Cryptographer Nets with Applications to Voting". In: "Towards Trustworthy Elections". Ps 231-241. Springer-Verlag Berlin, Heidelberg. ISBN:978-3-642-12979-7. 2010.
- [5] Chaum D.: "Blind Signatures for Untraceable Payments". Advances in Cryptology Proceedings of Crypto 82 (3):199203.
- [6] Fujioka A., Okamoto T., Ohta K.: "A Practical Secret Voting Scheme for Large Scale Elections". AUSCRYPT 1992. LNCS, Vol. 718. Ps. 244 - 251. Springer Heidelberg. 1993.
- [7] 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.
- [8] van de Graaf J., Montejano G., García P.: "Optimización de un Esquema Occupancy Problem Orientado a E-Voting". Memorias del XV Workshop de Investigadores en Ciencias de la Computación 2013 (WICC 2013). Ps. 749 - 753. ISBN: 9789872817961. 2013.
- [9] García P., van de Graaf J., Hevia A., Viola A.: "Beating the Birthday Paradox in Dining Cryptographer Networks". The third International Conference on Cryptology and Information Security in Latin America, Latincrypt 2014. Florianopolis, Brasil. Lecture Notes in Computer Science, Springer (2014).
- [10] García P., van de Graaf J., Montejano G., Riesco D, Debnath N., Bast S.: "Storage Optimization for Non Interactive Dining Cryptographers (NIDC)". 12th International Conference on Information Technology: New Generations (ITNG 2015). April 13-15, 2015, Las Vegas, Nevada, USA.

Improving the Generation of Labeled Network Traffic Datasets Through Machine Learning Techniques.

Jorge Guerra¹ and Carlos Catania²

¹ ITIC-CONICET, Universidad Nacional de Cuyo (UNCuyo), Mendoza, Argentina

² Ingeniera, Universidad Nacional de Cuyo (UNCuyo), Mendoza, Argentina

jguerra@uncu.edu.ar

ccatania@itu.uncu.edu.ar

Abstract. The problem of detecting malicious behavior in network traffic has become an extremely difficult challenge for the security community. Consequently, several intelligence-based tools have been proposed to generate models capable of understanding the information traveling through the network and to help in the identification of suspicious connections as soon as possible. However, the lack of high-quality datasets has been one of the main obstacles in the developing of reliable intelligence-based tools. A well-labeled dataset is fundamental not only for the process of automatically learning models but also for testing its performance. Recently, RiskID emerged with the goal of providing to the network security community a collaborative tool for helping the labeling process. Through the use of visual and statistical techniques, RiskID facilitates to the user the generation of labeled datasets from real connections. In this article, we present a machine learning extension for RiskID, to help the user in the malware identification process. A preliminary study shows that as the size of labeled data increases, the use of machine learning models can be a valuable tool during the labeling process of future traffic connections.

Keywords: Machine Learning, dataset generation, network security

1 Introduction

In the field of network security research, intelligence-based detection approaches emerged as a tool for dealing with the fast evolution of the different network scenarios. Probably the most significant challenge during the developing of such systems is the lack of appropriate public datasets [9]. Just before deploying in any real world environment, an intelligence-based Network Intrusion Detection System (NIDS) must be trained and evaluated using real labeled network traffic traces with an intensive set of intrusions or attacks [5].

One of the reason behind behind the lack of public datasets arises from the data's sensitive nature. It is no secret to anyone that bringing to light network

traffic can reveal sensitive information. Such information is generally related to confidential and personal communications coming from organization business data or in other cases user private access behaviors. It is understandable that in the face of such high risks, researchers frequently encounter insurmountable organizational and legal barriers when they attempt to provide datasets to the community [9].

For the previous reason, the main strategy points to create synthetic or benchmark dataset to deal with the data confidentiality problem. Synthetic datasets are created to represent certain problem domains. Specifically to cover specific needs or certain condition [5]. On the other hand, benchmark datasets are often very useful but suffer excessive preprocessing that separates them from real network environments. Examples of known benchmark datasets are: KDD-cup99 [10] who was built upon the data captured in the DARPA98 IDS evaluation program, DEFCON [2] that contains network traffic captured during a hacker competition called "Capture The Flag", CAIDA dataset [1] that contain particular kind of attack, among others.

Other solutions point to real life datasets. These datasets are created generally capturing traffic from institutional networks. Recently, the Stratosphere Intrusion Prevention System (IPS) project [3] has emerged as a project focused on providing an state of the art IPS to the Non Governmental Organizations (NGO). One of the goals of the project consists of generating high quality datasets for testing and developing new malware detection techniques. The particular encoding of the network behavior used by stratosphere IPS project facilitates the release of network data to the community.

However, the problem of labeling all the published data remains a very difficult task. The labeling process not only requires a considerable human effort but also the responsible of labeling must be a security specialist, who could not be always available. The fact is that the difficulty behind the labeling process could be the real reason behind the lack of high quality real life datasets.

With these issues in mind, we developed RiskID [7]. Still in an early stage, RiskID aims at being a collaborative labeling tool based on visual analytics and statistical techniques. In particular, RiskID is based on the combinations of several visualizations strategies with clustering algorithms working together for facilitating the recognition of malicious traffic. In this paper, we propose an extension for RiskID based on machine learning techniques. The general idea behind using machine learning techniques inside RiskID consists of generating a classifier trained on the subset of already labeled connections and use classifier output for helping the user in the decision process. Here, the goal is just to provide to the user another tool for supporting his decision process.

The rest of the article is organized as follows: Section 2 describes the problem statement, details the strategy for generating the behavioral models proposed by the Stratosphere project and shows the general aspects of the RiskID tool. Then section 3 exposes the proposed machine learning extension for RiskID. Section 4 presents the experiments design and the performance evaluation of the proposed

extension for RiskID. Finally, concluding remarks and future works are described in section 5.

2 Problem Statement

The RiskID tool takes a dataset containing network traffic as input and uses visual representation techniques to help users to identify malicious behaviors. The main goal behind RiskID is to provide to the network security community labeled datasets. We believe that by including machine learning techniques inside RiskID, the effort of identifying malicious behavior can be reduced. As an extension of current RiskID labeling tool, we propose the inclusion of a Machine Learning classification algorithm for helping the user during the labeling process. The idea is simple: Based on the subset of already labeled SC connections, the classification algorithm can provide the probability that a given connection. Notice that the final decision will still be user responsibility. We proceed to discuss the main aspects of the current version of RiskID.

2.1 The RiskID Tool

As you can see in Figure 1, RiskID is a visual analytics tool that combines visualization with clustering techniques to assist the user in the process of labeling connections [7]. The goal is to obtain a real life data set with, as much as possible, labeled connections. Specifically, the application aim at labeling traffic generated by botnet attacks. The process for labeling in RiskID is the following:

1. RiskID receives a JSON file that contains for every SC (from here referred simply as connection) some basic network information, such as IP addresses and Ports, together with its corresponding behavioral encoding.
2. The feature extraction module analyzes all connection behavioral encodings and creates for every connection a new vector summarizing the information in terms of periodicity, duration and size.
3. The cluster composition module analyzes the feature vectors and groups them according to a standard similarity measure.
4. The UI represents the list of connections with a Heatmap of the feature vectors, using different colors for each type of feature.
5. The user explores the connection list to discover common patterns through color similarity. During this process she can select potentially similar connections.
6. Upon selection, details about the connection composition are shown in order to facilitate comparison.
7. Eventually, when the user finds a high coincidence between selected connections she can proceed to label them as "Botnet" or "Normal".

It is worth noting that a correct labeling process mainly depends on the user selection strategy.

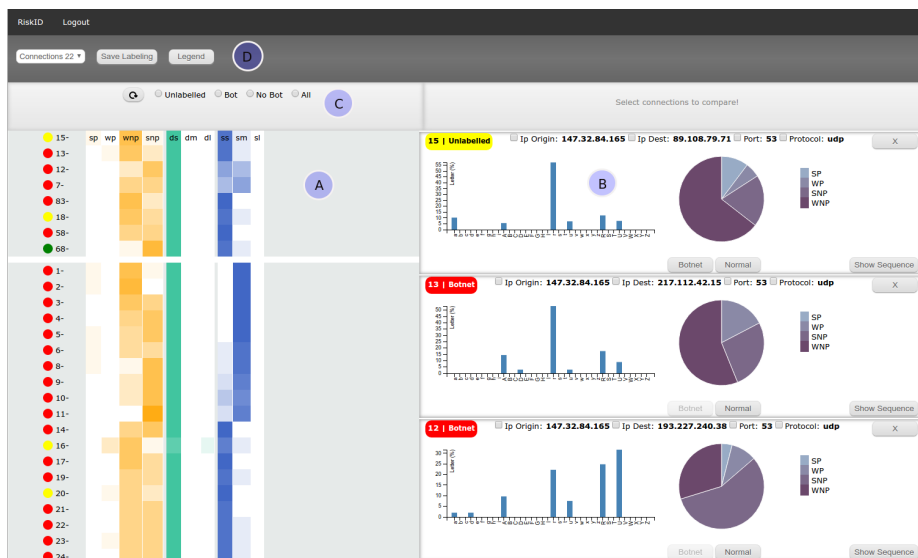


Fig. 1: RiskID User Interface. A: left panel collects descriptors for all connections with heatmaps of different colors for periodicity, size and duration features. B: right panel shows details for up to three connections. C: control panel to get a set of connections with similar filters selections. D: control panel to select Stratosphere datasets, get visual representation legends and save the labeled dataset resultant after connections labeling task by the user.

2.2 Behavioral Models

To deal with the confidentiality problem of network data, RiskID uses the encoding proposed by the Stratosphere IPS project. Such encoding only consider the size, duration and periodicity of network flows. The encoding start aggregating the flows according to a 4-tuple composed of: the source IP address, the destination IP address, the destination port and the protocol. They create *Stratosphere connection (SC)* putting together all the flows that match a tuple. From a traffic capture several of these SC are created. Each one of the these SC contains a group of flows. A sample **behavioral encoding** is shown in Fig. 2. The figure shows the symbols representing all the flows for a SC based on UDP protocol from IP address 10.0.2.103 to port 53 of IP address 8.8.8.8.

2.4.2*4.R.R*a*b*a*a*b*b*a*R.R*R.R*a*a*b*a*a*a*a*

Fig. 2: An example behavioral encoding of connection from IP address 10.0.2.103 to destination port 53 at IP address 8.8.8.8 using UDP.

RiskID as part of its data conversion generate for each connection a 10-dimensional numerical vector (denoted as feature vector) where the first four dimensions represent the periodicity (strong periodicity, weak periodicity, weak non periodicity and strong non periodicity respectively), the other three refer to duration (duration short, duration medium and duration large respectively) and the last three represent the size (size short, size medium, size large). The feature vector for a given connection is generated considering, for the complete symbol sequence, the cumulative frequency of the corresponding values associated with the behavioral encoding. At the end of the sequence, a percent of each feature is calculated and normalized between the values $[0,1]$. These final vector are used to create the heatmap visual representation where the intensity in the color scale indicates a given feature is predominant over the rest. To improve the heatmap the connections are organized making clusters using k-means algorithm [7]. For each connection we use the 10-dimensional feature vector and cluster ID to shape the training set.

2.3 Visualization Techniques

Main visualization tools used in RiskID consists of a heatmap representation supported by cluster strategies. Both components joined to filter options, histograms and pie graphs Figure 1(B) make up the bulk of the user interface. The heatmap representation is used for the list of Stratosphere connections Figure 1(A), that shows the connections grouped by the similarity in their encoding behavior. With the heatmap, it is intuitive to recognize the predominant features of each connection and, more importantly, relate connections with similar features. First, the connections are grouped by clusters. The clustering process helps the user get a first approximation of similar connections [7]. Once any connection is selected by a user the connection details section add this to the connection details list and show relevant information about it. A histogram graph displays their character distributions while a pie graph shows specifically their periodicity feature percent. One important advantage of RiskId is the possibility to compare two or more connections [7]. Each newly selected connection is placed under the previously selected one, and the details are stacked in the Detailed Connection View. Thus, the user can start a detailed comparison.

3 A Probability Estimator Based on Machine Learning Techniques

Machine Learning techniques are extensively used by malicious detection systems [9]. Their ability to learn with little data and then predict or detect similar behaviors make it a fundamental strategy in network detection field. In particular, we use a bagging strategies know like Random Forest (RF). RF is a general class of ensemble building method using a decision tree as the base classifier [8]. We use an RF to through the labeled connections to get a probability of botnet in unlabeled connections.

Once we get a botnet probability of whole unlabeled connections we help the user's decision for the next labeled. This information can be useful to represent in the connection list the output classification given by RF allowing the users get another evaluation criteria. We don't pretend to be determinants of the labeling process but yes be an influence in users decision. This new feature could reduce labeling time improving whole labeling task. We get first labeled connections by the RiskID users and create a training set to a machine learning algorithm and predict botnet behavior in remainder connections. Formally, suppose we have C like whole dataset and let C_r and C_t be the training set and test set respectively, we require $C_r \cap C_t = \emptyset$ and $C_r \cup C_t = C$.

4 Evaluation

For measuring the actual impact of the proposed machine learning extension, we need to consider not only a statistical evaluation but also the user interaction and confidence of the proposed extension. However, in this paper we will focus solely on the computational evaluation, leaving the evaluation with users for a later experiment. Specifically, we evaluate what is known as learning rate: the speed at which our extension learns new information or trends and updates the probability by connections accordingly [4]. We address the following questions:

1. Can random forest face at the beginning the disproportion between labeled and unlabeled and get a correct botnet probability for unlabeled connections?
2. How random forest behaves by types of connections?

The first question aims at the study of the classification performance of the proposed algorithm when it is trained with an small portion of labeled data and to predict the rest. Just like recommender system faces the cold start problem at first the model faces the challenge of giving a botnet probability from the few data labeled by users so far. Presumably, these probabilities will not be the best, but how well or poorly does the algorithm predict against this difficulty is the scenario we want to evaluate. For the second question we want to analyze the classification performance of the proposed algorithm considering the different type of traffic connection. It is clear that the traffic behavior of SMTP connections could be very different compared with HTTP. Such difference can certainly impact the in the performance of the learning rate of the classifier.

4.1 Dataset Description

For evaluating the performance of the proposed algorithm, we use the CTU-13 Dataset. The CTU-13 dataset consist of a group of thirteen different malware captures done in a real network environment taken from CVUT university campus networks. Datasets are publicly available as part of the Malware Capture Facility Project (MCFP) [6].

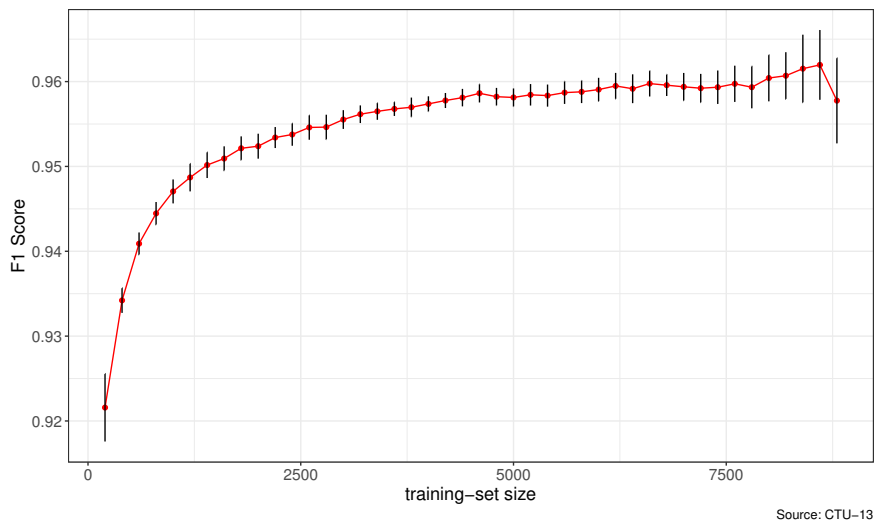
ID	IRC	SPAM	CF	PS	DDoS	FF	P2P	US	HTTP	Botnet	Conn.	Normal	Conn.	MCFP	IDs
A	X	X	X	-	-	-	-	-	-	-	911	415			CTU13-42
B	X	X	X	-	-	-	-	-	-	-	344	82			CTU13-43
C	X	-	-	X	-	-	-	X	-	-	13	292			CTU13-44
D	X	-	-	-	X	-	-	X	-	-	70	441			CTU13-45
E	-	X	-	X	-	-	-	-	X	-	31	147			CTU13-46
F	-	-	-	X	-	-	-	-	-	-	211	159			CTU13-47
H	-	-	-	X	-	-	-	-	-	-	29	135			CTU13-49
I	X	X	X	X	-	-	-	-	-	-	4000	513			CTU13-50
J	X	-	-	-	X	-	-	X	-	-	22	174			CTU13-51
K	X	-	-	-	X	-	-	X	-	-	2	21			CTU13-52
L	-	-	-	-	-	-	X	-	-	-	57	209			CTU13-53
M	-	X	-	X	-	-	-	-	X	-	479	231			CTU13-54

Table 1: Characteristic of botnet scenarios and general information about datasets.(CF: Click Fraud, PS: Port Scan, FF: Fast Flux, US: Compiled and controlled by us)

Table 1 provides brief information about each of the thirteen datasets. The first column shows the ID used for referencing the dataset. The next nine columns show the characteristics of the botnet scenarios. Then, in column ten and eleven, the number of connections labeled as botnet and normal. Finally, the last column shows the ID of the dataset in MCFP. For the purpose of the study, the three datasets were merged.

4.2 Experiment Design

During the training phase, we apply k-fold cross validation with 5 folds. To answer the first question, we trained the algorithm on different sizes portion of the dataset and tested on the remaining portion. In next iterations, we increase the number of data in the training set. This way, we simulate the use of the algorithm within the application over time. We started by taking only a random sample of 200 connections for the training set and tested with the remaining 8788 connections. Following this, we randomly took another 200 connections from the test set to add them to the training set. We perform this operation until there are approximately 200 connections in the test set. We use F1 Score to evaluate the RF performance at each iteration. This metric can be interpreted as a weighted average of the fraction of relevant instances among the retrieved instances (precision) and the fraction of relevant instances that have been retrieved over total relevant instances (recall). Each experimental scenario was simulated 30 times (i.e. 1320 simulations in total) to ensure the statistical robustness of results. As the final result, we use the mean value for F1 Score. To answer the second question we use the probabilities resultant of RF when was training with 2 percent, 50 percent, and 90 percent of the training set. These percentage values were selected to represent the initial, middle and final phase, respectively, of the labeling process. We perform a study of the algorithm behavior by types of connections. The connections were divided by port type, keeping only the most representative of all of them.



Source: CTU-13

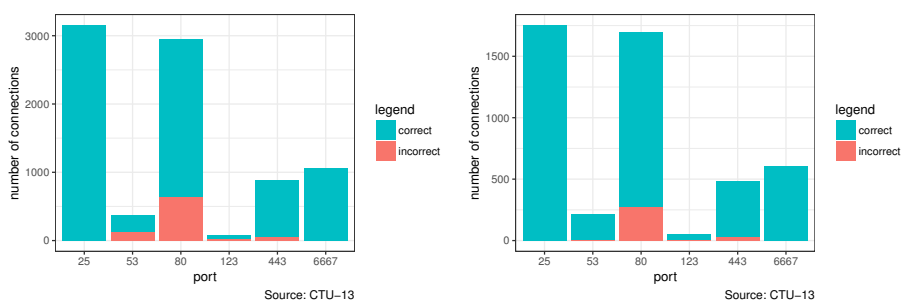
Fig. 3: Random Forest performance with incremental training data

4.3 Random Forest Performance with Incremental Training Data

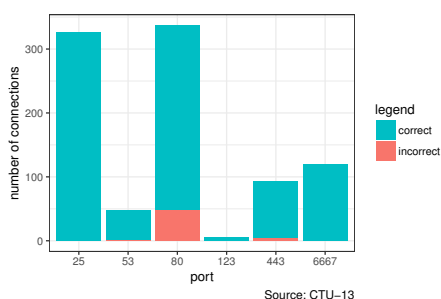
The idea behind this experiment is to analyze the impact of incrementing the number of data in training set. Specifically, we evaluate the F1 score metric by each iteration of RF. Figure 3 shows that as the number of data in the training set increases, the mean of F1 score improves. On the X-axis we have the different sizes in training set. Each point in the graph represents the number of times that RF is training and testing. On the Y-axis we have the mean of F1 score returned by RF in each iteration. When we training RF with the first 200 labeled connections (approximately 2 percent of the whole CTU-13 dataset) and then test with the remaining portion of the dataset we get a mean F1 Score about 0.92 value. This is a good first result but the F1 score still increases until 25 iterations when it stays oscillating close to 0.96 value. This result shows that approximately from the first 5000 (55 percent of the whole CTU-13 dataset) labeled connections, RF predicts with good results the rest of the connections.

4.4 Detection Performance Analysis of Random Forest by Type of Connections

In this section, we analyze the detection performance of RF by type of connections. Specifically we analyze the connections by ports. In the previous experiment, we evaluate RF with different training set size. The Figure 4 displays the proportion of connections correctly classified (show in color green) and incorrectly classified (show in color red) by the most representative ports when RF was training with 2 percent, 50 percent, and 90 percent of the training set. The information provided by Figure 4 exposes the fact that most of the traffic in CTU-13 come from 25 port. The port 25 refers to Simple Mail Transfer Protocol



(a) Prediction performance by port with 2 percent of the training set (b) Prediction performance by port with 50 percent of the training set.



(c) Prediction performance by port with 90 percent of the training set.

Fig. 4: Prediction performance by port.

(SMTP), used for email routing between mail servers. As can be seen, the RF detection model is able to detect the 100 percent of all the cases where port 25 is present. Even when the proposed algorithm was generated using only the 2 percent of the labeled data. This good performance happens because SMTP traffic may be very similar to each other and with just a few examples we can learn a lot. By the other hand the port 80 (refers to Hypertext Transfer Protocol "HTTP") present some detection error. This error is present due to the variability in the connections that share port 80. The remaining ports, although they have bad qualifications, most of the times connections are well identified.

5 Concluding Remarks and Future Work

In the present article we presented a machine based extension to RiskID [7], a tool for generating labeled network traffic datasets for research community. The preliminary study shows the viability of using a probability estimator generated from the subset of labeled data. To start using this extension inside RiskID tool, some connections need to be labeled previously. The Figure 3 has shown that

is not necessary much data for train our strategy and get good results. Once the new extension starts to suggest probability of botnet for each connection the users will have a new evaluation criterion. This favors the increase of connections that will be used to train and as we saw in Figure 3 this improve the detection performance. In this way, the new extension proposed in this article decreases the labeling time of the analyzed dataset. A similar analysis considering the type of connections displayed same results. Figure 4 showed decrease in error (bar with color red) when RF was trained with more data. In some cases, such as SMTP connections, the amount of needed labeled connections is considerable small (about 2 percent of the dataset). On the other hand, HTTP connections required a higher number of labels to reduce the classification errors. A user experience analyze is required to get a final evaluation for this new extension. We are working on some test to collect users experience and will be exposed in future works. Last observation is about the quality of training dataset, which determines that the proposed strategy be able to create a good prediction model. RiskID users could label connections with some errors and this could impact in RF results. Noise robustness is a very important problem in machine learning algorithms and our strategy don't escape this. The quality of first labels created by RiskID users is crucial for a good future prediction of our RF algorithm. Feature analysis about noise robustness of our new extension proposed for RiskID will be accomplished in next works.

References

1. Center for applied internet data analysis. <http://www.caida.org/>, October 2011. [Online; accessed May-2017].
2. The shmoo group. <http://cctf.shmoo.com/>, October 2011. [Online; accessed October-2016].
3. Stratosphere ips project. <https://stratosphereips.org/>, October 2015. [Online; accessed Jun-2017].
4. Iman Avazpour, Teerat Pitakrat, Lars Grunske, and John Grundy. Recommendation Systems in Software Engineering. 2014.
5. Monowar H. Bhuyan, Dhruva K. Bhattacharyya, and Jugal K. Kalita. Towards generating real-life datasets for network intrusion detection. *International Journal of Network Security*, 17(6):683–701, 2015.
6. Sebastian Garcia. *Identifying, Modeling and Detecting Botnet Behaviors in the Network*. PhD thesis, UNICEN University, 2014.
7. Jorge Guerra, Carlos Adrián Catania, and Eduardo Veas. Visual Exploration of Network Hostile Behavior. *Proceedings of the 2017 ACM Workshop on Exploratory Search and Interactive Data Analytics - ESIDA '17*, pages 51–54, 2017.
8. Ludmila I. Kuncheva. *Combining Pattern Classifiers: Methods and Algorithms: Second Edition*. 2014.
9. Robin Sommer and Vern Paxson. Outside the Closed World: On Using Machine Learning for Network Intrusion Detection. *2010 IEEE Symposium on Security and Privacy*, 0(May):305–316, 2010.
10. Irvine University of California. Knowledge discovery in databases darpa archive. <http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html/>, October 1999. [Online; accessed September-2016].

Observer effect: How Intercepting HTTPS traffic forces malware to change their behavior

María José Erquiaga^{1,2,3}, Sebastián García⁴ and Carlos García Garino^{1,3},

¹ ITIC, UNCuyo, Mendoza, Argentina

² Facultad de Ciencias Exactas y Naturales, UNCuyo, Mendoza, Argentina

³ Facultad de Ingeniería, UNCuyo, Mendoza, Argentina
merquiaga@uncu.edu.ar, cgarcia@itu.uncu.edu.ar

⁴ CTU University, Czech Republic
sebastian.garcia@agents.fel.cvut.cz

Abstract. During the last couple of years there has been an important surge on the use of HTTPS by malware. The reason for this increase is not completely understood yet, but it is hypothesized that it was forced by organizations only allowing web traffic to the Internet. Using HTTPS makes malware behavior similar to normal connections. Therefore, there has been a growing interest in understanding the usage of HTTPS by malware. This paper describes our research to obtain large quantities of real malware traffic using HTTPS, our use of man-in-the-middle HTTPS interceptor proxies to open and study the content, and our analysis of how the behavior of the malware changes after being intercepted. The research goal is to understand how malware uses HTTPS and the impact of intercepting its traffic. We conclude that the use of a interceptor proxy forces the malware to change its behavior and therefore should be carefully considered before being implemented.

Keywords: Malware, Botnets, HTTPS, Malware Traffic, Network Security, MITM, proxy, Malware Behavior

1 Introduction

In physics, the observer effect alludes to the influence of the observer on the phenomenon under observation. In most areas, this influence is often caused by instruments and modifies the behavior of what is being measured in some important manner. However, the observer effect can also appear in other contexts, such as malware execution and analysis for network security. Unfortunately, the impact of capturing techniques on the behavior of malware in the network was often overlooked. In this work, we study the observer effect in regard to the use of web TLS (Transport Layer Security) interceptor proxies for network malware analysis. The usage of web proxies TLS interceptors allows companies to view the content of its employees encrypted connections. This technology is nowadays widely implemented in large companies and it is supposedly used to protect the employees from infections. However, the inflicted changes upon the behavior of malware, the invasion of

employee's privacy and the implementation costs can be an inconvenience for some companies.

This research focuses on understanding the influence of web proxies TLS interceptors during the execution of malware. To achieve the analysis, we created a new and large TLS dataset of real malware samples extending for more than one year. The malware was run in a special infrastructure for selective interception and with complete access to the Internet. The average execution time of each malware is one week, with some executions lasting two months. The main goal of this work is to find how malware changed its behavior due to the filtering, blocking and interception of its web TLS connections. The most important factor of our research was the verification process, which consisted in executing the malware twice and simultaneously: once using the interceptor proxy and once without it.

The contributions of this work are: (1) The creation of a dataset. The dataset created is a large and modern dataset, it includes more than 80 malware captures and captured carefully with a methodology explained in this article. (2) Malware interception using https or port 443. The malware selected to captured was the one using encrypted communication or port 443 (the port assigned to encrypt the communication). Also, two scenarios were considered; with proxy and without proxy. (3) Publication of the dataset for the research community. The dataset was published in the stratosphere web site and twitter. (4) An analysis of the implications of intercepting the traffic of malware (with and without proxy), described in this paper.

The remainder of this work is organized as follows. Section 2 describes the background concepts, definitions and the previous work. Section 3 describes the dataset creation and features. This section also includes a description of the laboratory infrastructure and the methodology to obtain the dataset. Section 4 shows an analysis of the malware captures and a comparison of the same malware with and without using proxy. Finally, Section 5 presents the conclusions and future work.

2 Background and Previous Work

TLS is the standard security protocol for encrypting information in a network. It establishes an encrypted link using asymmetric and symmetric encryption algorithms. TLS is the widest used encryption protocol because it uses a trust chain to verify that the service is trustable and it doesn't require the client to have any special password. Recently, there seems to be a rise in the use of TLS by malware, causing new difficulties for the analysts and rising several new questions about what the malware is doing inside the encrypted channel, and how its attack strategy changes.

The study of TLS interception in the network and its impact on the users has been analyzed before from the perspective of privacy [1]. The practice of using TLS interceptor proxies is common inside companies, even when employees are unaware. The most important impact of this practice is that users get used to being intercepted and their general security protection measures decrease. The authors concluded that approximately 0.6% of all users inside companies are subject to interception.

The analysis of TLS usage by malware was previously studied by Cisco in an attempt to find their actions without decrypting the traffic [5]. The authors used 18

malware families that usually encrypt their traffic to understand the motivations behind encryption. The research found that 98.4% of the encrypted malware traffic used port 443/TCP for this purpose. However, there is no analysis of the malware using port 443/TCP with other protocols that are *not* TLS. Therefore, they mainly focused on TLS protocol on port 443/TCP.

Furthermore, Carné de Carnavalet et al [2], conduct investigations into the risks introduced by TLS interception tools, by analyzing the risks of using antivirus and parental control tools with a proxy.

Cisco researchers have found that the amount of malware using TLS has grown in the last years. Their analysis reveals that malware uses TLS differently than normal traffic. [6]. The author proposed to augment the common 5-tuple structure with TLS-based features. Among the additional features are the list of offered cipher suites, the selected cipher suite, the sequence of lengths and type codes of TLS records, and the time between TLS records in milliseconds. [7]

The most common analysis of the TLS protocol on malware traffic focus the detecting the malware. Common techniques include, for example, behavioral traffic analysis using a k-NN classification [3]. Another recent approach was to detect the malware by studying its encrypted HTTPs communication [4]

3 HTTPs Malware Dataset

The most important challenge of analyzing malware using HTTPs is the lack of a good public dataset. As part of our work we spent almost one year collecting real, and long term, malware traffic. The dataset created is part of our Nomad Project [9]. Our dataset consists of more than 80 network malware traffic captures. One of the goals of the dataset is to study the behavior of malware and how it changes in time. To obtain this type of data we executed the malware for long terms, up to 3 weeks or even months. The dataset contains malware capture of different types of malware (such as Trojans, Adware, botnets, etc.). For each capture, we generated several files to improve future analyses.

The process of creating the dataset can be described in four phases, (1) design and creation of the laboratory, (2) design of the capture methodology, (3) generation of experiments and output of information. The following subsections describe these phases in detail.

3.1. Malware Laboratory Infrastructure

The malware laboratory infrastructure consisted in a host Linux computer running more than 30 VirtualBox Windows 7 virtual machines. The host computer also ran a separate mitmproxy¹ implementation for each malware, allowing the complete isolation of results. The malware had unrestricted access to the Internet except for a limited bandwidth and an SMTP redirection to an e-mail honeypot. Fig. 1 shows a

¹ <https://mitmproxy.org/>

basic schema of the intercepting infrastructure. It can be seen how all most common ports for the web connections were redirected through the web TLS interceptor.

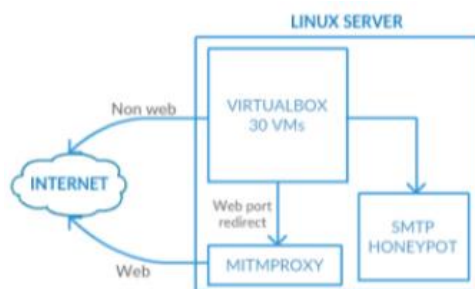


Figure 1. Malware capture laboratory infrastructure

3.2 Capture Methodology

The methodology implemented to capture malware traffic is based on the analysis of known malware that we use to infect machines in our laboratory, described before. The steps to capture malware traffic are: (1) find malware binary (2) copy the binary to the server (3) start the virtual machine and infect it (4) compute the start date and the infection date and monitoring the machine (5) stop and publish the capture. To find the malware binaries, we search in web sites specialized in malware samples such as virus total, hybrid analysis, and SSL blacklist. We monitor each machine using tcpdump, mitmproxy outputs and cacti. Thought those tools it was possible to detect if the machines were infected or not and if the communication was still alive. Once we considered we had enough information, we stopped the capture, generate the output of the dataset and published it in with the corresponding ID in the stratosphere website, blog and twitter.

3.3 Experiments and output files

The dataset consists of 80 different malware captures [8]. The malware captured includes botnet, trojans, backdoors and adware among others. Approximately 90% of these captures used port 443/TCP to communicate. However, only 83% of the captures used the TLS protocol over port 443/TCP. This means that at least 17% of malware used port 443/TCP for their own custom encryption protocol. Regarding the usage of these connections, from the complete set of captures, approximately 30% used the port 443/TCP (both with and without TLS) for their Command and Control channels. This means that most of the TLS connections generated by malware were not directly related to its remote control.

Each malware has assigned a unique ID. This ID allow us to identify each malware capture. If we execute the malware twice, we add a number to the ID. For instance, 260 is the capture ID, the first capture will be 260-1, the second one 260-2, and so on. The URL format to access to the capture information contains this ID. For example: <https://mcfp.felk.cvut.cz/publicDatasets/CTU-Malware-Capture-Botnet-260-1>.

4 Comparison and Evaluation of the Behavior of Malware with Intercepting TLS proxies.

In this section, we analyze the network behavior of malware captures when they are intercepted by an TLS proxy. We analyzed each malware traffic capture to evaluate the behavior in two scenarios: with proxy interception and without proxy interception. The goal is to detect the behavior differences in the same malware when the execution was made with or without proxy. As we explained in **Section 3.3**, each capture has a unique ID, the purpose of this ID is to identify each capture. For instance, the malware capture ID 169-1 belongs to a malware capture, the ID 169-2 is the same malware executed a second time. We will analyze malware captures with proxy and without proxy. Therefore, ten malware captures were analyzed: 169-1, 169-2, 192-1, 192-2, 219-1 and 219-2.

4.1 Miuref Capture (169-1)

This malware belongs to the Miuref family. It was executed three times, twice with an interceptor proxy (captures 169-1 and 169-2) and once without proxy (capture 169-3).

Miuref Capture with TLS Proxy

In this capture, Miuref first resolves the domains service8.org, 1.web-counter.info, timeservice24.com, 2.web-counter.info, 3.web-counter.info, 4.web-counter.info, and 5.web-counter.info. These are the main Command and Control (CC) channels of the malware. All the domains have one or more IP addresses and the malware starts to connect to them on port 443/TCP. Given that the protocol spoken by the malware on port 443/TCP is not TLS, the MITM proxy generates the follow error:

502 Bad Gateway. SSL handshake error: The client may not trust the proxy's certificate.

This is the first important conflict between the malware and the proxy, since the proxy does not allow the traffic to connect to Internet. Therefore, the C&C servers are not reached properly. After the unsuccessful connections, the malware contacts an IP address that is hardcoded in the binary: **185.118.67.195**. This IP is contacted on port 80/TCP and the protocol spoken is correct HTTP. This backup C&C mechanism worked well and allowed the malware to download a possible binary update.

The second most important conflict of the malware and the MITM proxy is that the malware keeps trying to connect to the 443/TCP ports in the C&C servers, generating a large amount of very suspicious traffic.

As the malware is related with ClickFraud and advertising, it also connects to twitter.com, youtube.com, google.com, facebook.com, bing.com, etc. These requests were done because certain webpages had links to them and not because the malware was abusing their services. We were able to verify this claim because it was possible to observe the HTTPs traffic and distinguish the real requests.

Then, the malware resolves the domain bam.nr-data.net, obtaining the following IPs 50.31.164.175, 162.247.242.18, 50.31.164.173, 50.31.164.166, 50.31.164.174, 162.247.242.19. It uses port 80 to contact the IP 50.31.164.175 and establishes a CC channel. For example, an unsuccessful connection to port 443 that were not SSL is the one to the IP 138.201.125.95.

Miuref Capture Without Proxy (169-2)

This is a capture of the Miuref malware but without using any MITM proxy. In this case, the malware tries to connect to its C&C servers using the port 443/TCP with its custom encryption protocols and it is successful. After contacting the C&C servers the malware also contacts its hardcoded IP address **185.118.67.195** and downloads a binary update. The operation of the malware remains similar to the capture with the proxy and it connects to several ClickFrauds and advertising sites (platform.twitter.com, connect.facebook.net, google.com and bing.com).

In this capture the behavior is similar to the previous capture (resolves the same domains, contacts the same IPs). However, the main difference is that the TLS connection is established. The differences in domains accessed by the two captures is also significant, although it is difficult to know exactly why the malware is connecting to hundreds of sites. From the 465 totals, unique domains accessed by the capture with MITM proxy and the capture without MITM proxy, 86% was requested by the malware without the MITM proxy. From the 206 unique domains requested in the capture without MITM proxy, 30% were unique to this capture. From all the traffic in the capture with the MITM proxy, 18% was sent to port 443/TCP, while only 8% of the traffic in the capture without the MITM proxy was sent to the same port. This difference is due to the use of an interceptor HTTPs proxy.

4.2 Remote Admin.Ammy Capture (ID 192)

This malware is a remote admin tool called Ammy. Ammy is a possible legitimate application, but it is usually abused for its remote administration capabilities. This malware was executed twice; with proxy, and without proxy.

Remote Admin.Ammy Capture 192-2 with proxy

This malware first resolves the domain *rl.ammy.com* and contacts the IP **176.56.184.37** in port 80/tcp. Then, it contacts four different IPs in the ports 443/TCP, 80/TCP and 8080/TCP in sequence. The IP address are **88.198.6.56**, **88.198.6.55**, **95.211.242.83** and **95.211.191.142**. Those IPs could be hardcoded in the binary file, or they may be delivered in the communication C&C channel with the server from the IP 176.56.184.37. The malware does not connect successfully to the IPs 88.198.6.56, 88.198.6.55, 95.211.242.83 and 95.211.191.142. This happens because the MITM

proxy cannot recognize the protocols used in those ports. It means, that the malware is using the protocol in an unconventional manner.

Remote Admin.Ammy Capture 192-1 and 192-3 without proxy

The malware resolves the domain `rl.ammy.com` and connects to its IP, `176.56.184.37`, in the port `80/tcp`. This CC channel works fine and receives a string of binary data. The malware then connects to a group of four IPs: `95.211.191.142`, `95.211.242.83`, `88.198.6.56`, `88.198.6.54`. The IPs are connected in an endless loop. In each loop, each IP is contacted on port `443/tcp`, `80/tcp` and `8080/tcp`, in that order mostly. These IPs could be sent in the binary answer from the first CC server or they could have been hardcoded in the malware binary. None of the looping servers were correctly executing the CC. This resulted in the malware not being able to activate or further receive orders.

Regarding capture 192-3, we observed; first, it resolves `www.msftncsi.com` and contacts the remote server in port `80`. Then, it contacts the same IP (`217.182.53.102`) on port `443`. It resolves the domain `rl.ammy.com`, and contacts the domain in port `80` and exchange binary information. Then it contacts the IP `95.211.191.142`, port `443 tcp` (it sends some binary information). Repeating this process in an endless loop.

4.3 Capture Kover.B (ID 219)

The malware executed is probably a trojan called Kover.B, it was executed with proxy and without proxy. In both cases, the malware contacted the remote Command and Control server in port `443/TCP` and established an encrypted channel.

Capture Kover.B (ID 219-1) with proxy

First, the malware tries to establish a Command and Control channel with the IP `42.2.231.204` on port `80/TCP`, with `149.80.126.178` on port `8080/TCP` and with `119.116.67.233` on port `443/TCP`. Most of them used an unknown encryption protocol. These connections failed because the proxy could not understand them. The only connection that is successful is to IP `42.2.231.204` because it used real HTTP. Then, the malware attempts to create C&C channels on port `8080/TCP` and `443/TCP` with several IP addresses. Most of these also failed for the same reason. When all these connections failed, the malware establish a C&C channel with the IP `185.117.72.90` on port `80/TCP` using HTTP protocol. The malware contacted more than 50 IP addresses using port `80/TCP`. Some of the domains are contacted but, after 10 seconds of non-response from the remote servers, it resets the connection. It establishes a CC SSL channel with the IP `168.150.126.63`, and then it closes the connection. Then it contacts IP `48.99.155.215` in port `443`, and exchanges binary information.

Capture Kover.B (ID 219-2) without proxy

The malware starts by creating a C&C channel with the IP address `23.4.249.223` on port `80/TCP` and port `443/TCP`. In the first case, the response from the remote server

is *HTTP/1.1 301 Moved Permanently*. In the second case, it exchanges certificates and establishes an SSL CC channel, after that, it exchanges binary information.

Then, the infected computer establishes a connection to www.microsoft.com, which is odd for any malware. Since the Windows computer is configured not to update, this may be a result of the malware opening another process. There are also connections to www.download.windowsupdate.com.

At the same time, the malware tries to connect to at least 50 different IP addresses on port 443/TCP. Most of them do not respond and after of 10 seconds of inactivity and it resets the connection. Eventually it establishes an SSL communication with the IP **198.144.30.128** on port 443/TCP, related with the domain *store.korfx.com*. It is clear that these IP are hardcoded in the malware.

The malware contacted different IP addresses on port 80/TCP trying to reach different C&C servers. However, most of them were not working. Only 4 IP addresses were contacted on port 443/TCP as C&C servers. It sends Sync packets up to ~5 packets per second to several IPs in ports 80, 8080 and 443

4.4 Capture Trojanized BitTorrent with Open Candy (ID 208)

This malware is probable a BitTorrent client that was *trojanized* with the adware Open Candy. It was executed twice; with and without proxy. In the two scenarios, the malware contacted the port 443/TCP and established a communication with a remote server. However, it was not an encrypted communication.

This capture is shown as an example of how there may be situations where the use of MITM proxy does not have an impact in the traffic. Since this malware does not have a Command and Control channel, it did not have any issue with the proxy.

Capture Trojanized BitTorrent with Open Candy (ID 208) with proxy

During this execution, the malware resolved different domains (*i-50.b-000.xyz.bench.utorrent.com*, *i-21.b-42606.ut.bench.utorrent.com* and *bittorrent.vo.llnwd.net*) related with the operation of BitTorrent. Most of the connections and updates are done using port 80/TCP. Some IP addresses are also contacted using UDP packets, due to the operation of the P2P protocol. Finally, it contacts other domains (previously contacted using the port 80), using the port 443/TCP. This process is repeated several times during the capture.

Capture Trojanized BitTorrent with Open Candy (ID 208) without proxy

This capture has the same behavior as the previous capture (208-1, with proxy). It contacts some domains on port 80, then it establishes an SSL communication with those IPs on port 443.

4.5 Discussion

After executing the malware with and without the web TLS proxy interception we discover a very important characteristic of the problem studied: the behavior of most malware changed when it was executed with a proxy. The most important findings are

three. First, while using the interceptor proxy, certain amount of malware, was not able to communicate to Internet at all, because the protocol used on port 443/TCP was not TLS. Therefore, the proxy refused to establish the connection. This forced the malware to new actions. Second, when the malware didn't achieve the connection to its remote servers, two new behaviors were observed: (1) the malware tried to reconnect continually to its C&C server, generating huge noise in the network, and (2) the malware seek another way to connect (choosing a different port, or looking for others servers). An example of this difference can be seen in Figure 2 and Figure 3. In particular, Figure 2 describes the behavior of the malware when was captured with mitmproxy and Figure 3 illustrates the behavior of the same malware without using a proxy. Third, some malware running with the web interceptor was not able to establish the C&C channel, while the capture without the web interceptor was able to do it. In these cases, the intercepted malware sent more than three times the amount of traffic compared with the not intercepted malware.

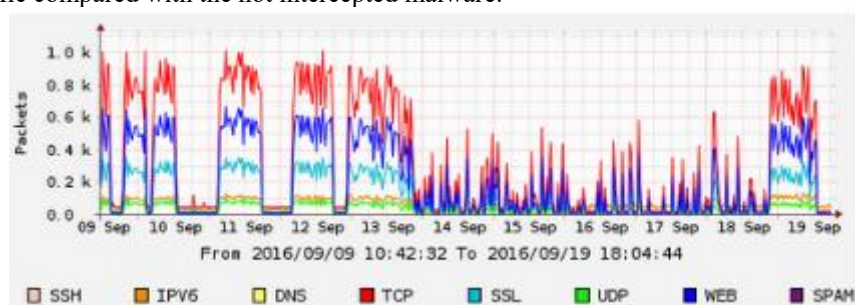


Figure 2. Malware captured with mitmproxy

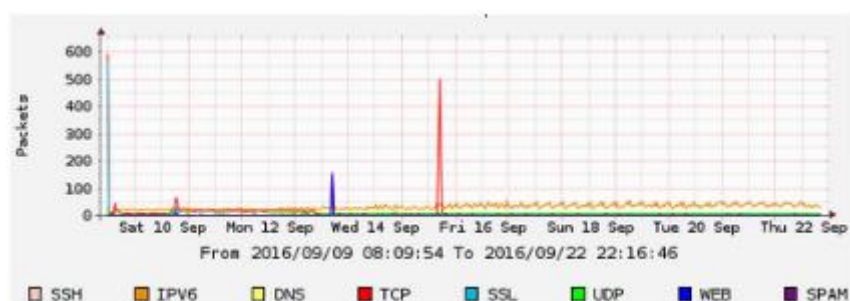


Figure 3. Malware captured without mitmproxy

5. Conclusion and future work

The analysis of the behavior of malware inside an encrypted connection is difficult. For this reason, we consider our malware dataset and analysis of encrypted behavior a step forward in this topic. There are two main conclusions regarding the impact of using a proxy for malware analysis. First, we observed a large amount of malware using a custom protocol on ports reserved for the HTTPS/HTTP protocols (ports 80,

8080 and 443). Blocking these connections forced the malware to generate new behaviors. Second, we noticed that malware's behavior can change in undefined ways when an intercepting proxy is used. Therefore, the implementation of a proxy should be carefully considered to detect malware in the network.

As future work, we will analyze more in detail other features in the captures. First, to detect which malware we are executing and second, to identify which version of the protocol the malware is using. When we executed the malware samples, we were provided of information from virus total and hybrid analysis. Even though this site is the more important virus database, it provides information from the possible name of the malware. For that reason, we consider that it is essential to verify which malware are we really executing. Second, we compared the differences between the malware that was using encrypted communication and not, but we did not analyze the version of the encryption protocol it was using.

References

1. O'Neill, M., Ruoti, S., Seamons, K., & Zappala, D. (2017). TLS Inspection: How Often and Who Cares? *IEEE Internet Computing*, 21(3), 22–29. <http://doi.org/10.1109/MIC.2017.58>
2. Xavier de Carné de Carnavalet and Mohammad Mannan. Killed by Proxy: Analyzing Client-end TLS Interception Software. 21-24 February 2016, San Diego, CA, USA. Copyright 2016 Internet Society, ISBN 1-891562-41-X. <http://dx.doi.org/10.14722/ndss.2016.2337>
3. Lokoč J., Kohout J., Čech P., Skopal T., Pevný T. (2016) k-NN Classification of Malware in HTTPS Traffic Using the Metric Space Approach. In: Chau M., Wang G., Chen H. (eds) Intelligence and Security Informatics. PAISI 2016. Lecture Notes in Computer Science, vol 9650. Springer, Cham
4. František Strásák. Detection of HTTPS Malware Traffic. Open Informatics, Computer and Information Science. May 2017. https://dspace.cvut.cz/bitstream/handle/10467/68528/F3-BP-2017-Strasak-Frantisek-strasak_thesis_2017.pdf?sequence=-1
5. Anderson, B., Paul, S., & McGrew, D. (2016). Deciphering Malware's use of TLS (without Decryption). Retrieved from <http://arxiv.org/abs/1607.01639>
6. Blake Anderson. Hiding in Plain Sight: Malware's Use of TLS and Encryption. January 25, 2016. <http://blogs.cisco.com/security/malwares-use-of-tls-and-encryption>
7. Blake Anderson, David McGrew, Alison Kendler. Cisco Systems, Inc. Classifying Encrypted Traffic with TLS-Aware Telemetry. January 2016. <http://resources.sei.cmu.edu/library/asset-view.cfm?assetID=449962>
8. Stratosphere Dataset. <https://stratosphereips.org/category/dataset.html>
9. Nomad Project. <https://stratosphereips.org/category/Nomad.html>

Posibles problemas en Qlink.it y librería CryptoJS

Antonio Castro Lechtaler¹, Marcelo Cipriano¹, Edith García¹, Pablo Lázaro², Julio Liporace¹, Eduardo Malvacio¹, Ariel Maiorano^{1,2}

¹ Grupo de Investigación en Criptografía y Seguridad Informática (GICSI),
Instituto Universitario del Ejército (IUE);

² Dirección de Gestión Tecnológica (DGT), Policía de Seguridad Aeroportuaria (PSA)
{acastro,marcelocipriano}@est.iue.edu.ar,
{editxgarcia,edumalvacio,jcliporace}@gmail.com,
{plazaro,amaiorano}@psa.gob.ar

Resumen. El objetivo del presente trabajo es describir los resultados preliminares obtenidos luego de una revisión del código fuente de la aplicación web Qlink.it. Este análisis inicial, aunque incluyendo aspectos relacionados a la seguridad de la aplicación de manera general, tuvo foco particularmente en lo relativo a su implementación de funcionalidades criptográficas. Si bien, salvo por la vulnerabilidad de XSS detectada, los potenciales problemas podrían no representar riesgos reales, el objetivo de esta publicación es también invitar a otros revisores a estudiar el sistema para confirmar si su utilización podría considerarse segura.

Palabras clave: seguridad informática, seguridad de aplicaciones, revisión de código fuente, criptografía, generación de números aleatorios, Qlink.it, CryptoJS.

1 Introducción

Recientemente se han publicado noticias [1,2] informando acerca de la disponibilidad del código fuente del sistema Qlink.it [3]. Considerando, por un lado, que el GICSI tiene por objetivo, entre otros, estudiar técnicas y mecanismos para la revisión de código fuente, con foco en los aspectos relacionados a la seguridad informática en general y a la criptografía en particular [4,5]; y por otro lado, que la DGT tiene por incumbencia, entre otras, evaluar periódicamente alternativas para la comunicación segura del personal de la Institución; se realizó conjuntamente una primera revisión general del código fuente de la aplicación web Qlink.it [6].

Al momento completar esta primera fase del análisis (mayo de 2017), los resultados preliminares de la revisión -aunque parcial e incompleta- advertirían posibles o potenciales problemas de seguridad, por lo cual se decidió consultar a los desarrolladores del sistema compartiendo estos resultados. Aunque se trate de una revisión que no abarcó a la totalidad del sistema, e inacabada, se solicitó el debido permiso para publicar esta información, en forma de artículo, con la intención de invitar a otros

revisores a estudiar el sistema, quienes podrían confirmar o rechazar estos potenciales problemas, y determinar si la utilización del sistema podría considerarse segura.

Es en este sentido también que un resumen limitado y sin mayores detalles de los primeros resultados fue publicado en el sitio dedicado a la seguridad de la información Segu-Info [17].

Se debe aclarar que uno de los potenciales problemas de seguridad tratados es “heredado” por la aplicación Qlink.it, al utilizar la librería CryptoJS [18].

Se aclara por último que algunos de los posibles problemas descriptos podrían no representar riesgo alguno, sin embargo se incluyen todos los resultados obtenidos para eventuales futuras revisiones por terceros y para a la vez dar cuenta del limitado alcance de la presente revisión.

1.1 Acerca de Qlink.it

De acuerdo a lo indicado en la documentación del proyecto [3,6], específicamente en su sección de preguntas frecuentes, “Qlink.it es una nueva manera, muy simple y segura, de enviar información confidencial a través de internet”.

Se detalla además que “Un qlink es un enlace normal de internet, pero con la característica especial de que se auto-destruye luego de que es leído por primera vez, aún si lo hubiese leído un robot. Como verás, esta característica especial es la pieza clave para enviar información confidencial a través de internet: si pones la información en un qlink, y envías el qlink a través de un correo o mensaje normal, entonces el destinatario sabrá que nadie más leyó la información si ésta estaba allí al pinchar el qlink.”

Modo de funcionamiento. Resumidamente, de acuerdo a lo informado también desde el sitio Web Qlink.it, en su sección de preguntas frecuentes avanzadas, el funcionamiento del sistema está descripto de la siguiente manera (en inglés en el original):

1. Cuando se ingresa un mensaje en qlink.it y se hace clic en el botón “qlink it!”, el navegador ejecuta código de Javascript que cifra el mensaje con una clave aleatoria dada, por ejemplo, YYYYYY.
2. Posteriormente, el mensaje cifrado se envía a través del protocolo https seguro al servidor Qlink.it.
3. En el servidor, el mensaje (ya cifrado con clave YYYYYY) se cifra de nuevo para ser almacenado, pero ahora con otra clave aleatoria, por ejemplo, XXXXXX.
4. Entonces, el servidor le devuelve un qlink preliminar, en este caso `https://qlink.it/XXXXXX`.
5. En ese momento, el navegador agrega al final del qlink preliminar la clave que sólo el navegador conoce para formar el qlink completo: `https://qlink.it/XXXXXX#YYYYYY`. Téngase en cuenta que el servidor Qlink.it no tiene acceso a la parte YYYYYY del qlink!
6. A continuación, es posible copiar y pegar el qlink completo y enviarlo al destinatario, ya sea por correo electrónico, chat, WhatsApp, o lo que sea.
7. Cuando el destinatario reciba el qlink completo y haga clic en él, el navegador sólo solicitará al servidor el qlink preliminar, `https://qlink.it/XXXXXX`,

porque la marca del carácter especial (#) indica que lo siguiente no debería ser enviado a través de Internet (Se puede comprobar esta función utilizando, por ejemplo, la opción de inspección en algunos navegadores como podría ser Chrome). Por lo tanto, el servidor Qlink.it nunca tiene acceso a la llave completa para leer el verdadero contenido del mensaje.

8. Cuando el servidor recibe la petición con el qlink preliminar, el qlink tiene en él la clave para buscar el mensaje encriptado y descifrarlo parcialmente. El servidor envía de nuevo a través del protocolo seguro https un mensaje que todavía está cifrado con la clave desconocida para el servidor YYYYYY. En ese momento el servidor hace una eliminación segura en el mensaje encriptado y ya no está disponible en el servidor.
9. Cuando el navegador del destinatario obtiene el mensaje cifrado, ya que mantuvo la última parte del qlink completo YYYYYY, ejecuta código Javascript para finalmente descifrar el mensaje usando esta última parte del qlink completo. Una vez que el mensaje se descifra totalmente, el navegador lo muestra en la pantalla del destinatario.

2 Posibles problemas en Qlink.it

Los siguientes potenciales problemas han sido detectados en una instancia del sistema instalada y configurada de acuerdo a las indicaciones provistas en [6]. Es por esto que los scripts que fueron desarrollados para las pruebas se encuentran configurados para trabajar con el servidor en “http://qlink/”. No se han realizado pruebas de ningún tipo en el servidor disponible en Internet, actualmente accesible en “https://qlink.it”.

2.1 Vulnerabilidad del tipo Cross-Site-Scripting (XSS)

Aunque se utilizan funciones Javascript para el filtrado de campos de entrada a completar por el usuario al generar un qlink –por ejemplo `filterXSS()` y `escapeHtmlEntities()`–, para el campo que contiene el mensaje, la entrada parece no verificarse o “filtrarse” correctamente. La salida del script de prueba anexada, “punto_2_1.py”, muestra que se simula lo que realizaría un navegador en cuanto a cómo es enviada la información al web-service que responde a estas peticiones en el servidor, especificando como mensaje, luego de un *tag* de cierre del elemento *textarea* que presenta el mensaje descifrado, un elemento `<script>` conteniendo una invocación a la función `alert()` para demostrar la vulnerabilidad frente a ataques del tipo XSS.

2.2 Otros parámetros de entrada y operaciones criptográficas

Luego de revisado el código en los archivos `public/js/application.js`, `app/src/Qlink/Models/Utils/RandomHasher.php` y `/app/src/Qlink/Controllers/LandingNewController.php`, se advierte que la primera parte parte del qlink, esto es, los primeros 10 caracteres, por ejemplo: `http://qlink/two/XXXXXXXXXX...` Son generados (no exclusivamente) en

base a un *timestamp* al milisegundo -resultado de la función Javascript `Date().getTime()` - que el navegador envía al servidor, y por lo tanto, que podría manipularse. Aunque el servidor registrará ese valor para usarlo en el próximo qlink, usando para la petición en curso el valor inmediato anterior, registrado previamente de la misma forma, el valor luego es sumado al *timestamp* del servidor, a la cantidad de microsegundos multiplicada por cien mil, y usado como semilla -mediante la función `mt_srand()` - para luego obtener valores a partir de la función `mt_rand()`.

Estas últimas funciones, basadas en el generador conocido como *Mersenne Twister*, no son aptas para utilizarse cuando existe la necesidad de números aleatorios para operaciones criptográficas, precaución explicitada también en su documentación oficial [7].

En el mismo sentido que lo anterior, ahora hablando del código que se ejecutará en el navegador mediante Javascript, aunque indirectamente, a través de la librería `CryptoJS` y su función `CryptoJS.lib.WordArray.random()`, la obtención de números aleatorios termina invocando a la función Javascript `Math.random()`, que es implementada por los navegadores en base a variantes del generador *Xorshift128+*, que tampoco es considerada segura o recomendable para la implementación de criptografía [8,9,10,11].

2.2.1. Estimación de fecha y hora de creación del qlink anterior

Sólo como un ejemplo que se desprende de lo anterior, se anexa también una salida de ejemplo de otro script de prueba, “prueba_2_2_1.py”, que requiriendo la generación de un qlink, tomará sus primeros diez caracteres para calcular la semilla utilizada, y realizar una estimación de cuándo fue creado el qlink inmediato anterior registrado. El script se vale del módulo o paquete `php_rand` [12].

Supóngase a x como el *unix epoch timestamp* del momento en que se generó el qlink anterior, en segundos; y a x_m al parámetro que se envió al servidor en tal momento, en milisegundos, por lo que a efectos de esta estimación aproximada, $1000x < x_m < 1000x + 999$. Por otro lado, asúmase t igual al *unix epoch timestamp* del momento en generamos el nuevo qlink, en segundos (suponiendo también que los relojes del cliente y servidor se encuentran sincronizados al segundo). Por último considérese a u como la cantidad en microsegundos utilizada en el código PHP de qlink, que de acuerdo a la implementación y recorte que sucede luego por número entero, se generaría tal que $0 < u < 99999$.

Por lo tanto, la semilla para el qlink que estamos generando, s , correspondería a $x_m + t + u$. Entonces, $s = x_m + t + u$, $s = 1000x + y + t + z$, con $0 < y < 999$ y $0 < z < 99999$. Luego, $x = (s - y - t - z)/1000$. Siendo esto una aproximación, se elimina el término $y/1000$, y $z/1000$ se reemplaza por un delta d , con $0 < d < 99$, entonces $x = [(s - t)/1000 - 99, (s - t)/1000]$. Por lo que la aproximación resultante corresponde a un rango de 99 segundos.

2.2.2. Obtención de "número DN" a partir de un qlink

También en base a los scripts de los ejemplos anteriores, se anexa la salida de “punto_2_2_2.py”, que muestra cómo sería posible obtener el código de seguimiento, o “número DN” de acuerdo a como es denominado en el sistema Qlink.it, a partir de los primeros diez caracteres de la parte variable de un qlink. Al arrancar, se genera un qlink para usar como ejemplo, pero el script podría adaptarse para funcionar con cualquier qlink que se encuentre vigente. El “número DN” es generado de manera similar a la primera parte de un qlink, sólo que el resultado corresponde a diez dígitos.

Se genera mediante una función muy similar a la anterior, radicando la diferencia en el conjunto de caracteres posibles para el mapeo de los números aleatorios. Se utiliza además el mismo *timestamp* que para la generación de la primer parte de un qlink. La función es invocada después de poco más de 50 líneas de código respecto a la generación de la primer parte del qlink. Por lo tanto, el script realizará pruebas para estimar el tiempo transcurrido entre la invocación a esas funciones para luego comprobar la existencia del código de seguimiento acotadamente, con la intención de reducir la cantidad de pruebas a realizar. Téngase en cuenta que puede ser necesario ejecutar varias veces el script para obtener resultados exitosos.

2.2.3. Generación de números aleatorios potencialmente insegura (Javascript, utilizando librería CryptoJS)

Volviendo a cuestiones relativas al código ejecutado en el navegador, específicamente en lo relacionado a la generación de números aleatorios, se anexa la salida de una prueba ejecutando “punto_2_2_3.py”, para demostrar potenciales problemas con la utilización de librerías basadas en la función que proveen los navegadores, implementando el generador *Xorshift128+*, de acuerdo a lo que fuera adelantado más arriba.

En este caso el script de prueba se trata de una adaptación de otro disponible en [14], que trabaja directamente con salidas de la función `Math.Random()`, valiéndose de la herramienta *Z3*, “*a high-performance theorem prover being developed at Microsoft Research*” [15], para la resolución simbólica del sistema de ecuaciones dada la información parcial conocida. El ejemplo de prueba fue adaptado para resolución con valores truncados por `CryptoJS.lib.WordArray.random()`. Se tomó a su vez como ejemplo la manera de generar el *salt* y el *iv* en qlink para estimar o adivinar valores posibles siguientes del generador. Si bien este ejemplo no representa de por sí un riesgo, debe considerarse que la misma función es utilizada para generar material de llave.

3 Posibles soluciones

Para los problemas descriptos en el apartado anterior, se proponen en lo que sigue, soluciones posibles para remediarlos o reducir el riesgo para los casos en que la amenaza fuera potencial. Adviértase de que de ningún modo deben considerarse a estas soluciones como correctas, completas o pertinentes. Se intenta únicamente comentar

resumidamente acerca de mecanismos de seguridad para sistemas, de manera general y en lo relativo a implementación de criptografía. Se sugiere a la vez la consulta de la referencia obligada en la materia, *The Open Web Application Security Project* (OWASP) [16].

3.1 Vulnerabilidad Cross-Site-Scripting (XSS)

Este puede tratarse de un caso a resolver especialmente ya que el servidor recibe un mensaje cifrado mediante el algoritmo AES. En principio entonces por supuesto no se podría verificar ni filtrar el contenido desde los scripts PHP que procesan, en el servidor, la información enviada. Por lo tanto la única alternativa sería aplicar filtros antes de presentar el contenido descifrado. Esto por supuesto mediante código Javascript a ejecutarse en el navegador del usuario. Revisando el código fuente del archivo `public/js/application.js`, se advierte que ya se estaría aplicando la función Javascript `filterXSS()`, entre otras, sin embargo se comprueba que no funcionaría como se espera. Una posible solución, al margen de confirmar la correcta invocación de la función utilizada actualmente, sería reemplazar ese mecanismo por otro que nunca permita la ejecución de código Javascript al presentar el contenido en la página de lectura del mensaje descifrado.

3.2 Otros parámetros de entrada y operaciones criptográficas

A la hora de generar números aleatorios, se sugiere no utilizar funciones no seguras ni utilizar *timestamps* como semillas. También se desaconseja utilizar (directamente) otra información como códigos de teclas presionadas o posición del puntero. Frente a la necesidad de números aleatorios seguros, en Javascript, se recomendaría utilizar la función o método `RandomSource.getRandomValues()` de la nueva Web Crypto API, ya soportada en la mayoría de las versiones recientes de los navegadores más populares. Para el reemplazo de `mt_srand()/mt_rand()` del lenguaje PHP, la sugerencia sería obtener números aleatorios leyéndolos desde el archivo especial `/dev/urandom` (Linux). Alternativamente a ésta última recomendación, podría considerarse también la función para generar números aleatorios provista por el framework elegido por qlink, que se vale de la librería OpenSSL (utilizándose ya en el sistema para la generación de `x_tokens`). Se sugiere por último asegurar el manejo de números aleatorios y/o material de llave, para no recortarlos o truncarlos indebidamente en procesos de formato o adecuación al medio o presentación.

Como consideraciones generales, destacaríamos la conveniencia de realizar la autenticación, además del cifrado, del mensaje; por el navegador, al igual que el cifrado, sin compartir la llave con el servidor. Por ejemplo mediante un código de autenticación de mensajes, como podría ser HMAC, asegurando de esa manera la imposibilidad de modificación o alteración de esa información. Por otro lado, aunque se mencionó anteriormente, el control de abuso o prueba por fuerza bruta limitado al registro de la dirección IP demorando un segundo a cada petición, podría no ser la mejor alternativa disponible; quizá otro mecanismo criptográfico entre cliente (del web-

service) y servidor podría evitar otros abusos o pruebas de fuerza bruta aún potencialmente posibles.

4 Conclusiones

Los ejemplos del punto 2.2 pueden no representar problemas de seguridad concretos, pero debe destacarse la importancia de la posibilidad de manipular algunos parámetros utilizados como parte de semillas. Al margen de que el sistema tiene protección contra fuerza bruta limitando las peticiones a una por segundo por dirección IP, debería confirmarse si podría estimarse la fecha y hora del servidor a la manera en que está descrito por ejemplo en [13], y así, entre otras posibilidades, generar un mismo qlink repetidamente. Respecto al punto 2.2.3, convendría confirmar si el servidor podría o no conocer el estado del generador de acuerdo a la información recibida para estimar parte del material de llave del qlink.

Si bien, salvo por el primero de los casos, detallado en el punto 2, las potenciales vulnerabilidades podrían no representar un riesgo real, la utilización de herramientas y mecanismos de desarrollo diferentes a las recomendadas en buenas prácticas en relación a la seguridad en general, y a la implementación de criptografía en particular, hacen que se concluya prudente esperar otras revisiones del sistema o información ampliada por parte de sus desarrolladores.

5 Agradecimientos

Se desea agradecer a los desarrolladores y responsables del sitio Qlink.it por su rápida respuesta a nuestras consultas y su permiso para la publicación de los resultados preliminares. A su vez, se agradece también a Cristian Borghello por su ayuda en la publicación inicial resumida [17] en su sitio dedicado a la seguridad de la Información, Segu-Info, <http://www.segu-info.com.ar>.

Referencias

1. El físico argentino que creó un sistema de seguridad para e-mails. REVISTA NOTICIAS. [En línea] <http://noticias.perfil.com/2017/04/09/el-fisico-argentino-que-creo-un-sistema-de-seguridad-para-e-mails/>, accedido por últ. vez en mayo de 2017.
2. El acceso a mensajes encriptados por agentes de inteligencia vuelve al foco de debate. AGENCIA TÉLAM. [En línea] <http://www.telam.com.ar/notas/201703/183809-el-acceso-a-mensajes-encriptados-por-agentes-de-inteligencia-vuelve-al-foco-de-debate.html>, accedido por últ. vez en mayo de 2017.
3. Repositorio de Qlink.it en Github. [En línea] <https://github.com/qlinkit>, accedido por últ. vez en mayo de 2017.
4. Castro Lechtaler, A., Liporace, J., Cipriano, M., García, E., Maiorano, A., Malvacio, E., Tapia, N. Automated Analysis of Source Code Patches using Machine Learning Algorithms. XXI Congreso Argentino de Ciencias de la Computación (Junín, 2015). ISBN: 978-987-3806-05-6. [En línea]

- http://sedici.unlp.edu.ar/bitstream/handle/10915/50585/Documento_completo.pdf-PDFA.pdf?sequence=1, accedido por últ. vez en mayo de 2017.
5. Proyecto AAP, Repositorio de GICSI en Github. [En línea] <https://github.com/gicsi/aap>, accedido por últ. vez en mayo de 2017.
 6. Proyecto webapp, Repositorio de Qlink.it en Github. [En línea] <https://github.com/qlinkit/webapp>, accedido por últ. vez en mayo de 2017.
 7. Función mt_rand, Manual de PHP. [En línea] <http://php.net/manual/es/function.mt-rand.php>, accedido por últ. vez en mayo de 2017.
 8. Referencia de Math.Random() en Mozilla Developer Neywork. [En línea] https://developer.mozilla.org/en-US/docs/Web/JavaScript/Reference/Global_Objects/Math/random, accedido por últ. vez en mayo de 2017.
 9. Implementación de función Random() en CryptoJS. [En línea] <https://github.com/jakubzapletal/crypto-js/blob/master/src/core.js>, accedido por últ. vez en mayo de 2017.
 10. Implementación del generador XorShift128+ en Mercurial de Mozilla. [En línea] <https://hg.mozilla.org/mozilla-central/file/tip/mfbt/XorShift128PlusRNG.h>, accedido por últ. vez en mayo de 2017.
 11. Implementación del generador XorShift128+ utilizada en Chrome, Repositorio en Github. [En línea] <https://github.com/v8/v8/blob/master/src/base/utls/random-number-generator.h>, accedido por últ. vez en mayo de 2017.
 12. Implementation of mt_rand and mt_srand functions for bruteforce and speed, Repositorio en Github del proyecto. [En línea] https://github.com/Gifts/pyphp_rand, accedido por últ. vez en mayo de 2017.
 13. Argyros, G., Kiayias, A.: I Forgot Your Password: Randomness Attacks Against PHP Applications. En 21st USENIX Security Symposium. [En línea] <https://www.usenix.org/conference/usenixsecurity12/technical-sessions/presentation/argyros>, accedido por últ. vez en mayo de 2017.
 14. Symbolic execution for the XorShift128+ algorithm, Repositorio en Github del proyecto. [En línea] <https://github.com/dougward/XorShift128Plus>, accedido por últ. vez en mayo de 2017.
 15. The Z3 Theorem Prover, Repositorio en Github de los proyectos. [En línea] <https://github.com/Z3Prover>, accedido por últ. vez en mayo de 2017.
 16. The Open Web Application Security Project (OWASP). [En línea] <https://www.owasp.org/>, accedido por últ. vez en mayo de 2017.
 17. Posibles vulnerabilidades en Qlink.it (análisis web). Segu-Info. [En línea] <http://blog.segu-info.com.ar/2017/05/posibles-vulnerabilidades-en-qlinkit.html>, accedido por últ. vez en mayo de 2017.
 18. CryptoJS. Google Code Archive. [En línea] <https://code.google.com/archive/p/crypto-js/>, accedido por últ. vez en mayo de 2017.

6 Anexo

6.1 Pruebas y ejemplos relativos a los potenciales problemas tratados

Se anexan, como ejemplos de cada caso, las salidas de los scripts de pruebas desarrollados para la prueba de los potenciales problemas descriptos en el punto 2. El código

fuente de estos scripts no se incluye en este anexo por cuestiones de espacio pero quien estuviera interesado podría solicitarlo a los autores.

```
$ python punto_2_1.py
mensaje del qlink: Mensaje de prueba
mensaje para alert(): Prueba XSS
qlink: http://qlink/two/L3rpir1tqRPV8DUzVSPPGu#123456
$
```

Listado 1. Salida de la ejecución del script de prueba relativo al punto 2.1.

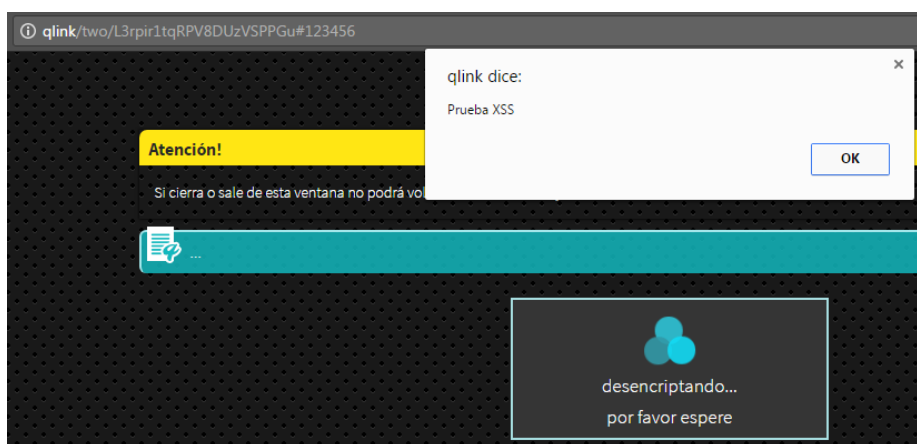


Figura 1. Captura de pantalla luego abrir el qlink generado.

```
$ python punto_2_2_1.py
qlink generado: http://qlink/two/f88cHwhwVDk14HN92XymdE#123456
obteniendo semilla...
semilla de qlink creado: f88cHwhwVD -> 1496390487982
el qlink anterior al recientemente creado fue generado, aproximada-
mente, entre Mon May 15 21:44:53 2017 y Mon May 15 21:46:32 2017
$
```

Listado 2. Salida de la ejecución del script de prueba relativo al punto 2.2.1.

```
$ python punto_2_2_2.py
qlink generado (ejemplo):
http://qlink/two/lQ9PvYWZrJGyTmqEvPeiX#123456 dn obtenido:
8576041210
obteniendo semilla...
semilla de qlink generado: lQ9PvYWZrJ -> 1496555563119
```

```

qmlink para timing #0 generado:
http://qmlink/two/IoePfzTwYziTuniQBUMreN#123456 dn: 4523295259
qmlink para timing #1 generado:
http://qmlink/two/lvaaKCsFbIUb8tBgSAuIqk#123456 dn: 7097969091
[...]
qmlink para timing #9 generado:
http://qmlink/two/cNR7bGgGzjdDPJ9hRN4N4t#123456 dn: 6566803103
diferencia entre semilla para qmlink/dn #0: 1110
diferencia entre semilla para qmlink/dn #1: 192
diferencia entre semilla para qmlink/dn #2: 103
[...]
diferencia entre semilla para qmlink/dn #8: 95
diferencia entre semilla para qmlink/dn #9: 134
probando semilla: 1496555563214 trk: 3000169741 (0/1015)...
status: FAIL
probando semilla: 1496555563215 trk: 5334132004 (1/1015)...
status: FAIL
[...]
probando semilla: 1496555563652 trk: 7677823134 (438/1015)...
status: FAIL
probando semilla: 1496555563653 trk: 8097925954 (439/1015)...
status: OK
$

```

Listado 3. Salida de la ejecución del script de prueba relativo al punto 2.2.2.

```

$ python punto_2_2_3.py
se estimarán valores siguientes considerando salida conocida:
f7f195a46835dd93 f37e44f2afe372308cb09ca5c102323d
BROWSER: chrome
resolviendo...
modelo resuelto:
[c1480396913101804 = True,
 ostate0 = 14459375271365247648,
 ostate1 = 12664944705117373901,
 c149958108320120 = True,
 c1353346471225766 = True,
 c476870980001021 = True,
 c525365255680336 = True,
 c1015901882049230 = True]
siguientes 200 x 4bytes estimados: 7af3b8eac[...]
encontrados siguientes 16bytes generados en browser? True
$

```

Listado 4. Salida de la ejecución del script de prueba relativo al punto 2.2.3.

Post-Quantum Cryptography Using Hyper-Complex Numbers

Jorge Alejandro Kamlofsky¹ – Juan Pedro Hecht²

¹ CAETI - Universidad Abierta Interamericana.
Av. Montes de Oca 725 – Buenos Aires – Argentina.
Jorge.Kamlofsky@uai.edu.ar

² Universidad de Buenos Aires, Facultades de Ciencias Económicas, Ciencias Exactas y Naturales e Ingeniería. Maestría en Seguridad Informática, Buenos Aires, Argentina.
phecht@dc.uba.ar

Abstract. Encrypted communications are performed using symmetric ciphers, which require asymmetric cryptography for safe initiation. Asymmetric cryptography was seriously weakened after the presentation of Shor's algorithm (1997) and others for quantum computers. New algorithms are generalized as post quantum cryptography. Asymmetric cryptography based on non-commutative algebra is a growing trend arising as a solid choice that strengthens these protocols. Hyper-complex numbers generated by the Cayley-Dickson construction forms non-commutative algebras. This paper focus on the use of these numbers in post-quantum cryptography.

Keywords: octonion's cipher, quaternion's cipher, non-commutative cryptography, post-quantum cryptography, PCQ with hyper-complex numbers.

1 Introducción

1.1 Trabajos Relacionados

La criptografía se refiere a la ciencia o arte de diseñar criptosistemas. Su principal propósito es la protección de los intereses de las partes de una comunicación. Un criptosistema es un dispositivo diseñado para brindar tal protección. Se encripta la información de manera que solo pueda ser utilizada por quien esté autorizado [1]. Se distinguen dos instancias: el intercambio seguro de claves, y el encriptado y desencriptado del mensaje [2]. La criptografía puede dividirse, en dos ramas: simétrica, que encripta y desencripta los mensajes con una única clave compartida y asimétrica, que por un lado logra el intercambio seguro de claves y por el otro el cifrado de mensajes usando una clave pública y otra privada, hoy día con herramientas de la teoría de números.

Whitfield Diffie y Martin Hellman fueron los pioneros de la criptografía asimétrica: en 1976 presentaron el revolucionario concepto de criptografía de clave

pública [3] cuya seguridad radica en el problema de la intratabilidad del logaritmo discreto [4] (*DLP: Discrete Logarithm Problem*). Sin embargo, por facilidad de implementación práctica, RSA [5] es hoy el esquema criptográfico de clave pública más usado. Su seguridad radica en el problema de la intratabilidad de la factorización de grandes números enteros (*IFP: Integer Factorization Problem*). Es responsable en gran parte, de la seguridad en Internet, en las transacciones bancarias electrónicas o en la firma digital de correos electrónicos [6], entre otros.

En 1997 Peter Shor presentó un algoritmo que reduce drásticamente la complejidad computacional del problema IFP mediante una computadora cuántica [7]. Los trabajos de Kitaev [8] y Proos-Zalka [9] presentaron ataques eficaces a los problemas DLP y DLP para curvas elípticas también mediante computadora cuántica. A pesar que este dispositivo aún no se había inventado, la sola existencia de estos algoritmos, debilitó notoriamente a esta rama de la criptografía.

Desde inicios de este siglo ha crecido el interés por el desarrollo de criptosistemas asimétricos alternativos que sean resistentes a ataques de complejidad sub-exponencial y ataques a través de computadora cuántica [10 – 12]. A estos esquemas se los denomina colectivamente como criptografía post-cuántica (PQC: de sus siglas en inglés) [13]. Se han desarrollado diversas líneas de investigación PQC, conocidas hoy día como criptografía basada en polinomios multicuadráticos, en hashings, en retículos y en códigos [10, 14 - 16]. Sin desmedro de esas corrientes de estudio, se han buscado soluciones vinculadas con estructuras algebraicas no conmutativas y no asociativas [17]. Nuestros estudios se ubican en esta última corriente.

En general se usan estructuras algebraicas en las que el producto no cumple la propiedad conmutativa. Gracias a su naturaleza algebraica, a esta línea se la denomina criptografía no conmutativa [18]. No se conocen ataques concretos a esta línea que hayan logrado resultados exitosos. Dentro de esta línea, en [19] se presentó un esquema de distribución de claves Diffie-Hellman basado en un anillo de polinomios matriciales, inspirado en [20].

Dentro de esta clasificación, un subconjunto de elementos algebraicos es el conjunto de los números hipercomplejos creados a partir de la construcción de Cayley-Dickson: cuaterniones [21], octoniones [22], y otros. Dicha construcción [23] tiene la peculiaridad que en cada álgebra generada, se pierde una propiedad. A partir del álgebra de cuaterniones, el producto no es conmutativo. Criptosistemas no conmutativos con números hipercomplejos se presentaron en [24 - 26].

Como este proceso permite formar infinidad de álgebras no conmutativas, es de suponer que podrían desarrollarse infinidad de criptosistemas no conmutativos. Sin embargo, propiedades que se pierden en instancias de esta construcción hacen que ciertas álgebras carezcan de interés criptográfico. En este trabajo, se presenta un enfoque acerca del uso de números hipercomplejos en PQC.

1.2 Motivación y Alcance

Pareciera ser que la criptografía no conmutativa usando números hipercomplejos obtenidos a partir de la construcción de Cayley-Dickson no tuviera límite. Sin

embargo, como a partir de elementos de dimensión 16 (sediniones) las álgebras generadas presentan divisores de cero, su interés criptográfico quedaría así acotado. Este trabajo muestra virtudes y límites en el uso de números hipercomplejos en PCQ.

1.3 Objetivo del Trabajo

La finalidad de este trabajo es presentar un enfoque que muestre las ventajas del uso de números hipercomplejos en criptografía no conmutativa pero a su vez pretende acotar su uso al armado de esquemas criptográficos con cuaterniones y octoniones.

1.4 Relevancia del tema

En los últimos lustros se han realizado desarrollos en línea con PQC. En 2016 la NIST declaró a PQC como un tema de interés, y llamó a presentación de propuestas para la implementación de algoritmos PCQ eficiente [27].

1.5 Estructura del Trabajo

En la Sección 2 se presenta el marco teórico. En la Sección 3 se presenta el enfoque propuesto con datos experimentales. La Sección 4 contiene las conclusiones.

2 Marco Teórico

2.1 La Criptografía y la Seguridad de las Comunicaciones

Nociones Básicas de la Criptografía Simétrica. A partir del inicio del cifrador, el mensaje se transforma en el punto de emisión mediante operaciones matemáticas de manera que sea imposible de interpretarlo mientras viaja en el canal inseguro, o bien su costo en tiempo y/o recursos sean tan altos que su descubrimiento carezca de sentido. Se usan algoritmos criptográficos muy robustos que permiten que la información se encripte bit a bit o en grupos de n-bits permitiendo que puedan cifrarse comunicaciones en tiempo real [4], sin que la encriptación demore la transmisión del mensaje. Usan la misma clave para el cifrado y descifrado. Pueden iniciarse con claves de 128 bits, aunque recientes recomendaciones [28] imponen una longitud de claves para cifradores simétricos de 192 y 256 bits (AES: advanced encryption standard NIST).

Nociones Básicas de Criptografía Asimétrica. Usa elementos públicos que se comparten, y elementos privados que se mantienen en secreto. Tradicionalmente usan propiedades y operaciones de aritmética modular en estructuras algebraicas de

campos de números enteros. Así se brindó soluciones al problema de presentar en forma segura claves para su uso en cifradores simétricos: mientras que mediante RSA [5] y ElGamal [29] se pueda enviar una clave simétrica cifrada a otro usuario que descifra usando su clave pública, con Diffie-Hellman [3] ambas partes pueden generar la misma clave intercambiando elementos.

Amenaza a la Criptografía: El Algoritmo de Shor y la Computación Cuántica.

En 1997 Peter Shor presentó un algoritmo para computación cuántica basado en la transformada rápida de Fourier que logra resolver en tiempo polinómico el problema IFP [7]. Es decir, permite reducir drásticamente la complejidad del problema (considerado de clase NP) a niveles atacables [30]. Una computadora cuántica usa qubits en lugar de bit. Un qubit posee los estados clásicos 0 o 1 en superposición, realizando cómputos en forma paralela sobre ambos y que finalmente se resuelven en bits clásicos por decoherencia. Por ello, se puede realizar una cantidad exponencial de operaciones en paralelo en relación exponencial con la cantidad de qubits del computador cuántico. La computación cuántica prácticamente arrasa con todo lo conocido en la criptología actual: con ello desaparecen de escena prácticamente todos los criptosistemas de clave pública actualmente en uso.

Criptografía Post-Cuántica Basada en Álgebra no Conmutativa. En este trabajo, se utilizan estructuras de anillos de polinomios de matrices cuadradas o de cuaterniones, entre otros, con elementos finitos, por lo tanto, su seguridad radica en la complejidad del tratamiento del problema DLP. Nuestros esquemas se basan en la dificultad de resolver el problema SDP (*Symmetric Decomposition Problem*) [19] en un anillo no conmutativo de polinomios matriciales. Desde el punto de vista criptográfico, solo se necesita estar seguro que no exista fórmula que permita reducir la complejidad del problema DLP (incluso con computadora cuántica). Y esto está garantizado ya que en los anillos no conmutativos no existe forma conocida de relacionar el determinante de una matriz o bien sus autovalores con la potencia de la matriz [31], parte de la clave privada, independientemente de la cantidad de qubits que pudiera tener una computadora cuántica que ejecute el ataque.

2.2 Los Numeros Hipercomplejos

Son una extensión de los números complejos construidos mediante herramientas del álgebra abstracta, tales como terniones, cuaterniones, tesarines, octoniones, sedeniones, y demás. Tienen más de una componente compleja. Muchos conforman álgebras que carecen de interés dentro de nuestro análisis [32].

La Construcción de Cayley-Dickson. La construcción de Cayley-Dickson produce una secuencia de álgebras sobre el cuerpo de los números reales. Cada álgebra producida posee el doble de la dimensión de la anterior. En cada etapa, al generarse una nueva álgebra, esta pierde una propiedad algebraica específica [22].

Proceso General. Como lo notó Hamilton [21], el complejo $a + bi$ puede entenderse como el par de números reales (a, b) . La suma se hace componente a componente y el producto con otro complejo (c, d) se hace de la siguiente forma:

$$(a, b).(c, d) = (ac - db, ad + cb)$$

También puede definirse el conjugado de un complejo como: $(a, b)^* = (a, -b)$. De esta forma ya se han obtenido los números complejos, a partir de los reales. Los cuaterniones pueden definirse de forma análoga. Es decir, pueden ser pensados como un par de números complejos y obtener suma, producto, conjugación y norma. Repitiéndose el proceso, se pueden hallar las álgebras de dimensión superior.

Propiedades Perdidas en el Proceso. Al aplicarse el proceso de Cayley-Dickson al cuerpo de los reales y obtenerse el cuerpo de los complejos, se pierde la relación de orden. Su aplicación sobre el cuerpo de los complejos forma el álgebra de cuaterniones: les hace perder la propiedad conmutativa. Los cuaterniones forman entonces, estructura de anillo de división. Aplicando el proceso sobre el anillo de cuaterniones se genera el álgebra de octoniones: se pierde la propiedad asociativa. Los octoniones, forman un álgebra de división normada. La conformación del álgebra de Sedinitiones (16 componentes) le quita la propiedad de ser álgebra de división, lo que se hereda a las posteriores álgebras superiores generadas tras aplicaciones sucesivas del proceso, haciéndolas poco interesantes para su uso en PQC.

Finalmente, el uso de números hipercomplejos que forman álgebras no conmutativas, para su uso en PQC se limita entonces, a cuaterniones y octoniones.

Anillo de Cuaterniones. Un anillo $(A; +; \cdot)$ es una estructura algebraica (un conjunto A con las operaciones suma y producto) donde $(A; +)$ forman estructura de grupo, y $(A; \cdot)$ de semigrupo. Y el producto es distributivo (por los dos lados) respecto de la suma. Un anillo es no conmutativo si no se verifica la propiedad conmutativa entre todos los elementos de A para la operación producto.

El primer anillo de división (cuyos elementos no nulos son inversibles) no conmutativo fue el anillo de los cuaterniones [21].

Definición: Cuaternión. Sea $(H; +; \cdot)$ un anillo conmutativo con unidad. Un cuaternión q con coeficientes en H es un número hipercomplejo de la forma: $q = a + b.i + c.j + d.k$, donde $a, b, c, d \in \mathbb{R}$; i, j, k son unidades imaginarias que verifican que: $i^2 = j^2 = k^2 = -1$, y además: $i.j = -j.i = k$; $j.k = -k.j = i$; $i.k = -k.i = j$.

Operaciones Básicas: suma, resta y producto de un escalar por un cuaternión se realiza de la misma forma que con cualquier vector de 4 dimensiones. Para el producto entre cuaterniones debe tenerse en cuenta el producto entre unidades imaginarias. Puede encontrarse información más detallada en [24].

Álgebra de Octoniones. Los octoniones forman un álgebra de división normada. Fueron descubiertos por John T. Graves en 1843, e independientemente por Arthur

Cayley. Se usaron recientemente en Física, para dar sustento a la teoría de cuerdas [33]. El conjunto de Octoniones no nulos con la operación producto forman una estructura de bucle (loop) de Moufang no asociativo [34].

Definición: Octonión. Es una expresión de la forma $o = \alpha_0 + \alpha_1.e_1 + \alpha_2.e_2 + \alpha_3.e_3 + \alpha_4.e_4 + \alpha_5.e_5 + \alpha_6.e_6 + \alpha_7.e_7$ donde α_i son números reales y e_i unidades imaginarias. Puede ser representado como un vector de 8 componentes.

Operaciones Básicas: Suma, Resta y Producto de un Escalar por un Octonión se realiza de la misma forma que con cualquier vector de 8 dimensiones. Para el producto entre octoniones debe tenerse en cuenta el producto entre unidades imaginarias [22]. Puede encontrarse información más detallada en [26]. El producto de octoniones es cerrado, tiene elemento neutro, pero no cumple las propiedades conmutativa ni asociativa.

3 Implementaciones Criptográficas con Números Hipercomplejos

Se limitan al uso de cuaterniones y octoniones. En esta sección se hace referencia a esquemas de intercambio de claves Diffie-Hellman: una vez logrado esto, es fácil derivar a otros esquemas asimétricos [20]. En [24] se presentó un esquema Diffie Hellman inspirado en [19], implementado con polinomios de cuaterniones: llamado DHCQ. Junto con la mejora propuesta en [25] se logró reducir notablemente los tiempos de proceso. En [26] se presenta un esquema similar implementado con octoniones: llamado DHECO. Ambos sistemas se los denomina compactos ya que son aptos para procesadores de pequeño porte gracias a no requerir de librerías de precisión extendida, y se presentan versiones para procesadores de 8 bits, 16 bits y 32 bits. DHECO funciona solo con operaciones suma-producto, lo que permite que sea implementado en procesadores muy elementales. Promete ser bastante más robusto. Pero sus tiempos de procesamiento son mayores. Ambos esquemas son inmunes a ataques subexponenciales y de computadora cuántica.

En esta sección se presentan los conceptos básicos de ambos esquemas con datos experimentales. Más detalles pueden obtenerse de los trabajos originales [24 - 26].

3.1 DHCQ: Esquemas de Intercambio de Claves Usando Cuaterniones

El Protocolo. Alice elige dos cuaterniones aleatorios A y B , con elementos de $Z_{2^k \times 8}$ (con $k = 1, k = 2, k = 4$) y los normaliza: q_A y q_B . Elige como clave privada dos números enteros aleatorios m y n en Z_{16} , y un polinomio entero $f(x)$ con coeficientes y exponentes en Z_{16} tal que $f(q_A) \neq 0$ y envía a Bob por el canal inseguro los elementos q_A y q_B . Bob elige como clave privada dos números enteros aleatorios r y s en Z_{16} , y un polinomio entero $h(x)$ con coeficientes y exponentes en Z_{16} tal que $h(q_A) \neq 0$. Ambos realizan las normalizaciones de $f(q_A)$ y $h(q_A)$: $f'(q_A)$ y $h'(q_A)$. Alice calcula su

token: $r_A = f(q_A)^m \cdot q_B \cdot f(q_A)^n$. Bob calcula el suyo: $r_B = h'(q_A)^r \cdot q_B \cdot h'(q_A)^s$; y se los intercambian para el cálculo de las claves: $k_A = f(q_A)^m \cdot r_B \cdot f(q_A)^n$ (Alice), $k_B = h'(q_A)^r \cdot r_A \cdot h'(q_A)^s$ (Bob). Se modularizan: $K_A = k_A \cdot 2^{k \cdot 16} \pmod{2^{k \cdot 16}}$, $K_B = k_B \cdot 2^{k \cdot 16} \pmod{2^{k \cdot 16}}$ con $K_A = K_B$.

3.2 DHECO: Esquemas de Intercambio de Claves Usando Octoniones

Resumen del Protocolo. Alice elige dos octoniones aleatorios o_A y o_B , con elementos de $Z_{k \cdot 8\text{bits}}$ (con $k = 1, 2$ o 4). Elige como clave privada dos números enteros aleatorios m y n en Z_{16} y un polinomio entero $f(x)$ con coeficientes y exponentes en Z_{16} tal que $f(o_A) \neq 0$. y los envía a Bob por el canal inseguro. Bob elige como clave privada dos números enteros r y s en Z_{16} , y un polinomio entero $h(x)$ con coeficientes y exponentes en Z_{16} tal que $h(o_A) \neq 0$. Alice calcula su token: $r_A = (f(o_A)^m \cdot o_B) \cdot f(o_A)^n$. Bob calcula el suyo: $r_B = (h(o_A)^r \cdot o_B) \cdot h(o_A)^s$ y se los intercambian para el cálculo de las claves: $K_A = (f(o_A)^m \cdot r_B) \cdot f(o_A)^n$ (Alice), $K_B = (h(o_A)^r \cdot r_A) \cdot h(o_A)^s$ (Bob). Luego: $K_A = K_B$.

La figura 1 muestra un esquema del protocolo propuesto.

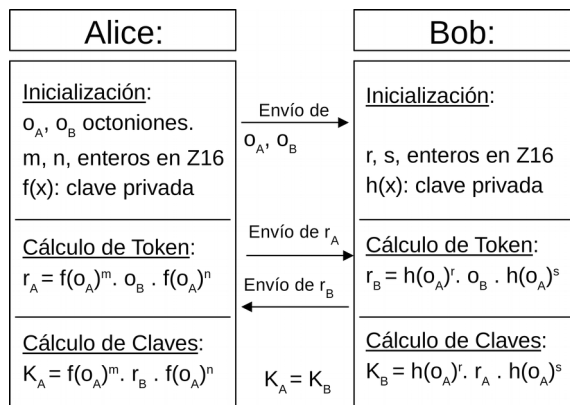


Fig. 1: Esquema del protocolo propuesto.

3.3 Seguridad Adicional.

Si bien el uso tanto de cuaterniones como de octoniones garantiza la ausencia de divisores de cero, el álgebra en el anillo de enteros (en las componentes de los cuaterniones o de los octoniones) posee divisores de cero si el módulo no es primo, presentando efectos no deseados. Para evitar este inconveniente, se trabaja con módulo: el mayor primo menor a $K \cdot 8\text{bits}$. Esto es: si $k = 1$, las componentes del cuaternión u octonión son elementos en Z_{251}^* . Si $k = 2$ sus componentes son elementos en Z_{65521}^* . Y si $k = 4$ sus componentes son elementos en $Z_{4294967279}^*$.

La seguridad de los protocolos presentados en [24 - 26] se ha fortificado: En lugar de que ALICE publique los enteros m y n , aquí se los usa como parte de su clave

secreta. BOB por su parte elige sus propios enteros secretos r y s en lugar de repetir los m , n que publicó ALICE. Ello dificulta aún más el problema SDP.

3.4 Datos Experimentales

Equipamiento Usado. El computador usado tiene un procesador Intel® Core™ i3-2328M CPU @ 2.20GHz \times 4 y 3,7 GiB RAM. Sistema operativo Kali GNU/Linux 64-bit, con un núcleo Debian. Los algoritmos fueron programados en Python 2.7.10.

Versiones Propuestas. Cada valor de k genera una versión para cada criptosistema: 8, 16 y 32 bits. Se presentan en 8 bits: DHCQ8 y DHECO8. En 16 bits: DHCQ16 y DHECO16. En 32 bits: DHCQ32 y DHECO32.

Resultados Experimentales. Se presenta una comparación de los tiempos de ejecución para la obtención de 1000 claves de 256 bits entre los diferentes esquemas y versiones. La tabla 1 muestra los resultados experimentales.

Tabla 1. Tiempos de ejecución para la obtención de 1.000 claves de 256 bits.

N° Test	CPU Time (s)								
	DHCM8 8-bits	DHCM16 16-bits	DHCM32 32-bits	DHCQ8 8-bits	DHCQ16 16-bits	DHCQ32 32-bits	DHECO8 8-bits	DHECO16 16-bits	DHECO32 32-bits
1	4,9232	2,6201	4,9177	3,7838	1,9273	1,0624	10,8278	5,4771	4,9928
2	4,8215	2,6124	5,0230	3,8360	1,7862	0,9280	10,8237	5,4119	5,4673
3	4,9672	2,5244	4,9549	3,5639	1,8374	0,8861	11,0357	5,5073	5,0466
4	4,9166	2,6128	4,9277	3,6088	1,8602	0,9133	11,0351	5,4587	4,9441
5	4,8174	2,6365	5,0176	3,7249	1,8298	0,9343	11,2985	5,5615	4,9406
6	4,9174	2,6104	4,9433	3,6658	1,8519	0,9102	10,8646	5,5769	5,0594
7	4,9207	2,5530	5,2199	3,6511	1,7609	1,0421	11,1051	5,5226	4,9823
8	4,8164	2,6122	4,9217	3,7963	1,8615	0,9048	11,0446	5,4649	5,0472
9	4,8456	2,6124	5,0184	3,6306	1,8013	0,9627	11,1739	5,4660	4,8783
10	4,8163	2,5378	5,0295	3,6095	1,8660	0,9449	11,4447	6,2083	4,9739
Promedio	4,8762	2,5932	4,9974	3,6871	1,8383	0,9489	11,0654	5,5655	5,0333

3.5 Ventajas de los Esquemas Propuestos

Mayor Velocidad del Esquema DHCQ. En todas sus versiones presenta menores tiempos de ejecución en comparación con implementaciones con matrices [19].

Mayor Robustez del Esquema DHECO. Los esquemas basados en álgebra no asociativa son aún más resistentes que los cripto-sistemas no conmutativos. Por tratarse de un sistema algebraico no conmutativo y no asociativo, los octoniones no admiten representación matricial. Las potencias se deben calcular en forma exclusivamente recursiva, impidiendo la localización de potenciales generadores u

órdenes multiplicativos por medio del algoritmo de Shor, base del ataque cuántico que requiere que la estructura sea un campo, es decir producto asociativo.

Esquemas Aptos para Procesadores de Pequeño Porte. No requerir librerías de precisión extendida los hace apto para procesadores pequeños. En especial, DHECO solo usa operaciones suma-producto, haciéndolo apto para procesadores elementales.

Inmunidad Frente a Ataques de Complejidad Sub-Exponencial o de Computadora Cuántica: No se conocen aún ataques de estos tipos a estructuras algebraicas no conmutativas que hayan sido efectivos y que debiliten su seguridad.

4 Conclusiones

La construcción de Cayley-Dickson a partir de los números complejos genera álgebras de números hipercomplejos que verifican el no cumplimiento de la propiedad conmutativa, lo cual es de interés en PCQ. Se concluye que solo las álgebras de cuaterniones y octoniones poseen interés criptográfico.

Las soluciones presentadas en [24 - 26] y aquí mencionadas, son esquemas de intercambio de claves Diffie Hellman implementados con cuaterniones y octoniones. Sin embargo, en general, lograr otros esquemas asimétricos una vez logrado uno de ellos, no suele presentar grandes dificultades.

Referencias

1. Barreno, Marco A.: The Future of Cryptography Under Quantum Computers. Dartmouth College Computer Science Technical Reports (2002).
2. Marrero Travieso, Yran: La Criptografía como elemento de la seguridad informática. ACIMED 11.6 (2003).
3. Diffie W., Hellman M.E: New directions in cryptography. IEEE Transactions on information theory, 22, 644-654, (1976).
4. Menezes, Alfred J., Paul C. Van Oorschot, and Scott A. Vanstone: Handbook of applied cryptography. CRC press (1996).
5. Rivest, Ronald L., Adi Shamir, and Len Adleman: A method for obtaining digital signatures and public-key cryptosystems. Communications of the ACM 21.2, 120-126. (1978)
6. Navarro Robles, Pedro Ramón: Intercambio de claves sobre anillos no conmutativos: End $(\mathbb{Z}_p \times \mathbb{Z}_p^2)$ y extensiones. (2014).
7. Shor P.: Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer, SIAM J. Comput., 5, 1484-1509 (1997).
8. Kitaev A.: Quantum measurements and the abelian stabilizer problem, Preprint arXiv/quant-ph., 9511026 (1995).
9. Proos J., Zalka C.: Shor's discrete logarithm quantum algorithm for elliptic curves. Quantum information and computation, 3, 317-344 (2003).

10. Bernstein, Daniel J., Johannes Buchmann, and Erik Dahmen.: Post-Quantum Cryptography. (2007).
11. Magliveras S.S., Stinson D.R., van Trung T.: New approaches to designing public key cryptosystems using one-way functions and trapdoors in finite groups. Technical Report CORR, 2000-2049 (2000).
12. Shpilrain V., Zapata G.: Combinatorial group theory and public-key cryptography, Preprint arXiv/math.gr, 0410068 (2004).
13. Barreto, P. et al.: Introdução à criptografia pós-quântica, Minicursos do XIII Simpósio Brasileiro em Segurança da Informação e de Sistemas Computacionais, SBSeg, (2013).
14. Kaya K. Open Problems in Mathematics & Computational Science, Springer Verlag, (2014).
15. Chen L. et al, NISTIR 8105, Report on Post-Quantum Cryptography, NIST. [En Línea], (2006). Disponible en: <<http://nvlpubs.nist.gov/nistpubs/ir/2016/NIST.IR.8105.pdf>> . Fecha de consulta: 20/04/2017.
16. Moody, D.: Update on the NIST Post-Quantum Cryptography Project, 2016 <http://csrc.nist.gov/groups/SMA/ispab/> (consulted April 20, 2017).
17. Kalka, A.: Non-associative public-key cryptography. arXiv:1210.8270, (2012).
18. Geritzen L. et al (Editors): Algebraic Methods in Cryptography, Contemporary Mathematics, AMS, Vol. 418, (2006)
19. Hecht J.: Un modelo compacto de criptografía asimétrica empleando anillos no conmutativos. V Congreso Iberoamericano de Seguridad Informática, Montevideo (2009).
20. Cao Z., Xiaolei D., Wang L. New public-key cryptosystems using polynomials over non-commutative rings. Preprint arXiv/cr, eprint.iacr.org/2007/009.pdf (2007).
21. Hamilton, W. R. Lectures on Quaternions: Containing a Systematic Statement of a New Mathematical Method. Hodges and Smith, (1853).
22. Baez, J.: The Octonions. Bulletin of the American Mathematical Society 39.2 (2002)
23. Kornilowicz, Artur.: Cayley-Dickson Construction. Formalized Mathematics 20.4 (2012)
24. Kamlofsky J., Hecht J., Abdel Masih S., and Hidalgo Izzi, O.: A Diffie-Hellman Compact Model Over Non-Commutative Rings Using Quaternions. VIII Congreso Iberoamericano de Seguridad Informática, Quito (2015).
25. Kamlofsky, J.: Improving a Compact Cipher Based on Non Commutative Rings of Quaternion. En XXII Congreso Argentino de Ciencias de la Computación CACIC, (2016).
26. Kamlofsky, J., Hecht, J. y Abdel Masih, S.: Post-Quantum Cryptography: An Elementary and Compact Key Exchange Scheme Based on Octonions. Presentado en: IX Congreso Iberoamericano de Seguridad Informática CIBSI. (2017).
27. National Institute of Standards and Technology, Information Technology Laboratory – Computer Security Division. “Post-Quantum Crypto Project”. [En línea], (2016) Disponible en: <<http://csrc.nist.gov/groups/ST/post-quantum-crypto/>>. Fecha de consulta: 11/06/2017.
28. Barker E., Roginsky, A.: Transitions: Recommendation for Transitioning the Use of Cryptographic Algorithms and Key Lengths. NIST Special Publication 800-131A (2015).
29. EGamal, Taher. A Public Key Cryptosystem and a Signature Scheme Based on Discrete Logarithms. En Advances in cryptology. Springer Berlin Heidelberg, pp. 10–18 (1984).
30. Hecht, JP.: Fundamentos de Computación Cuántica. Editorial Académica Española, (2005).
31. Eftekhari, M.: A Diffie–Hellman key exchange protocol using matrices over noncommutative rings. Groups-Complexity-Cryptology, 4(1), pp. 167–176 (2012)
32. Kenneth O. May.: The impossibility of a division algebra of vectors in three dimensional space; American Mathematical Monthly, Vol 73, No. 3 (1966) 289-291.
33. Baez, J, et Huerta, J.: Des octonions pour la théorie des cordes. Pour la science 406 (2011).
34. Belousov, V. D.: "Moufang loops", Hazewinkel, Michiel, Encyclopedia of Mathematics, Springer (2001).

Two-Way Continual Authentication Model

Carlos Fabián Lugani

LIA - Laboratorio de Informática Aplicada – Sede Atlántica –
Universidad Nacional de Río Negro
clugani@unrn.edu.ar

Summary. Authentication is a process by which a person using a personal computer, information system or device, defines its identity, and therefore, such system grant access based on predefined permissions. The use of biometric devices, IoT devices, wearable technology and other similar devices, define present authentication methods as weak and requires a level of mutual authentication so that both actors trust each other granting who they claim or seem to be. This paper addresses that information systems and/or devices that should validate their identity in the same way users do. Additionally, due to the increasing information security incidents, there is a risk that one of the actors, during the connection process, might be an impostor instead of a truly actor. For this reason, the authentication method should be a continuous and ongoing process in order to counteract with the security risks mentioned above.

Key words: authentication, identity management, security, IoT

1 Introduction

Need-to-Use is the primary baseline that defines who can and who cannot access and use a specific resource. At first, there is a person with a genuine interest who is allowed to use a system or a personal computer. Normally, it is well understood that the person knows where such system is, because the system is always located in the same place or can be unambiguously identified. There are systems that are not located at a defined site, but they normally provide a service for multiple devices and they are constantly changing. This is why the difficulty in the identification of systems will increase due to the different scenarios and the technological progress.

Normally, authentication involves the action that a person executes on an information system so that such system gives access to a specific information or allows to execute several predefined actions. All the actions that a person can execute have been previously programmed or enabled and therefore, the scopes of those actions are already known.

The verification of the user's identity more frequently used implies a simple method based on a username and a password. However, this identity verification is a very basic method because there does not exist strong verifications to validate if the password used truly belongs to the person who claims to be. Even though there are several

biometric technologies in order to increase the level of trust to verify a person's identity, they have not been widely adopted due to different factors, such as:

- Limitations for achieving cost-benefit implemented solutions.
- Lack of knowledge from actors or lack of information assessment leading to a loss of confidentiality, integrity and availability.
- Overconfidence when no incidents related to information security are occurring (originated by intruders or intrusion actions), or lack of information related to previous registered incidents caused by intruders in the past.

Taking into account all the issues mentioned above, this paper focus on a new approach on the authentication concept, by considering it as a method where every actor has the responsibility to provide confidence over other actors about its authenticity.

This is the reason why present systems are becoming more complex and not only do they have data about people and companies, but also they can control key functions of a human being as its heart system [1] or when people's lives depend on such key systems as: airplane flight self-control, drones traffic, automated control of robots, street traffic and motorized self-driving vehicle [2].

2 Major Difficulties

The model proposed can be better understood with the following example where a variety of actors with different interests can be appreciated regarding the authentication process among stakeholders.

- In the first place there is a vehicle equipped with state-of-the-art technology such as: driver/owner identity recognition, self-driving technology, access to the Internet and the traffic control information system, medical and security emergencies. There would be, as well, associated systems providing data which would identify the car and its driver as part of the traffic control system network.
- The driver gets out of his car at a shopping centre, he gives his car specific instructions for self-parking. When the driver leaves the shopping centre, he indicates his car to wait for him at the front door by means of a device connected to the Internet (cell phone or a wearable device).
- This particular example presents several types of associated identity verifications:
 - o The car system must verify that the real owner is sending the message.
 - o The car system must verify that the real owner is approaching the car so that letting him open the car's door and,
 - o The owner must verify that the car is his own and no other person's car. This is the reason why computer/system authentication is the primary baseline for the owner's physical and personal security.
- The examples mentioned above are part of the authentication method established between two actors, however there would exist some other systems such as those one person can interact with, as the following example:

- You are driving your car and the system gets a message that there is a traffic jam and the car system gives the driver another alternatives to make the better driving choice.
- In this case, it is important to take into account the authentication process from the car to the traffic control system and the authentication from the traffic control system to the car. Any kind of mistake in these authentication process would make the system unreliable putting the actors at risk.
- It is also possible that the driver has a car accident leaving him unconscious, and in this case, the car's system sends information about the driver's health condition to a medical emergency system. This is the reason why the verification of the information reliability and user's identity is a primarily baseline.

According to the items mentioned above, the authentication process not only should it be only in one-way, but also a two-way technique is required, because both actors should verify the other actor effectively and unmistakably. This example shows the first concept of this paper that deals with the authentication process on both directions or named as two-way authentication technique.

Continuing with the difficulties exposed above it should be taking into account an important issue related to the information security point of view. Regarding information security, it is possible that a device/system or part of them can be forged once the communication was established. This can happened due to two reasons, the presence of multiple devices interfering instructions or signal messages that are not for the proper recipient, or the presence of fake devices replacing the original ones by the use of fraud, scams or even more serious situations such as spying or terrorism activities.

2 People Identification

In order to identify people into personal computers and different devices, there are several methods of authentication. There are software-based and hardware-based identification systems. These identification systems are being used when a person need to access a personal computer, an ATM, some Internet of Things devices, when accessing a building or a country. For all these cases, the identity must by verified. The identity is assessed and verified according to the established security levels. Therefore, every authentication action must be previously analysed regarding its associated risks. It is important to analyse these concepts in order to continue with other advanced or strong authentication methods or considering a combination of different methods. Particularly, identity verification risks should be assessed over weak methods of verification leading to an information leakage and also unauthorized physical and/or logical access. In this case, there existed not only the error of granting an improper access, but the user confidence of having granted this access to the proper user. This example shows that this authentication control is deficient.

For example, when a person logs in a personal computer he/she uses a password that could be a word, passphrase or a pattern drawn on the cell phone display – also some devices have fingerprints readers. Therefore, we have two types of authentication methods.

At a primary level of authentication, a person can be identified by some information he knows, although someone else can obtain and/or guess this information by different tricky methods. Consequently, the identity of a person cannot be truly verified.

In stronger levels of authentication, where is possible to identify different parts of the body, the authentication is called biometric authentication. Although this method is safer regarding its verification phase, it shows some difficulties, for example some people's fingerprints impressions are difficult to be clearly identified, and in the case of facial recognition systems, such systems can be modified, or when regarding to voice recognition systems, the voice to identify can be recorded for future cyber-crimes. We have to take into account that biometric readers are sensitive to external agents such as water or dirt, making them useless and, therefore, this type of authentication cannot be used in every situation.

The following items show the different techniques that can be used for the present example [3]:

Digital fingerprints: this is the oldest identification technique, fingerprints reading and error recognition rates have been improved, and however, this technique cannot be used in every case.

Iris recognition [4]: this is the most accurate method for recognition because the algorithms used lead to an efficient matching search. The iris does not change and it is protected from accidents or dirt compared to other parts of the body.

Face recognition: this system is not frequently used because it presents different kind of false positive authentications due to several factors as deficient amount of light, distance and age as well.

Hand geometry recognition: this system reads the whole hand including fingers length, wrinkles and, in recent investigations, the shape of veins. [5]. However, there exist some problems regarding people suffering from specific conditions where hands or fingers cannot be accurate recognised making it difficult an accurate reading of the hand, also hand readers are too big.

Voice recognition: there are multiple applications and researches where voice is recognised by the means of different patterns. A training process is required before starting with this authentication method but it presents a high accuracy rate of 100%. [6] [7] [8].

Smell recognition: this method does not need any kind of physical contact but the person should be at a short distance from the reader making this a secondary verification phase [9].

Some other verification methods can be wearable devices. Such devices are electronic devices attached to clothes or accessories and they interact with the person carrying them. These electronic devices can be connected to some other devices in order to carry out an action previously programmed. Some examples are smartwatches, sneakers with GPS system, bracelet controlling vital signs, or clothing.

Taking into account the examples mentioned above, it is necessary to carry out a detailed analysis on the use of these devices or biometric systems for personal recognition. Such analysis should bear in mind that some people are reluctant to these type of biometric recognition by considering them an intrusive method. In this case, the

best verification method could be the use of different verification methods, so that the error rate would not be significant.

3 Components verification

Normally, the systems do not identify themselves to the people. In fact, people consider that if the system is where is supposed to be, that is enough for verifying its identity. For example, when using an ATM located at a bank, the user considers it as a trusted ATM. However, sometimes there are occasions when people can be deceived by fake ATMs.

When we are dealing with mobile devices and the Internet of Things (IoT), we do not consider its physical location, since these devices are portable and can be found in different locations.

Additionally, The Internet of Things devices are supposed to interact with other similar devices or central systems. Therefore, it is of paramount importance the verification of these devices with the other ones in order to assure the proper operation of the entire system. As an example of a significant authentication method, we can mention the airplane identification system and its interaction with the air traffic control system at land, where security and verification of the different actors is of critical importance.

These systems should consider other systems authentication methods when connecting to such systems in order to provide or deliver information. These verifications should have, as in the case of people, multiple sources of verification. For example, although existing a software authentication, it is needed a physical component (i.e. a short range radio label on most devices). This label should send a verification by the use of radio frequency (RFID: Radio Frequency Identification, or NFC: Near Field Communication) so that the different systems would be verified and managed by other devices, as if they were managed by the actors located at a short range distance. According to the information mentioned above, authentication is the most appropriate verification method using, at the same time, a combination of devices from different origins.

4 Two-Way Identification Technique

So far, the authentication case of people interacting with other people or other devices has been presented. According to this information the topic of two-way authentication method is introduced.

¿Why is it important that people and systems need to identify themselves reciprocally?

We have an evident answer if we are taking into account the new technologies in mobile devices, IoT devices and, specially, those devices connected to the human body, helping and/or directly performing some vital functions. Nowadays, bio-prosthesis are being designed by private companies and universities [10] to help the

handicapped perform different activities or to maximize such activities (exoskeleton). These devices would become parts of the human body and would have their independent mobility. At the same time, these bio-prosthesis would be communicating their activities to other systems, and receiving information from these activities with different degrees of priorities. These inter-connections, its authentication, action capability, security maintenance for exchanging information between the different actors and the ability of these actors, will be part of that security device. This aspect is very important regarding its use by the people.

Considering these difficulties, it is important to contemplate that the verification performed by a person needs to verify he/she is using the correct device. In this case, the person should verify and check identities in an easy but secure way. This previous situation can include third parties connected to the same device such as some kind of networks or the Internet. For this reason, it is of primary importance to consider security issues related to connectivity.

5 Connection Continuity (Availability)

The second important aspect presented in this paper is related to continual authentication. Normally, systems need a logging process or entering a password to have access to such system only one time at the beginning of a session or work day. After that, it is supposed that the person who has accessed the system is always the same user, but the session could expire due to a long period of inactivity. However, even though the control process of inactivity can take a short time, during this period, the user could have changed as well as the original system.

This is the reason why a continual authentication process is required, as we have not to be totally confidence that after having performed the first contact, it would be not necessary to do further authentications in order to effectively be sure that the original person and system remain connected and no identity theft has occurred.

It is also important the way the connection is established, and depending on the methods used it is necessary to implement different control levels. In the first place, the connection between two different actors is considered safe or unsafe.

In the first case, we have confidence that the transport channel could not be hijacked, and be unnoticed of such interrupted connection. However, telecommunications protocols ignore transmission micro-outages or even signal power drops due to communication interferences.

Nowadays, it is possible to tap any type of communication, including fiber optic networks, interrupting the connection for a short period of time and causing a damage in the fiber optic cable or causing a no significant decrease signal quality so that it will be necessary the use of special equipment to detect the network tap. There are some wireless methods such as evanescent coupling or optic dispersion [11].

Consequently, in order to avoid the possible loss of connection continuity of the original actors, it is necessary that the communication system carries out constant verifications and controls that such system is not being intercepted or hijacked by other devices.

6 Conclusion

The following paper presents the different stages that makes it necessary to develop an authentication model that takes into account these principal factors:

- It must be a two-way method and involves people, communication systems, network connections, and devices.
- The authentication process must not only take place at the beginning of communication but also ongoing verifications should also take place in a continuous way.

The aspects developed in this paper present the major difficulties and some aspects regarding scenarios, different examples and several components involved over the different authentication models. It also shows that a deeper analysis must be carried out in order to develop a detailed model or authentication outline to be used for different types of devices according to their use.

Additionally, different phases have been identified such as authentication and continuity, and their related alert/incident management procedures, so that the user is aware that system's unauthorized access events have occurred or the system can perform some actions in case that some suspicious activity is detected. According to this, perhaps it is necessary to develop artificial intelligence systems associated to the authentication process. But this discussion will be managed in a further analysis.

The market is focused on the development of independent but interconnected management devices, in order to use them for different needs, some of them related directly to the human being. However, this type of needs are growing faster than the security systems when using these particular devices, implying the need of a solution. This solution must take into account the variety of components of verification both for people and devices, in order to design appropriate authentication protocols for dealing with these major difficulties.

Finally, this paper could be taken as a start on an original proposal for a more specific continuous authentication model.

References

- [1] The University of Tokyo – Department of Biomedical Engineering. Artificial heart – Available at: http://www.bme.gr.jp/Research_info_E/AH.html, last accessed April 10 2017.
- [2] CES – Consumer Technology Association – Self Driving Technology. Available at: <http://www.ces.tech/Show-Floor/Marketplaces/Self-Driving-Technology>, last accessed April 10 2017.
- [3] Olufemi Sunday Adeoye, A Survey of Emerging Biometric Technologies - Department of Computer Science University of Uyo, International Journal of Computer Applications (0975 – 8887) Volume 9– No.10, November 2010 . Available at: <https://pdfs.semanticscholar.org/2c6c/4f84046be0a9b317f398bf1783e32e5ad771.pdf>, last accessed April 10 2017.

- [4] John Daugman. How Iris Recognition Works. IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 14, NO. 1, JANUARY 2004. Available at: <https://www.cl.cam.ac.uk/~jgd1000/csvt.pdf>, last accessed April 10 2017.
- [5] Antonio Iula, Alessandro, Stuart Savoia, Giosue Caliano – An ultrasound technique for 3D palmpoint extraction, Sensors and Actuators A: Physical Volume 212, 1 June 2014, Pages 18–24. Available at: <http://www.sciencedirect.com/science/article/pii/S0924424714001162> , last accessed April 10 2017.
- [6] A Winda, W R E Byan, Sofyan, Armansyah, D L Zariantin and B G Josep - Motorcycle Start-stop System based on Intelligent Biometric Voice Recognition OP Conf. Series: Materials Science and Engineering 187(2017) 0120 39 doi:10.1088/1757-899X/187/1/012039. Available at: <http://iopscience.iop.org/1757-899X/187/1/012039>, last accessed April 10 2017.
- [7] Joshua Wheeler, Brigitte Richardson, Scott Amman, Ranjani Rangarajan - Systems Engineering Approach for Voice Recognition in the Car - SAE International Journal of Passenger Cars - Electronic and Electrical Systems 10(1) · March 2017 - DOI: 10.4271/2017-01-1599
- [8] Fumihiko Adachi, Ryosuke Isotani, Ken Hanazawa - Voice recognition system, voice recognition method, and program for voice recognition . Patent: US8639507B2 · License: USPTO TOS Date of Patent: Jan 28, 2014. . Available at: <https://www.google.com/patents/US8639507>, last accessed April 10 2017.
- [9] P.Inbavalli , G.Nandhini, Body Odor as a Biometric Authentication - / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (5) , 2014, 6270-6274 Available at: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.661.6164&rep=rep1&type=pdf> , last accessed April 10 2017.
- [10] MIT –Massachusetts Institute of Technology – d'Arbeloff Laboratory for Information Systems and Technology. Available at: http://darbeloff-lab.scripts.mit.edu/darbeloff-lab/?page_id=296, last accessed April 10 2017.
- [11] MillerSandra Kay, Fiber optic networks vulnerable to attack - Revista Information Security – November 2006. Available at: <http://searchsecurity.techtarget.com/news/1230106/Fiber-optic-networks-vulnerable-to-attack>, last accessed April 10 2017.

Verificabilidad “End to End” para OTP - Vote

Pablo García¹, Silvia Bast¹, Germán Montejano^{1,2}

¹ FCEyN (UNLPam) – Uruguay 151, Santa Rosa, La Pampa
{pablogarcia, silviabast}@exactas.unlpam.edu.ar

² FCFMyN (UNSL) – Ejercito de Los Andes 50, San Luis, Argentina
gmonte@unsl.edu.ar

Abstract- OTP – Vote es un modelo para votación electrónica basado en One Time Pad que utiliza claves múltiples y distribuidas que funcionan finalmente como una llave única. El esquema garantiza anonimato incondicional en la medida en que se verifiquen las condiciones iniciales exigidas y, simultáneamente, permite llevar la seguridad computacional del escrutinio a los niveles que se deseen. Como cualquier propuesta relacionada con voto electrónico, uno de los principales inconvenientes a la hora de la implementación pasa por convencer a la sociedad de que el esquema es seguro. Aportando en ese sentido, el presente documento expone una técnica para aplicar verificabilidad “End to End” al esquema original.

Key words: E-Voting – OTP - Vote – Parallel Channels – Verificabilidad “End to End” – Funciones de Hash – MCDU – Birthday Paradox.

1. Introducción

La implementación de sistemas de voto electrónico se encuentra en la actualidad en un momento de transición. Existe una cantidad de argumentos atendibles que dan lugar a que desde muchos sectores^{1,2} se afirme que los productos de software diseñados para tal fin no permiten garantizar la transparencia del proceso electoral.

En general, gran parte de la sociedad percibe como una enorme “caja negra” lo que pueda ocurrir con los sufragios una vez que se complete la emisión de los mismos. Si bien no se justifica razonablemente que esa incertidumbre sea mayor que en un sistema de votación manual tradicional, la realidad nos indica que ésa es la percepción dominante. Como consecuencia de lo anterior, todos los aportes que puedan realizarse en búsqueda de garantizar la transparencia del proceso de votación electrónica adquieren relevancia.

¹ <http://www.infobae.com/opinion/2016/10/31/problemas-que-el-voto-electronico-trae-y-los-que-no-evita/>

² <http://www.laizquierdadiario.com/Voto-Electronico-un-sistema-vulnerable-que-no-garantiza-el-voto-secreto>

En particular, la Verificabilidad “End to End” (E2E, [1], [2], [3]) es uno de los puntos de mayor valor a los efectos de agregar transparencia al proceso de votación electrónica. Se define mediante las tres condiciones siguientes:

- Verificabilidad individual: cualquier votante puede verificar que su sufragio fue incluido en el recuento.
- Verificabilidad universal: cualquier persona puede determinar que el recuento total de los votos es correcto.
- Secreto del voto: ningún votante podrá demostrar cuál fue la opción que eligió, a los efectos de evitar maniobras relacionadas con el “clientelismo político”.

2. OTP - Vote

El Modelo OTP-Vote, presentado en [4], se basa en la premisa de que en cualquier sistema de voto electrónico es necesario proteger:

- Indefinidamente la privacidad del votante ([5]).
- La seguridad de los datos del escrutinio mientras dure el proceso electoral.

La protección de los sufragios anónimos sólo debe soportar el lapso de tiempo que corresponda al proceso de votación ([6]). Luego, la información se hará pública.

En *OTP - Vote*, el proceso eleccionario consiste de tres grandes etapas:



Figura 1: Etapas del Proceso Eleccionario

La Figura 2 expone las etapas con sus datos de entrada y salida.

Los elementos de datos que aparecen en el Modelo son:

- Claves: el modelo hace uso de claves One Time Pad (OTP, [7]), que está basado en el *Secreto Perfecto* de Shannon [8]: las mencionadas claves son totalmente aleatorias y tan largas como el mensaje claro.
- Archivos de datos que almacenan bits. son elementos fundamentales en el Modelo propuesto y se van modificando durante el proceso eleccionario:
 - Clave de Descifrado (*CD*): se genera a partir de operaciones XOR (\oplus) de claves OTP.
 - Archivo Binario de Votos (*ABV*) se genera en base al modelo de almacenamiento Múltiples Canales Dato Único (MCDU) propuesto en [9], que se muestra en la Figura 3. El esquema apunta a resolver las limitaciones presentadas por Birthday Paradox [10] y se analiza en profundidad en [11] y [12].
 - Tablas Relacionales: se usan para almacenar los datos de los cargos, candidatos e identificadores de votos. También los votos emitidos, una vez finalizado el proceso eleccionario

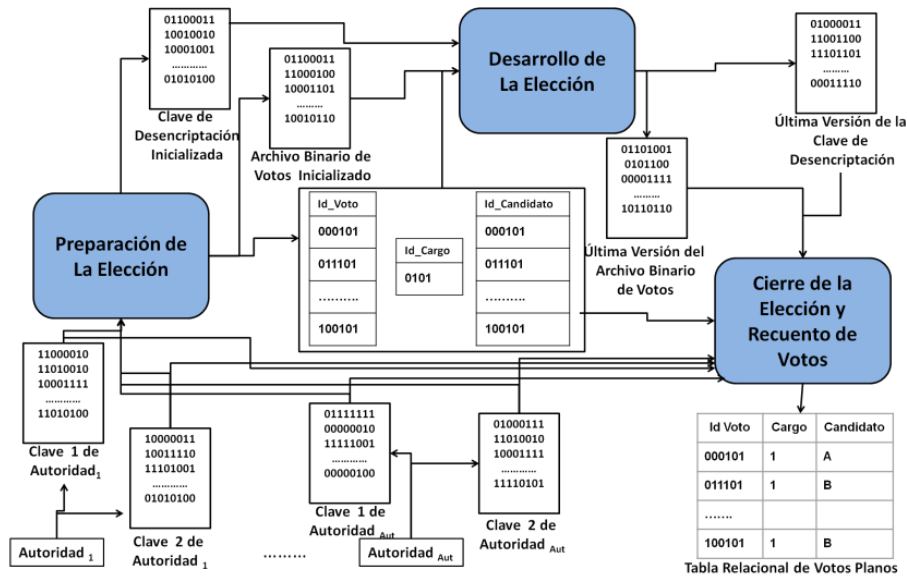


Figura 2: Etapas del Proceso Eleccionario con Datos de Entrada y Salida

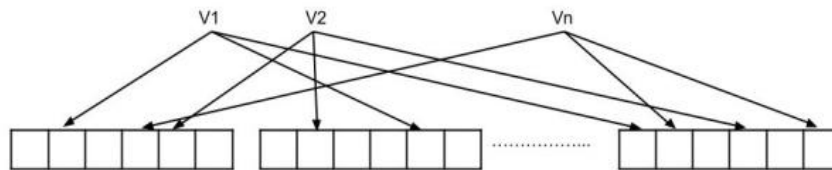


Figura 3: Propuesta de Almacenamiento de MCDU

Se describen a continuación las actividades que se llevan a cabo en cada una de las etapas del proceso eleccionario:

Etapas de Preparación de la Elección: las acciones que deben llevarse a cabo en esta etapa son:

- Especificación de la semántica de la tupla en la que se almacenarán los votos:
 - #bits de cada slot (*TBslot*).
 - #bits del Identificador de Voto (*TBId*) y su ubicación.
 - #bits asignados para almacenar el cargo (*TBcargo*) y su ubicación.
 - #bits asignados para almacenar código del candidato votado para cada cargo de la elección (*TBCandidato*) y su ubicación.
- En cuanto a la definición de las dimensiones de los atributos, es necesario evaluar la probabilidad de que algún intruso pueda obtener un dato válido de entre todos los posibles. A través del aumento de la redundancia en la cantidad de bits usados para el almacenamiento de cada uno de los atributos, la probabilidad de obtener una tupla válida de entre todas las combinaciones de valores posibles puede llevarse a cualquier valor deseado.

- Definir las dimensiones del ABV y CD , según las fórmulas establecidas en [9].
- Generar las tablas: *Cargos*, *Candidatos* e *Identificadores de Votos*, con las características propuestas en [4]
- En el momento previo al comienzo del acto eleccionario se inicializan el ABV y la CD , con el aporte de las Claves OTP de las autoridades electorales, las mismas se usarán en las etapas de preparación y cierre de la elección. En el transcurso del proceso debe garantizarse que las mismas se encuentren seguras y aisladas.
- Las claves K_{i1} aportadas por cada una de las CA autoridades servirán para dar valores a ABV . Inicialmente el mencionado archivo tendrá asignado cero en cada una de sus posiciones, luego se aplicará el XOR de cada una de las K_{i1} aportadas por las autoridades de la Junta electoral:

$$ABV = ABV \oplus K_{i1} \quad \forall i \ 1 \leq i \leq CA \quad (1)$$

- Las claves K_{i2} de las autoridades inicializarán CD , que en un principio se encuentra con todos sus elementos en cero.

$$CD = CD \oplus K_{i2} \quad \forall i \ 1 \leq i \leq CA \quad (2)$$

Etapas de Desarrollo de la Elección: se llevan a cabo las actividades que se detallan a continuación

- Autenticación del elector: consiste en verificar que quien va a votar figure dentro del padrón electoral, es decir que sea un votante válido. El modelo propone proceder de la forma habitual, registrándose el usuario en el lugar de la elección, con la presencia de una autoridad de la Junta Electoral. Para cada votante la autoridad habilitará un único sufragio.
- Emisión del voto: para el almacenamiento de los Votos se sigue el esquema MCDU.
- Para cada voto, el sistema genera una clave OTP (KV_v) de dimensión $TBslot$ bits que tendrá un doble propósito:
 - Aportar a la CD final de los votos, mediante operaciones XOR.
 - Cifrar la información del voto.

Debe incluirse un protocolo que garantice que KV_v se mantenga inalterable para los dos usos en los que es necesaria.

El elector genera su voto que combinado con la clave produce la Contribución Final de Voto (CFV).

El aporte de la clave de voto a la CD se lleva a cabo a través de la siguiente operación:

$$CD = CD \oplus KV_v \quad \forall v \ 1 \leq v \leq N \quad (3)$$

Para cada Voto se genera la Cadena de Voto CV_v formada por: Id_de_Voto (asignado aleatoriamente), Id_de_Cargo y además el $Id_Candidato$ seleccionado.

El sistema genera una cadena de TBSlot bits CI_v , con todos sus elementos en cero. Produce además un conjunto de números aleatorios $CjtoQ = \{q_i\}$ para cada uno de los Q canales, donde q_i representa el lugar donde se almacenará el voto en el canal i -ésimo.

Se realiza el XOR de la CV_v con los slots que corresponden a los q_i de la CI_v . Esto es:

$$Contribución_{vi} = CV_v \oplus CI_{vi} \quad \forall i \in CjtoQ \quad (4)$$

Finalmente se aplica:

$$CFV_v = Contribución_v \oplus KV_v \quad (5)$$

$$ABV = ABV \oplus CFV_v \quad (6)$$

Etapas de Cierre de la Elección y Recuento de Votos: Al momento de cierre de la elección se requiere la intervención de las Autoridades de la Junta Electoral.

El proceso de descifrado de los votos consta de tres sub-procesos:

- Aplicación de las K_{i1} de las Autoridades (las mismas que se usaron en la etapa inicial) a la última versión del ABV .

$$ABV = ABV \oplus Ki1 \quad \forall i \leq CA \quad (7)$$

- Aplicación de las K_{i2} de las Autoridades las mismas que se usaron en la etapa inicial) a la última versión de la CD .

$$CD = CD \oplus Ki2 \quad \forall i \leq CA \quad (8)$$

- XOR entre el ABV y la CD resultantes de los pasos anteriores que genera el Archivo Binario de Votos Descifrado ($ABVD$).

$$ABVD = ABV \oplus CD \quad (9)$$

- Finalmente, el recuento se realiza de la siguiente manera:
 - Se eliminan las tuplas que corresponden a votos vacíos.
 - Se eliminan las tuplas con votos que se generaron por colisiones y no se corresponden con la información de ninguno de los Id de voto.
 - Se recorre el $ABVD$ generándose la tabla de Votos Planos.
 - Luego se produce el conteo de los votos por medio de una consulta SQL.

3. Propuesta para verificabilidad “End to End”

Habiendo expuesto el funcionamiento detallado de OTP – Vote, se presenta a continuación una técnica que permite proveer al mismo de Verificabilidad *E2E*:

1. Se agrega un campo H a cada fila del archivo ABV de $OTP-Vote$. Nótese que será necesario modificar también el tamaño de CD .
2. Para cada sufragio, se genera un número aleatorio grande G , que se informa al votante y se hace público.
3. Se almacena en H el valor de una función $HASH(G)$.
4. Al finalizar el proceso se verifica que los números generados y publicados son coherentes con los valores almacenados en H .

Es necesario plantear una serie de condiciones iniciales que resultarán imprescindibles para garantizar la transparencia del proceso:

- Las claves K_{i1} y K_{i2} de la autoridad i -ésima se elegirán de una cantidad de claves mucho mayor que las que efectivamente se utilizarán. De ese grupo, serán seleccionadas de manera auténticamente aleatoria. El contenido de todas esas claves es información fundamental que debe ser protegida de manera incondicional.
- Cada clave incluirá una serie de commitments (que podrían colocarse en sobre cerrado). Por ejemplo, podría incluirse el resultado de h funciones de hash con $(h \in Z^+) \wedge (1 \leq h)$ aplicadas al contenido de dicha clave. Al finalizar el acto eleccionario, se publica cuales fueron las funciones utilizadas. Estas funciones serán similares para todas las claves de tipo K_{i1} y K_{i2} .
- El éxito del modelo exige que las siguientes informaciones críticas sean protegidas de manera inviolable:
 - El conjunto total de claves de tipo K_{i1} .
 - El subconjunto de claves de tipo K_{i1} que será efectivamente utilizado.
 - El conjunto total de claves de tipo K_{i2} .
 - El subconjunto de claves de tipo K_{i2} que será efectivamente utilizado.
 - El estado de los archivos ABV y CD luego de aplicar las claves K_{i1} y K_{i2} .

En esas condiciones, definimos:

Y : #Tuplas generadas.

N : #Votantes.

X_i : Valor obtenido en la i -ésima posición de la tupla.

C : #Componentes de la tupla. En todos los ejemplos posteriores se le asigna un valor 4, pero debe analizarse si otro valor es más conveniente.

Para ilustrar el funcionamiento del modelo propuesto, se implementa un procedimiento que, dado un valor de N , contabiliza cuántas de todas las tuplas posibles cumplen con la condición necesaria:

$$\sum_{i=1}^C X_i = N \quad (10)$$

Luego, utilizando el software online SECANU³ ([13]), por aplicación de los métodos de interpolación de Lagrange, se obtiene la fórmula para definir la cantidad de cuádruplas diferentes que pueden generarse en función de N :

$$Y = \frac{1}{6} N^3 + N^2 + \frac{11}{6} N + 1 \quad (11)$$

Suponiendo que la función de hash elegida sea $H = G \bmod 4$, la Tabla 1 muestra la cantidad de tuplas posibles para diferentes valores de N .

N	Y
12	455
24	2925
36	9139
48	20825
60	39711
72	67525
84	105995
96	156849
108	221815
120	302621
132	400995
144	518665
250	2667126

Tabla 1: Variación de Y para diferentes valores de N .

Un valor de $Y = 250$ se aproxima a la cantidad de votantes promedio de una mesa electoral en la Argentina en la actualidad. Las cuádruplas no son equiprobables, pero todas tienen probabilidad mayor que cero.

A los efectos de analizar la validez de la propuesta, se realizó una simulación con 10.000.000 de corridas para elecciones con $N = 11$ votantes. Aplicando (11), se obtiene un valor de $Y = 364$ tuplas. En la Figura 4 se numeran las mismas desde 1 (11,0,0,0) hasta 364 (0,0,0,11) y se observa la distribución obtenida.

Es importante destacar algunos elementos que ocurrieron en la simulación:

- La totalidad de las tuplas posibles aparecieron en la simulación.
- Las que aparecen con mayor frecuencia son aquellas en las que los cuatro valores de la tupla (que sumados deben dar Y , en este caso 11) se aproximan a $N/4$.
- La suma de las 38 tuplas más frecuentes supera el 50% de las apariciones.
- La cuádrupla más frecuente (*CMF*) apareció 220.291 veces, es decir menos de 23 veces cada mil corridas.

³ <http://secanu.exactas.unlpam.edu.ar/>

Parece lógico afirmar que si se aumenta el valor de N , se observará una disminución en la frecuencia de aparición de la CMF . Se implementa un nuevo simulador que contabiliza la cantidad de veces que aparece la misma, los resultados se muestran en la Tabla 2.

Cuando se realiza una sesión del simulador, es lógico que las cuádruplas más probables sean aquellas en las que todos los valores son cercanos al promedio. Esto se debe a la distribución uniforme de la variable aleatoria:

$$X_i = \text{"Valor de la función de Hash para el } i\text{-ésimo voto"}$$

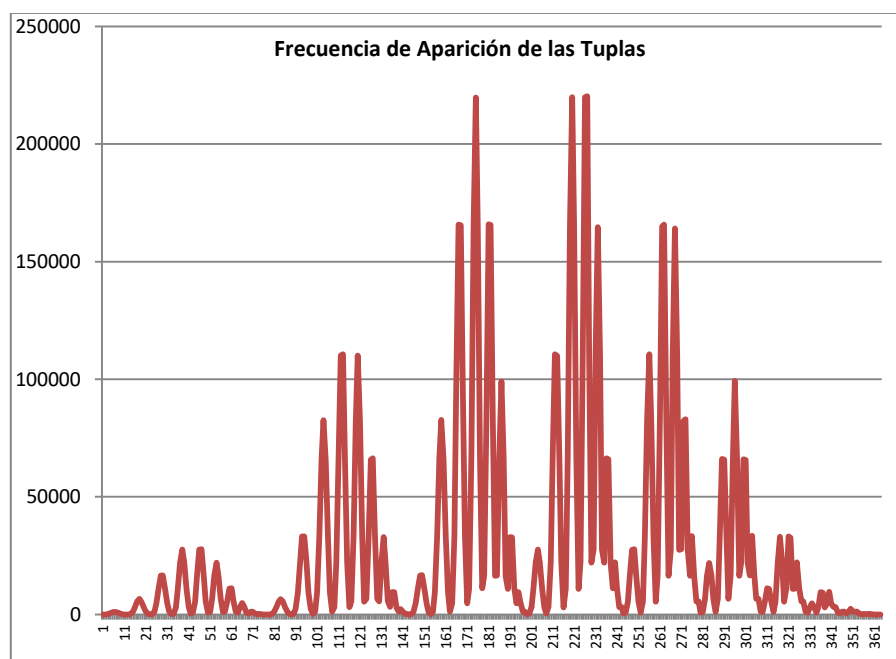


Figura 4: Frecuencia de Aparición de las Cuádruplas

N	CMF	$F(CMF)$	%	Repeticiones
5	2-1-1-1	5888	0,05888	100.000
10	3-3-2-2	2517	0,02517	100.000
15	4-3-4-4	1506	0,01506	100.000
20	5-5-5-5	1132	0,01132	100.000
25	6-6-7-6	787	0,00787	100.000
30	7-8-8-7	597	0,00597	100.000
35	9-9-9-8	460	0,0046	100.000
40	11-10-10-9	399	0,00399	100.000
45	11-12-11-11	343	0,00343	100.000

Tabla 2: Variación de los Valores de CMF en Función de N

Sin embargo, aún la CMF tiene un valor muy bajo. Que, además, se va reduciendo cuando aumenta el valor de N . Esto es promisorio. Para completar la evaluación de esta situación, en el simulador se realizan dos sesiones complementarias. En la primera se hace una sola sesión y se anota la Cuádrupla obtenida, (CO) y a continuación se realiza una corrida con 100.000 repeticiones y se observa en cuántas de ellas el resultado coincide con CO . Los resultados de las simulaciones se muestran en la Tabla 3. En ella, se define TC como el número total de cuádruplas posibles.

N	CO	$F(CO)$	%	TC
20	2-7-3-8	89	0,00089	1771
20	3-6-6-5	582	0,00582	1771
20	3-5-9-3	161	0,00161	1771
20	5-2-5-8	184	0,00184	1771
20	4-4-7-5	601	0,00601	1771
20	3-5-6-6	608	0,00608	1771
20	7-3-4-6	416	0,00406	1771
20	2-2-6-10	31	0,00031	1771
30	5-6-9-10	221	0,00221	5456
30	10-5-11-4	55	0,00550	5456
30	7-9-9-5	307	0,00307	5456
30	7-10-7-6	362	0,00362	5456

Tabla 3: Frecuencias de Aparición de Cuádruplas Específicas

4. Conclusiones y Trabajos Futuros

Se presenta en el presente documento una propuesta integral para proveer de verificabilidad $E2E$ al sistema de voto electrónico OTP - Vote. La misma debe interpretarse de manera integral, dado que incluye una serie de componentes que proveen seguridad solamente si todos ellos se aplican de manera conjunta.

Desde el punto de vista práctico, la publicación del número aleatorio grande generado para cada sufragio, otorga la garantía al votante de que su voto fue tenido en cuenta. Simultáneamente, el chequeo de corrección de la tupla obtenida al final de la elección, contra la totalidad de los valores publicados da la seguridad de que la totalidad de los votos fue computada correctamente si se verifican las condiciones exigidas.

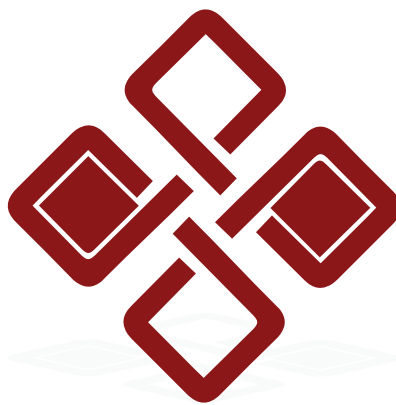
A futuro, es necesario hacer un profundo análisis sobre las funciones de hash que se aplicarán y sobre los parámetros correspondientes. Si bien aún la combinación más probable tiene una probabilidad muy baja, debe trabajarse para optimizar ese comportamiento. En el nivel ideal, todas las combinaciones serían equiprobables. En cualquier caso, debe procurarse minimizar la diferencia entre las combinaciones más y menos probables. El camino a seguir pasa por investigar el comportamiento de otras funciones de hash y, simultáneamente, verificar el comportamiento aumentando el valor de C . Si bien intuitivamente parece lógico que aumentar ese valor dará como resultado la obtención de mejores resultados, debe controlarse que no aparezcan “votos aislados” que permitirían probar por quién votó un elector específico. Ese

fenómeno se produce si todos los votantes que generaron el mismo valor de hash votaron por el mismo candidato. Obviamente, el riesgo de que eso ocurra aumenta cuando el valor de C es mayor.

Reconocimientos. A Jeroen van de Graaf, PhD., por su enorme generosidad para con el Mg. Pablo Garcia durante sus estadias en la Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Minas Gerais, Brasil, en 2012 y 2016.

Referencias

1. Ryan P., Schneider S., Teague V.: "End-to-End Verifiability in Voting Systems, from Theory to Practice". Voting Systems, from Theory to Practice. IEEE Security & Privacy, 13(3):59–62, 2015.
2. Rabin, M., Rivest, R.: "Efficient End to End Verifiable Electronic Voting Employing Split Value Representations" Bregenz, Austria. EVOTE 2014. ISBN 978-9949-23-688-6.
3. Awad M., Leiss E.: "End-to-End Cryptography: Spreading Democracy". International Journal of Applied Engineering Research. Volume 11, Issue 11. Ps. 7391-7394. 2016.
4. Bast, S. "Confidencialidad e Integridad de Datos en Sistemas de E-Voting – Un Modelo para la Implementación Segura de un sistema de Voto Presencial" - Editorial Académica Española.- ISBN 978-3-639-53793-2. 2017
5. Van de Graaf J.: "Anonymous One Time Broadcast Using Non Interactive Dining Cryptographer Nets with Applications to Voting". Towards Trustworthy Elections". Ps. 231-241. Springer-Verlag Berlin, Heidelberg. ISBN:978-3-642-12979-7. 2010.
6. Bast S., Montejano G., García P., Fritz E.: "Evaluación de la Integridad de Datos en Sistemas de e-Voting". XVII Workshop de Investigadores en Ciencias de la Computación (WICC 2015). 2015. Universidad Nacional de Salta. Ps. 827 -831. ISBN: 978-987-633-134-0. <http://sedici.unlp.edu.ar/handle/10915/46871>.
7. N. Nagaraj, V. Vaidya, and P. G. Vaidya, "Revisiting the one-time pad," International Journal of Network Security, vol. 6, no. 1, pp. 94-102, 2008.
8. Shannon, C. E.: "Communication Theory of Secrecy Systems"- Bell System Technical Journal - N° 28 (1949) 656–715.
9. García, P.: "Una Optimización para el Protocolo Non Interactive Dining Cryptographers" - ISBN-13: 978-3-639-85270-7. ISBN-10: 3639852702. EAN: 9783639852707. Editorial Académica Española. <https://www.eac-publishing.com/> - 2017.
10. García, P., van de Graaf J., Hevia A., Viola A.: "Beating the Birthday Paradox in Dining Cryptographers Networks". En "Progress in Cryptology – Latincrypt 2014". Springer International Publishing. ISSN: 0302-9743. ISSN (electrónico): 1611-3349. ISBN: 978-3-319-16294-2. ISBN (eBook): 978-3-319-16295-9. Ps. 179 – 198. Octubre, 2014.
11. García P., van de Graaf J., Montejano G., Riesco D., Debnath N., Bast S.: "Storage Optimization for Non - Interactive Dining Cryptographers (NIDC)". The International Conference on Information Technology: New Generations. 2015. Las Vegas, Nevada, USA. <http://ieeexplore.ieee.org/document/7113449/>.
12. García P., Bast S., Fritz E., Montejano G., Riesco D., Debnath N.: "A Systematic Method for Choosing Optimal Parameters for Storage in Parallel Channels of Slots". IEEE International Conference on Industrial Technology (ICIT 2016). 14 - 17 March 2016 / Taiwan, Taipei. <http://ieeexplore.ieee.org/document/7475019/>.
13. Ascheri M., Pizarro R., Astudillo G., García P., Culla M., Pauletti C.: "Software Educativo para la Resolución Numérica y Gráfica de Temas de Cálculo Numérico". WICC 2017. Instituto Tecnológico Buenos Aires. 2017.



CACIC 2017



UNIVERSIDAD
NACIONAL
DE LA PLATA

www.cacic2017.info.unlp.edu.ar | cacic2017@lidi.info.unlp.edu.ar

UNIVERSIDAD NACIONAL DE LA PLATA | FACULTAD DE INFORMÁTICA | 50 y 120 - LA PLATA - 427-7270